

Copyright © 2001, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

**COMPLEMENTARY SILICIDE THIN-BODY
SILICON-ON-INSULATOR CMOS DEVICES**

by

Jakub Tadeusz Kedzierski

Memorandum No. UCB/ERL M01/21

1 March 2001

**COMPLEMENTARY SILICIDE THIN-BODY
SILICON-ON-INSULATOR CMOS DEVICES**

by

Jakub Tadeusz Kedzierski

Memorandum No. UCB/ERL M01/21

1 March 2001

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Complementary silicide thin-body silicon-on-insulator CMOS devices

by

Jakub Tadeusz Kedzierski

B.S. (Ohio State University) 1995

M.S. (University of California, Berkeley) 1999

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering
and Computer Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Jeffrey Bokor
Professor Chenming Hu
Professor Ronald Gronsby

Spring 2001

Complementary silicide thin-body silicon-on-insulator CMOS devices

Copyright 2001

by

Jakub Tadeusz Kedzierski

Abstract

Complementary silicide thin-body silicon-on-insulator CMOS devices

by

Jakub Tadeusz Kedzierski

Doctor of Philosophy in

Engineering – Electrical Engineering
and Computer Sciences

University of California, Berkeley

Professor Jeffrey Bokor, Chair

The thin-body silicon on insulator (SOI) transistor is a promising design for the 10-50nm gate length regime. One of its major challenges is the large series resistance of the thin SOI layer. In this work a new thin-body device structure is presented that reduces this resistance by fabricating the source/drain regions out of two low-barrier silicides, one for NMOS and one for PMOS. This device is fabricated with gate lengths as small as 15nm using aggressive electron beam lithography techniques, to demonstrate its immunity to short channel effects. The two complementary silicides used are: PtSi for PMOS, and ErSi_{1.7} for NMOS. The devices are fabricated without any doping in the source, drain, or body. A secondary structure is also proposed; adding doped extension regions to the complementary silicide source/drains decreases the influence of the Schottky barrier on current transport. This doped complementary silicide source/drain thin-body structure functions in a manner that is similar to conventional thin-body transistor.

The fabrication of the undoped structure is described in detail, in particular the silicide formation and the electron beam lithography steps. Electrical results for both NMOS and PMOS devices are presented with functional devices down to 15nm gate-length. A transmission model is used to fit the experimental data and to extract the silicide barrier height. This model is also used to examine the influence of oxide thickness scaling and extension doping on expected device performance.

A 2d device simulator, Fielday2d, is used to examine the design space of the thin-body complementary silicide source/drain devices. The doped and undoped designs are compared and the influence of the relevant structure parameters is studied. Simulations suggest that a fully depleted source complementary silicide thin-body structure may exhibit the lower leakage current of the undoped structure and the higher on current of the doped structure.

I would like to dedicate this work to a man whose
calm thoughtfulness has always inspired me,
my grandfather Professor Tadeusz Kowalak.

Table of Contents

1.0	Introduction	1
	References	8
2.0	Background/Literature	11
2.1	Metal-semiconductor junctions	11
2.1.1	Conventional metal-semiconductor junction resistance model	12
2.1.2	Equivalent barrier lowering model	14
2.1.3	Flat-band barrier height definition	17
2.2	Synopsis of work relevant to the quantitative understanding of thin-body source/drain engineering requirements	18
2.3	Synopsis of previous silicide source/drain device research	22
2.3.1	Design space for the silicide source/drain device	24
	References	26
3.0	Thin-body complementary silicide source/drain devices	28
	References	35
4.0	Description of Device Fabrication	36
4.1	Overview of process flow	36
4.2	Critical process windows	38

4.3	Specific processing modules	39
4.3.1	Lithography – Electron beam lithography fundamentals	40
4.3.2	Lithography process	45
4.3.2.1	Resist processing	45
4.3.2.2	Alignment process	51
4.3.3	Silicide formation – Physical vapor deposition fundamentals	52
4.3.3.1	Surface preparation and clean	52
4.3.3.2	Overview of metal deposition techniques	53
4.3.3.3	Film coverage during deposition	54
4.3.3.4	Deposition vacuum requirements	55
4.3.3.5	Details of evaporation	56
4.3.3.6	Structure zone model	57
4.3.3.7	Silicidation reaction	59
4.3.4	Silicide formation process	60
4.3.4.1	Erbium silicide process	61
4.3.4.1.1	Vacuum requirements and Er evaporation	61
4.3.4.1.2	Erbium silicide thermal processing	64
4.3.4.1.3	Erbium wet etching	67
4.3.4.1.4	Erbium thickness considerations	68
4.3.4.2	Platinum silicide process	69
4.4	Devices at the end of processing	71
	References	73

5.0	Thin-body complementary silicide device results and modeling	76
5.1	UCS CMOS device performance	76
5.2	Transmission model	79
5.2.1	Schottky barrier lowering model	82
5.2.2	Transmission model fitting	85
5.2.3	Methods to improve UCS device performance	86
5.3	TiSi ₂ source/drain devices	89
5.4	Fielday2D simulation of thin-body CS devices	91
5.4.1	Undoped CS device simulation	91
5.4.2	Doped CS device simulation	96
	References	101
6.0	Discussion	102
6.1	Comparison of DCS and UCS structures	102
6.2	The design of DCS doped extension regions	105
6.3	Geometry considerations for UCS devices	109
6.4	3D integration of UCS structures	112
6.5	CMOS integration of CS devices	112
6.6	Conclusion	113
	References	115
	Appendix A – Details of fabrication procedure	116

Acknowledgements

The first person I would like to thank dearly is my advisor professor Jeffrey Bokor. Throughout the years he has given me good advice when I asked for it, and independence when I needed it. His guidance has helped me to intensify my desire to function as a part of the scientific community, while at the same time giving me the skills I need to fulfill this desire.

I would also like to extend my thanks to professors Chenming Hu and Tsu-Jae King, for their invaluable advice during our four-year interaction through the 25nm transistor project. Not many students have the benefit of getting the perspective of three professors on their research work. This experience has helped me to put my work into a proper context, and critically evaluate my own ideas and those of others.

I have many fellow graduate students to thank for their help and support. In particular I would like to thank Peiqi(Patrick) Xuan, our discussions have been and I hope will continue to be both fun and fruitful. Patrick's amazing command of the physical reality has always given me a venue from which to bounce off my research ideas. It is during one of our discussions that the idea for this work was first conceived. The other students I would like to thank for their friendship as well as advice are: Mark Cao, Nick Lindert, Wen-Chin Lee, Stephen Tang, Leland Chang, Xuejue Huang, Troy Clear, Qiang Lu, Yee-Chia Yeo, Pushkar Ranade.

Thanks also to Erik Anderson, Bruce Harteneck from the Lawrence Berkeley National Laboratory. Without their help with the challenging lithography process, this project would be impossible. The consistent help they provided, for this project and

others within our group, is greatly appreciated. Erik Anderson's world-class electron beam lithography expertise has made it possible for our group to make some of the world's smallest silicon devices.

I would like to thank the Berkeley Microlab staff for providing the tools with which we can do high quality research. Katalin Voros and Robert Hamilton do a great job of running such a complex lab. In particular, I wish to extend a special thanks to Charlie Williams and Steve Gidden, they have helped me on countless occasions by fixing critical equipment. They put in long hours making sure the experiment can progress, and I truly appreciate that.

I would like to acknowledge financial support from the Intel Fellowship, and research support from DARPA, through the 25-nm transistor project. Many thanks also go to MeiKei Jeong at IBM for his help with Fielday2d. Working with Meikei is the main reason I really enjoyed my collaboration with IBM, an experience that has influenced my career choices to no small degree.

Finally I would like to thank my wife, Raya, who has made these years a great joy. Every morning I wake up looking forward to another day of our life together.

1.0 Introduction

Rapid scaling of silicon devices has been a catalyst for the speed and power-consumption improvements of consumer electronics. These improvements are made possible by the simple fact that smaller transistors work faster and can be made to consume less power than larger transistors.

As transistor dimensions shrink new operating conditions have to be chosen, so that the design can be optimized around them. There are two approaches to choosing new operating conditions: constant voltage scaling, when V_{dd} is kept constant, and constant field scaling, when the gate field is kept constant. Constant voltage scaling of transistors primarily produces gains in speed, while constant field scaling primarily produces gains in power consumption. Regardless of the scaling method, the most challenging aspect of scaling is maintaining a low off-current.

Although the exact device design optimizations are complex, in general low off-current is maintained by decreasing oxide thickness and source/drain junction depth. The silicon dioxide used for gate dielectric and doped silicon used for the source/drain regions have scaled with each technology generation for 30 years, proving their robustness. However if scaling trends continue at the current pace fundamental challenges for both SiO_2 gate oxide and doped silicon source/drain will be reached, Fig. 1.1[1].

	2001	2002	2003	2004	2005
Oxide thickness (nm)	1.5-1.9	1.5-1.9	1.5-1.9	1.2-1.5	1.0-1.5
Gate leakage (nA/um)	8	10	13	16	20
Drain ext. X_j (nm)	30-50	25-43	24-40	20-35	20-33
Drain sheet res. (Ω/sq)	280-730	250-700	240-675	220-650	200-625
Drain abruptness (nm/dec)	3.4	2.9	2.7	2.4	2.2

Figure 1.1: Relevant sections of the International Technology Roadmap for Semiconductors: 1999. Red(dark) regions indicate targets that have no known solutions. Yellow(gray) regions indicate targets which are being pursued.

Oxide scaling will be limited by the finite tunneling current through the thin oxide; such current grows exponentially with oxide thickness and will limit SiO₂ scaling to a thickness around 12Å[2,3,4]. Scaling of the junction depth will probably be limited by junction leakage current[5,6], which goes up as body doping increases. These trends will make it difficult to scale silicon bulk technology below 40nm gate-length.

The search for a replacement to bulk silicon technology has been a vibrant topic of research for many years. Designs with relatively small deviations from the traditional structure, such as silicon-on-insulator(SOI) technology[7,8,9], have already been adapted by a part of the electronics industry. Research on fundamentally different devices such as quantum computers[10,11] and carbon nanotubes[12,13] is promising, but fabrication and integration problems seem very challenging.

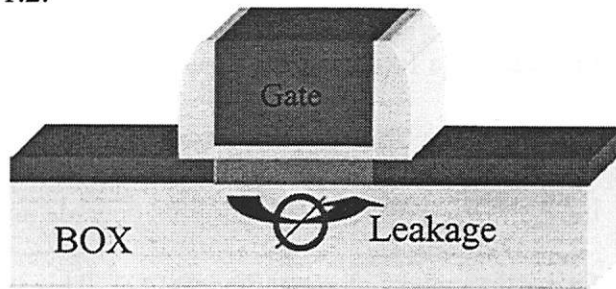
Due to the complexity of reliably integrating millions of transistors on a chip, large deviations from the bulk design are often rejected by the industry. It seems likely that silicon devices will evolve to new forms only when the magnitude of the change is justified by the new design's gain in performance. An example of such a change is SOI. In order to limit source-to-body capacitance SOI transistors eliminate a large part of the body under the device and replace it with an insulator, usually silicon dioxide. Although the change from bulk to SOI involves very little modification of the basic transistor structure and is basically a change in substrate, many design and fabrication problems had to be overcome for the SOI structure to be implemented in VLSI.

The current implementation of SOI devices uses a relatively thick silicon layer of 500~1000Å[8]. This design leaves a part of the body under the gate undepleted, and is therefore called partially depleted SOI design. The major disadvantage of partially

depleted SOI devices is that minority carriers generated at the drain by hot majority carriers tend to be trapped in the body, lowering the threshold voltage[14]. Partially depleted SOI devices are sufficiently similar to bulk-silicon devices that the similar materials and device geometry can be used for both. Therefore the threshold voltage is adjusted with body doping, and the drains are silicided with a mid-gap silicide to reduce series resistance.

SOI technology is also interesting because it may prove to be a stepping stone to a design that scales beyond the bulk-Si limit of 50-30nm gate-length. Decreasing the SOI silicon layer thickness to a value smaller than the gate-length results in a fundamentally different SOI structure shown in Fig. 1.2.

Figure 1.2: Generic single gate thin-body structure. The SOI thickness is 4x smaller than the gate-length, for good leakage control.



This thin-body, or fully depleted design, can be scaled to shorter channel lengths than bulk transistors because it doesn't rely on body doping to control off-current[15,16,17]. Off-current is reduced by eliminating the part of the body that conducts leakage current and replacing it with an insulator; only the silicon near the gate that carries the on-current remains. Thus the thin-body transistor doesn't suffer from the junction leakage that causes scaling problems for bulk-transistors. Thin-body designs come in two basic categories, with either a single or a double gate. Double-gate devices have better short-channel performance for the same body thickness, but are more significantly affected by the series resistance of the thin-body. This is due to the fact that a double-gate device has

the potential to conduct twice the current of a single-gate device. Simulations indicate that double-gate silicon devices can be scaled to the sub-10nm gate-length regime, where they are limited by the source-to-drain tunneling current, and may represent the smallest transistor design that can be implemented in silicon[18,19,20].

In addition to their superior scaling properties thin-body devices may have another fundamental advantage over traditional bulk transistors. Thin-body transistors can be designed to have a significantly smaller transverse field than bulk devices for the same inversion charge, because they don't rely on body doping to control short channel characteristics. A lower transverse field may lead to a higher current for the same gate overdrive voltage ($V_{gs}-V_t$). The classical explanation for this effect is that mobility degrades as a function of transverse electric field [21]. However in transistors shorter than 100nm mobility is not expected to play a significant role, still recent experiments have shown that transistor currents degrade as a function of transverse electric field even for ballistic short channel devices[22]. This is probably due to the influence of the transverse potential profile in the body on the ability of a carrier to transit from the low field source region into the channel.

Thin-body devices suffer from two fundamental challenges, and a myriad of fabrication and integration difficulties. One of the fundamental challenges of thin-body devices is the control of threshold voltage. Threshold voltage(V_t) of a transistor determines the fundamental trade-off between the off-current and on-current, and must be carefully engineered for specific technologies, taking into consideration power and speed requirements. Being able to obtain a reproducible and specific V_t is a requirement of any device technology. In bulk transistors V_t is set by doping the body and using a poly-

silicon gate, if the same V_t control scheme were to be used for an ultra-thin body transistor the body doping level would have to be very high, since there is much less of the body to dope. Any variations in the body thickness would result in large variations in the threshold voltage. High doping level in the body would also introduce a large transverse field at the channel decreasing the device current.

The favored method for controlling the V_t of thin body transistors is to use a gate material of different workfunction than polysilicon. Making the gate out of a material that has a work-function in the middle of the silicon band-gap for both NMOS and PMOS thin-body transistors would result in $|V_t|=0.45V$ [23]. While this value may be acceptable for low power applications it is too high for high-performance logic. To lower the V_t to the desired value of 0.2V two different gate materials have to be used, with workfunctions straddling the silicon mid-gap energy. The final tuning of the V_t could be performed with a body doping, or by properly engineering the gate material. Finding such materials and integrating them into a device manufacturing process flow is a challenge that still needs to be addressed.

The other significant design challenge of the thin-body devices is reducing the series resistance of the thin-body region. As shown in Fig. 1.2 thin-body devices in their simplest form lack the deep source/drain regions and silicide that reduce the series resistance of bulk transistors. The high series resistance of thin-body devices can be reduced by several different approaches: changing the source/drain geometry (method 1), and changing the source/drain material (method 2).

Method 1 is often associated with elevated source drain design, in which the source/drain regions are thicker outside of the spacers, Fig. 1.3a. Fabrication of such

structures is difficult since it requires selective growth techniques that may be impossible to integrate into a VLSI fabrication due to their high defect density levels[24,25].

Elevated source/drain designs also suffer from increased Miller capacitance, a problem that can be eliminated by using a different geometry, Fig. 1.3b, however methods for fabricating such a structure have so far resulted in low quality channel material[26,27].

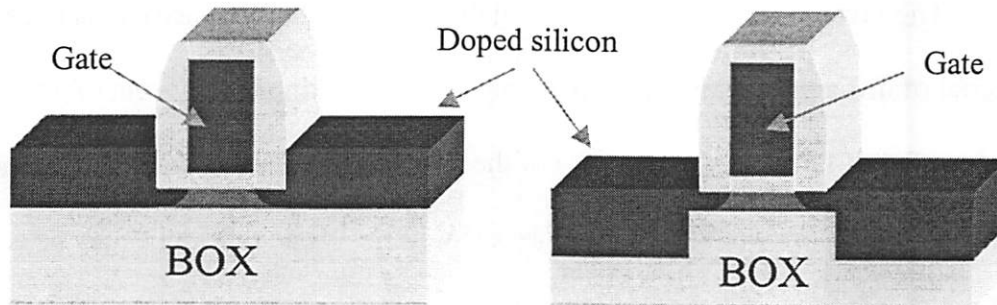


Figure 1.3a: An elevated source/drain thin-body structure. Doped silicon regions shown in dark gray, undoped silicon shown in light gray.

Figure 1.3b: A solid phase epitaxy thin-body structure. Body region is deposited as amorphous and crystallized using source/drain as seeds.

Method 2 relies on changing the source/drain material to a metal. The simplest structure that implements this solution is shown in Fig. 1.4. Unfortunately, this device design, when implemented with a traditional silicide, results in high contact resistance at the silicon/metal interface even for highly doped source drain extensions.

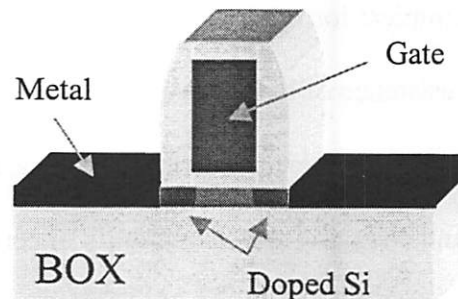


Figure 1.4: A simple metal or silicide source/drain device. Black regions are metallic, either elemental or silicide.

The resistance problem of silicided thin-body devices can be seen by simply extrapolating the contact resistance(R_c) from the state of the art specific contact resistance(ρ_c). Degenerately doped NMOS devices with $TiSi_2$, or $CoSi_2$ silicide source/drain have a ρ_c of as low as $10\Omega\text{-}\mu\text{m}^2$ [28]. Assuming a body thickness of $0.01\mu\text{m}$ gives a R_c of $1\text{k}\Omega\text{-}\mu\text{m}$, this resistance is larger than channel resistance of $\sim 700\Omega\text{-}\mu\text{m}$, and will significantly reduce the device performance.

Since the contract resistance is a strong function of silicide-to-silicon barrier height reducing the barrier produces a large reduction in contact resistance. Unfortunately reducing the barrier for electrons increases the barrier for holes and vice-versa making it impossible to reduce the barrier for electrons and holes with one silicide. The obvious solution is to use two complementary silicides(CS), one for NMOS and another one for PMOS, each having a low barrier to its respective carrier. The investigation of the viability of such an approach is the subject of this work.

References:

- [1] SIA roadmap 1999; http://public.itrs.net/files/1999_SIA_Roadmap/FEP.pdf
- [2] W-C. Lee, and C. Hu, "Modeling gate and substrate currents due to conduction and valance band electron and hole tunneling," Proceeding of the 2000 symposium on VLSI technology, p.98-99, June 13-15 2000
- [3] Bin Yu, H. Wang, C. Riccobene, Q. Xiang and M.R. Lin, "Limits of gate-oxide scaling in nano-transistors," Proceeding of the 2000 symposium on VLSI technology, p. 90-91, June 13-15 2000
- [4] G. Timp, K.K. Bourdelle, J.E. Bower, F.H. Baumann, T. Boone, R. Cirelli, K. Evans-Lutterodt, J. Garno, A. Ghetti, H. Gossman, M. Green, D. Jacobson, Y. Kim, R. Kleiman, F. Klemens, A. Kornblit, C. Lochstampfor, W. Mansfield, S. Moccio, D.A. Muller, L. E. Ocola, M.L. O'Malley, J. Rosamilia, J. Sapjeta, P. Silverman, T. Sorsch, D.M. Tennant, W. Timp, B.E. Weir, "Progress toward 10nm CMOS devices" Proceedings of the International Electron Device Meeting 1998, p. 615-618, Dec 6-9 1998
- [5] T. Ghani, K. Mistry, P. Packan, S. Thompson, M. Stettler, S. Tyagi, and M. Bohr, "Scaling challenges and device design requirements for high performance sub-50nm gate length planar CMOS transistors," Proceeding of the 2000 symposium on VLSI technology, p. 174-175, June 13-15 2000
- [6] Robert Chau, Jack Kavalieros, Brian Roberds, Richard Schenker, Dan Lionberger, Douglas Barlage, Brian Doyle, Reza Arghavani, Anand Murthy, and Dilbert Dewey, "30nm physical gate length CMOS transistors with 1.0ps n-MOS abd 1.7ps p-MOS gate delays," Proceedings of the International Electron Device Meeting 2000, pp. 45-48, Dec 10-13 2000
- [7] Shahidi, G.C. "Mainstreaming of the SOI technology," Proceedings of the 29th European Solid-State Device Research Conference, p.3-10, 13-15 Sept. 1999
- [8] P. Smeys, V. McGahay, I. Yang, J. Adkisson, K. Beyer, O. Bula, Z. Chen, B. Chu, J. Culp, S. Das, A. Eckert, L. Hadel, M. Hargrove, J. Herman, L. Lin, R. Mann, E. Maciejewski, S. Narasimha, P. O'Neil, S. Rauch, D. Ryan, J. Toomey, L. Tsou, P. Warekamp, R. Wachnik, T. Wagner, S. Wu, C. Yu, P. Agnello, J. Connolly, S. Crowder, C. Davis, R. Ferguson, A. Sekiguchi, L. Su, R. Goldblatt, and T.C. Chen, "A high performance 0.13um SOI CMOS technology with Cu interconnects and low-k DEOL dielectric," Proceeding of the 2000 symposium on VLSI technology, p. 184-185, June 13-15 2000
- [9] Antoniadis, D.A.; Wei, A.; Lochtefeld, A. "SOI devices and technology," Proceedings of the 29th European Solid-State Device Research Conference, p.81-7, 13-15 Sept. 1999

- [10] S. Lloyd, "Quantum-mechanical computers," *Scientific American (International Edition)*, vol.273, (no.4), p.44-50, Oct. 1995
- [11] David P. DiVincenzo, "Prospects for quantum computing," *Proceedings of the International Electron Device Meeting 2000*, pp. 12-15, Dec 10-13 2000
- [12] P.L. McEuen, "Single-wall carbon nanotubes," *Physics World*, vol.13, (no.6), p.31-36, June 2000
- [13] Fuhrer, M.S.; Nygard, J.; Shih, L.; Forero, M.; Young-Gui Yoon; Mazzoni, M.S.C.; Hyoungh Joon Choi; Jisoon Ihm; Louie, S.G.; Zettl, A.; McEuen, P.L. Crossed nanotube junctions. *Science*, vol.288, (no.5465), p.494-7, 21 April 2000
- [14] S. K. H. Fung, N. Zamdmer, P. J. Oldiges, J. Sleight, A. Mocuta, M. Sherony, S-H. Lo, R. Joshi, C. T. Chuang, I. Yang, S. Crowder, T.C. Chen, F. Assaderaghi, and G. Shahidi, "Controlling floating-body effects for 0.13um and 0.10um SOI CMOS," *Proceedings of the International Electron Device Meeting 2000*, pp. 231-234, Dec 10-13 2000
- [15] Bin Yu, Yeh-Jiun Tung, Stephen Tang, Elliot Hui, Tsu-Jae King, and Chenming Hu, "Ultra-thin-body silicon on insulator MOSFETs for terabit-scale integration," *Proceedings of the International Semiconductor Device Research Society (ISDRS) 1997*, pp. 623-628, 1997
- [16] Digh Hisamoto, Tsuyosi Kachi, Shinpei Tsujikawa, Akihiro Miyauchi, Kikuo Kusukawa, Noriyuki Sakuma, Yoshio Homma, Natsuki Yokoyama, Fumio Ootsuka, and Takahiro Onai, "A compact FD-SOI MOSFET fabrication process featuring SixGe1-x gate and damascene-dummy SAC," *Proceeding of the 2000 symposium on VLSI technology*, p. 208-209, June 13-15 2000
- [17] Jakub Kedzierski, Peiqi Xuan, Erik H. Anderson, Jeffrey Bokor, Tsu-Jae King, and Chenming Hu, "Complementary silicide source/drain thin-body MOSFETs for the 20nm gate length regime," *Proceedings of the International Electron Device Meeting 2000*, pp. 57-60, Dec 10-13 2000
- [18] X. Huang, W-C. Lee, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, Y-K. Choi, V. Subramanian T-J. King, J.Bokor, C. Hu, "Sub 50-nm FinFET: PMOS," *Proceedings of the International Electron Device Meeting 1999*, pp. 67-70, Dec 1999
- [19] H.-S. Wong, David J. Frank, and Paul M. Solomon, "Device design considerations for double-gate, ground plane, and single-gated ultra-thin SOI MOSFETs at the 25nm channel length generation," *Proceedings of the International Electron Device Meeting 1998*, pp. 407-410, Dec 6-9 1998
- [20] H. Kawaura, "Carrier Transport in Ultrasmall MOSFETs," *Superlattices and Microstructures*, Dec 2000 (To be published)

- [21] A. Sabnis and J. Clemens, "Characterization of electron mobility in inverted <100> silicon surface," Proceedings of the International Electron Device Meeting 1979, pp.18-21
- [22] G. Timp, J. Bude, F. Baumann, K.K. Bourdelle, J. Garno, A. Ghetti, M. Green, H. Gossmann, Y. Kim, R. Kleiman, A. Kornblit, F. Klemens, S. Moccio, D.A. Muller, P. Silverman, T. Sorsch, W. Timp, D.M. Tennant, R. Tung, and B.E. Weir, "The Ballistic Nanotransistor and the Scale of Complexity in an Integrated Circuit," Superlattices and Microstructures, Dec 2000 (To be published)
- [23] Leland Chang, Stephen Tang, Tsu-Jae King, Jeffrey Bokor, and Chenming Hu, "Gate length scaling and threshold voltage control of double-gate MOSFETs," Proceedings of the International Electron Device Meeting 2000, pp. 719-722, Dec 10-13 2000
- [24] R. Gwoziecki, M. Jurczak, T. Skotnicki, J.-L. Regolini, M. Paoli, "Suitability of elevated source/drain for deep submicron CMOS," Proceedings of the 29th European Solid-State Device Research Conference, p.384-7, 13-15 Sept. 1999
- [25] Gerold W. Neudeck, Tai-Chi Su, and Jack P. Denton, "Novel silicon epitaxy for advanced MOSFET devices," Proceedings of the International Electron Device Meeting 2000, pp. 169-172, Dec 10-13 2000
- [26] V. Subramanian, J. Kedzierski, N. Lindert, Y. Su, J. McHale, K. Cao, Tsu-Jae King, Jeffrey Bokor, and Chenming Hu, "A bulk Si compatible ultra thin body SOI technology for sub-100nm MOSFETs," DRC 1999, p28-29
- [27] Peiqi Xuan, Jakub Kedzierski, Vivek Subramanian, Jeffrey Bokor, Tsu-Jae King, Chenming Hu, "60nm Planarized Ultra-thin Body Solid Phase Epitaxy MOSFETs," Proceedings of the 58th Device Research Conference 2000, p. 67
- [28] Osamu Nakatsuka, Tetsuo Ashizawa, Hirotaka Iwano, Shigeaki Zaima, and Yuki Yasuda, "Contact resistivities and electrical characteristics of Co/Si contact by rapid thermal annealing," Advanced Metallization Conference 1998, (Material Research Society, 1999), pp. 605-611

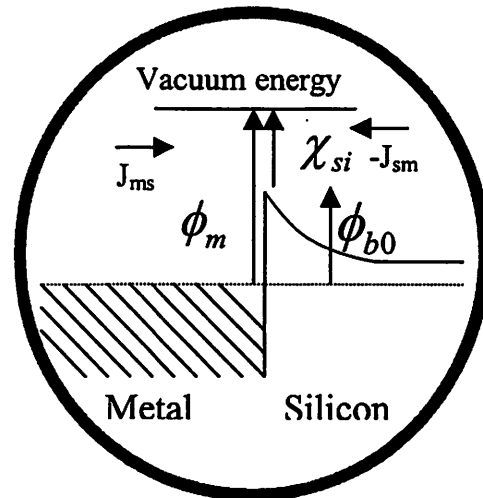
2.0 Background / Literature

This section will present the theoretical and experimental background material needed to understand the issues relevant to thin-body device silicidation.

2.1 Metal-semiconductor junctions

The theory of metal-semiconductor junctions is of special importance when considering the source/drain engineering of the silicide thin-body device, since the contact resistance plays a dominant role in determining the overall device performance. The simplest band diagram of a metal-semiconductor junction is shown in Fig. 2.1. In this low field model, the Schottky barrier height is correlated to the difference between the workfunctions of the metal and silicon, and is dependent only on those two material properties. The current across the junction is just the sum of the two thermionic emission currents $J=J_{sm}+J_{ms}$ [1].

Figure 2.1: Energy band diagram for a simple model of a low-field Schottky barrier. The barrier height $\Phi_{b0} = (\Phi_m - \chi_{si})$.



In a real junction some intrinsic interface charge is always present and it tends to move the barrier height to the mid-gap of the semiconductor[2]. The impact of this effect is to pin the barrier height around 0.6eV for $\Phi_m=4.5\text{eV}$, and reduce the influence of Φ_m on Φ_{b0} to $\Phi_{b0}=0.3(\Phi_m - \chi_{si})$, where Φ_{b0} is the flat-band barrier height, χ_{si} is the electron affinity of

silicon, and Φ_m is the metal workfunction. As the electric field perpendicular to the interface increases secondary effects start to become significant. Carrier tunneling across the junction and barrier lowering due to image charge induction are the two most important effects. Image charge induction causes the effective barrier to be lowered by high electric fields, increasing the thermionic current[3]. High E-fields also make the barrier more narrow increasing the tunneling current through it.

It is common to divide metal-semiconductor barriers into two categories. If the interface fields are low the metal semiconductor junction functions like a diode, since the Schottky barrier blocks current flow in reverse bias. Such a metal-semiconductor junction is called a Schottky junction. If the fields are high the tunneling probability is large and the junction behaves more like a resistor, and is called an Ohmic junction. Usually in this regime a linearization of the transport mechanisms is made to extract the equivalent specific contact resistance, R_c [1].

2.1.1 Conventional metal-semiconductor junction resistance model

Accurately modeling the specific contact resistance for a highly doped semiconductor-metal interface is important when designing transistor contacts. As mentioned in the previous section tunneling is a dominant transport mechanisms through an Ohmic contact. A quantitative relation for the specific contact resistance can be obtained by using the WKB approximation to calculate tunneling probability. Then the tunneling current can be calculated by integrating the probability over the appropriate carrier distributions. Linearizing the resulting relation gives the tunneling contact resistance as [4]:

$$R_C \propto \exp \left[\frac{2\sqrt{\epsilon_s m^*}}{\hbar} \frac{\Phi_{b0}}{\sqrt{N_{doping}}} \right]$$

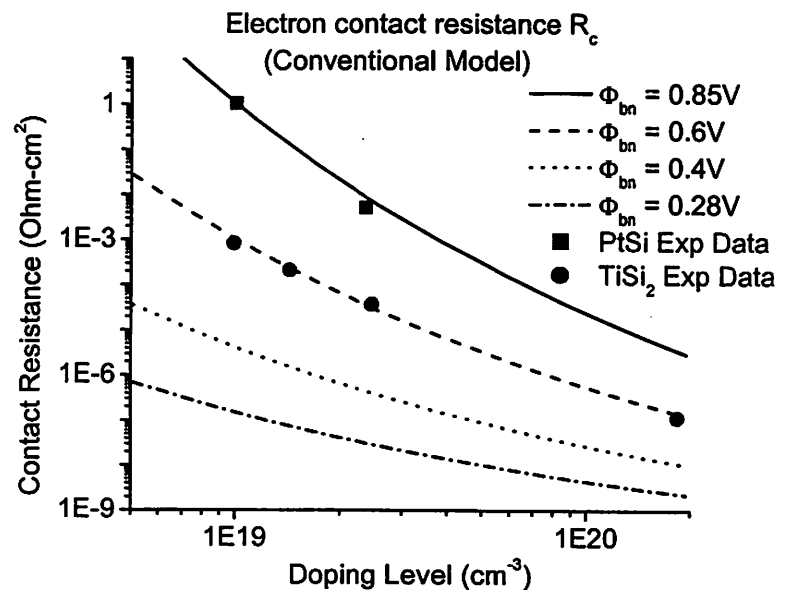
A similar relation can be calculated assuming only thermionic emission transport over the barrier:

$$R_c = \frac{k}{qA^*T} e^{\frac{q\Phi_{bc}}{kT}}$$

Since in the actual junction both thermionic emission and tunneling transport mechanisms occur simultaneously the resistances can be added in parallel to give the actual contact resistance. In relevant design space for ohmic contacts to silicon the tunneling resistance is generally significantly smaller than the thermionic resistance, and therefore dominates current transport.

Several significant details in this model are left to the fitting parameters. One is the variation of the m^* on the doping concentration in the silicon. The second is the influence of the induced charge barrier lowering on the tunneling barrier profile. The omission of image charge barrier lowering is not significant if the barrier is large, and other fitting parameters absorb the effect. The conventional theory makes just such a simplification. Fig. 2.2 is a plot of the conventional model after fitting to data for TiSi_2 and PtSi electron barriers[1].

Figure 2.2: Specific contact resistance according to the conventional model. Model parameters fit to the experimental data for TiSi_2 and PtSi.



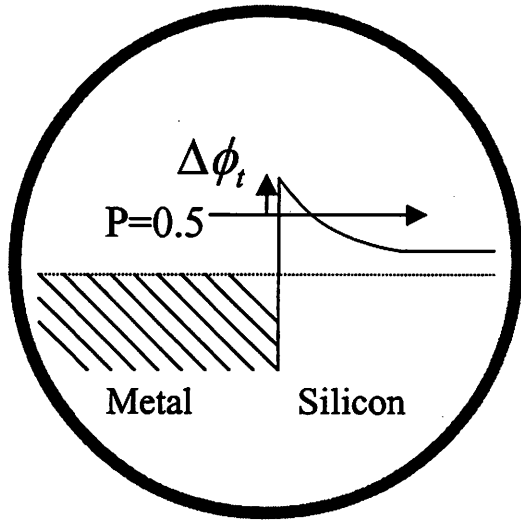
The doping is extrapolated to $2E20\text{cm}^{-3}$, which is close to the electrically active solubility limit. The lowest contact resistance that can be obtained with a 0.6V barrier silicide is around $1E-7 \Omega\text{-cm}^2$, in agreement with experimental results[5]. As expected lower barriers lead to lower contact resistances. A barrier height of 0.28eV would make it possible to lower the resistance to $2E-9 \Omega\text{-cm}^2$, an almost two orders of magnitude improvement.

A disadvantage of the conventional model is that it seeks to absorb an important effect, the barrier lowering due to induced image charge, into the fitting parameters. It is not clear how accurate this approach is especially for low-barrier metals where barrier lowering can be a significant portion of the total barrier height. An alternative model for the current in a Schottky barrier is to translate the tunneling into an effective barrier lowering, combine it with image force barrier lowering and then treat all current through the barrier as thermionic.

2.1.2 Equivalent barrier lowering model

This section will present an original method to quantitatively combine the tunneling and image charge barrier lowering effects in one model. The equivalent barrier-lowering model seeks to compensate for the lack of tunneling in the thermionically emitted current by lowering the barrier by an appropriate amount. The advantage of this approach is that it is simple to combine barrier lowering that results from image charge induction with the equivalent barrier-lowering that results from tunneling. The equivalent barrier-lowering model is also easy to incorporate in a ballistic transport model.

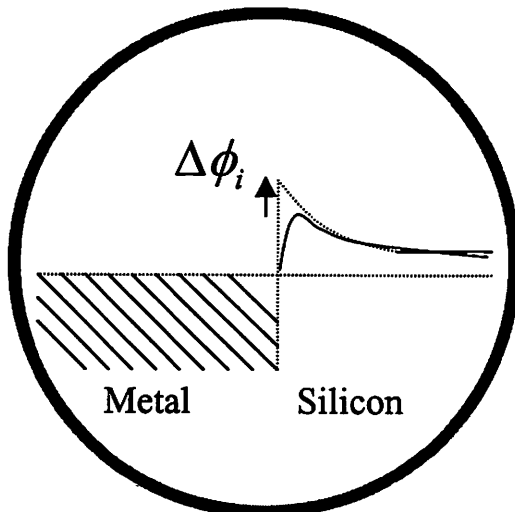
A rough approximation can be made to convert the tunneling effect into a effective barrier lowering mechanism by calculating the energy at which the transmission coefficient is equal to 0.5 and lowering the barrier to that energy. Since the transmission probability is an exponential function of barrier height and width the implied assumption of this approximation, that for all energies where $T < 0.5$ $T=0$, and where $T > 0.5$ $T=1$. Fig. 2.3 shows the relevant potentials and the effective barrier lowering formula.



$$\Delta\phi_t = \frac{1}{e} \left(\frac{3e\hbar(\ln 2)}{4\sqrt{2}m^*} \right)^{2/3} (E_y)^{2/3}$$

Figure 2.3: Band diagram of equivalent barrier lowering to carrier tunneling through the Schottky barrier. Relation above gives this lowering in terms of the electric field E_y at the metal interface.

Image charge induction lowers the apparent barrier to electrons by taking into consideration the effect of the charge induced in the metal by the electron moving across the Schottky barrier. Image charge induction effectively reduces the barrier height, as shown in Fig. 2.4[6].



$$\Delta\phi_i = \left(\frac{e}{4\pi\epsilon_s} \right)^{1/2} (E_y)^{1/2}$$

Figure 2.4: Band diagram of equivalent barrier lowering due to image charge induction. Relation above gives this lowering in terms of the electric field E_y at the metal interface.

To find the effective barrier height, the two barrier lowering terms are subtracted from the low field barrier Φ_{b0} :

$$\Phi_b = \Phi_{b0} - \Delta\phi_i - \Delta\phi_t$$

The details of the justification for the simple equation above are theoretically interesting, in that it is not immediately obvious if an electron that is in the process of tunneling induces the same barrier lowering as an electron being thermionically emitted over the barrier. This is because the magnitude of the wavefunction of a tunneling electron is significantly less than one, in the steady state solution. If a tunneling electron induced a lower image charge than a thermionic electron it would be improper to add the two barrier lowering mechanisms. However such reasoning assumes that the decoupled picture, of the electron traveling in a fixed Hamiltonian, applies. Such a picture is not applicable even for thermionically emitted electrons, since it would imply that the physical width of the electron wavepacket has an influence on the image barrier lowering magnitude. So just as the coupled solution corrections produce a full barrier lowering in thermionically emitted electrons, I expect that a coupled treatment will produce a correction of the same magnitude to tunneling electrons. This is because the position operator of the tunneling and thermionically emitted electron becomes an operator in the Hamiltonian of the induced charge with the same magnitude[21].

The fact that barrier lowering is just as effective for tunneling electrons as it is for thermionically emitted electrons is a justification of the addition of the equivalent barrier-lowering for tunneling and induced charge components to obtain the total equivalent barrier lowering. Image charge lowering does however change the shape of the barrier,

making it more quadratic than triangular at the top, in general such a barrier will be slightly more difficult to tunnel through than a triangular barrier.

2.1.3 The flat-band barrier height definition

The actual low-field Schottky barrier height, Φ_{b0} , (Φ flat-band) is determined largely by the charge at the metal-semiconductor interface. The presence of this charge effectively pins the barrier closer to midgap. Taking into consideration the presence of these states an empirical formula can be derived, and fit to the available data:

$$\Phi_{b0} = C_1(\Phi_m) - C_2 \quad \text{Experimentally for silicon } C_1 = 0.27, C_2 = 0.55 \text{ eV [7]}$$

The available Φ_{b0} data is shown in Fig. 2.5

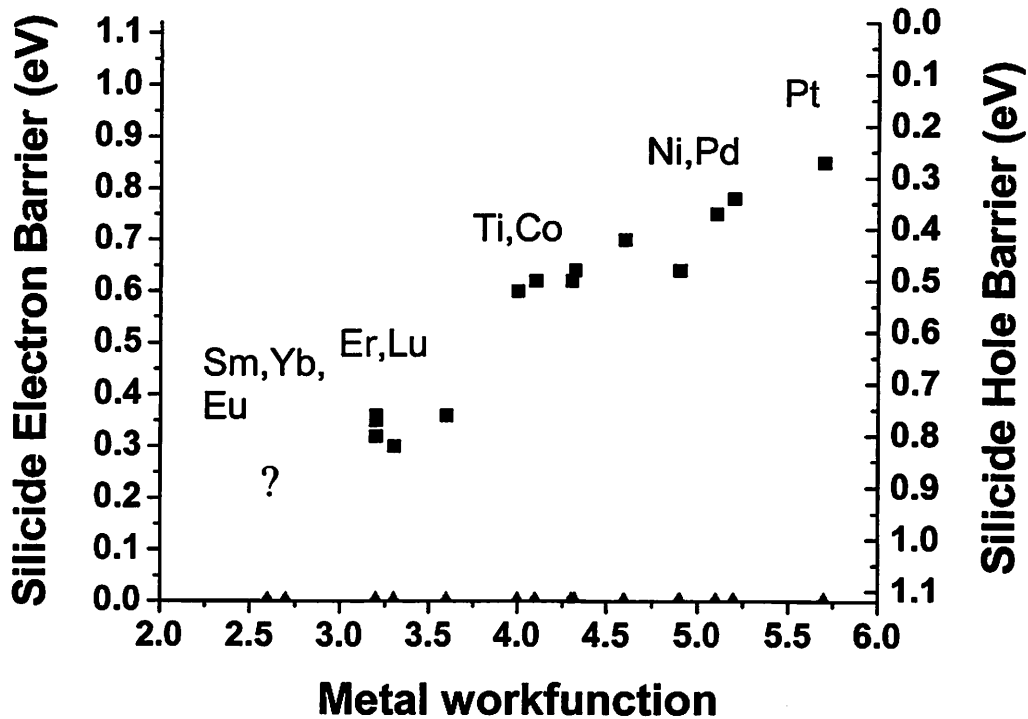


Figure 2.5: The silicide flat-band electron barrier height, Φ_{bn0} , as a function of metal work-function. Metal work-functions in red, corresponding silicide Φ_{bn0} in black.

Since the workfunction (Φ_m) for metals varies from Europium $\Phi_m=2.5\text{eV}$, to Platinum $\Phi_m=5.5\text{eV}$ according to this empirical theory it should be possible to obtain barriers between 0.13eV and 0.93eV to n-type silicon. Platinum silicide barrier height is fairly close to the expected value at 0.88eV[8], and the Europium silicide barrier height has not yet been measured. As can be seen in the figure there is a large variation from the empirical linear fit between workfunction and barrier height. This variation is likely caused by the exact nature of the interface between the silicide and the silicon. Imperfections such as trap states and grain boundaries in either silicon or silicide may cause such variation, for example epitaxial NiSi has a barrier height to electrons of 0.78eV, exactly what would be expected from the empirical relation, however poly-NiSi has a barrier height of 0.65eV[9]. In general the introduction of extra trap states tends to pin the barrier in the middle of the silicon band-gap.

2.2 Synopsis of work relevant to the quantitative understanding of thin-body source/drain engineering requirements

In order to better understand the relevant issues and solutions in siliciding the thin-body transistor it is useful to review the source/drain engineering studies used for traditional bulk transistors, as well as other efforts of thin-body source/drain engineering. Silicide has been used to reduce the source/drain and gate resistance in bulk transistors for several technology generations. Cobalt and Titanium disilicides are typically used due to their low resistance and their symmetric barrier height to both electrons and holes[10]. They are fabricated on highly doped silicon regions to limit contact resistance. A standard source/drain design is shown in Fig. 2.6 with all the significant sources of series resistance listed.

Bulk design

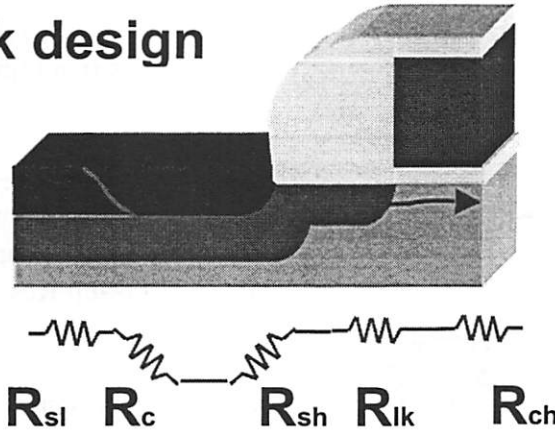


Figure 2.6: Series resistance components in a bulk transistor:

Parasitic:

- R_{sl} Silicide Sheet Res.
- R_c Specific Contact Res.
- R_{sh} Silicon Sheet Res.
- R_{lk} Link-up Res.

Reference:

- R_{ch} Equivalent Channel Res.

The sheet resistance of silicides in 0.18 μm technology is typically on the order of 10-20 Ω/sq [11], corresponding to a silicide thickness of 250-350 \AA . This resistance is much lower than is required for the connection between the source and the metal via, low enough to use the silicide as an interconnect for short distances. The metal-semiconductor contact resistance is also small for bulk devices, even down to 30nm gate lengths $R_c \sim 40\Omega\text{-}\mu\text{m}$ [5]. This is because the contact area is large, nominally covering the entire source/drain region. Current spreading effects and resistance in the silicon limit the useful contact area to a region around the gate, however if silicon resistance is low this conduction region is still large. Other parasitic resistances are also not a problem; by engineering the correct profile of dopants in the source/drain R_{sh} can be kept in the 20 $\Omega\text{-}\mu\text{m}$ range, and the R_{lk} for an abrupt doping gradient of 8nm/dec is $\sim 40 \Omega\text{-}\mu\text{m}$. All these resistances add up to less than 120 $\Omega\text{-}\mu\text{m}$ and are therefore smaller than the equivalent channel resistance of $\sim 700 \Omega\text{-}\mu\text{m}$ (NMOS per gate). Therefore bulk devices are expected to lose only $\sim 17\%$ of current drive capability to series resistance problems. The same is not true for the thin-body device, if it were to use a similar source/drain metalization technology, as in Fig. 2.7.

Thin body

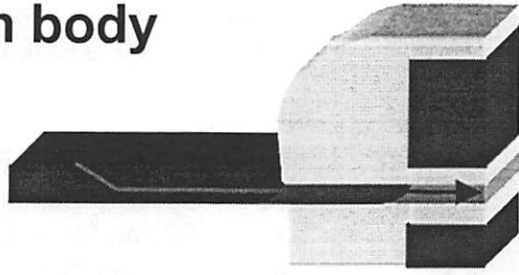


Figure 2.7: Series resistance components in a thin body transistor:

Parasitic:

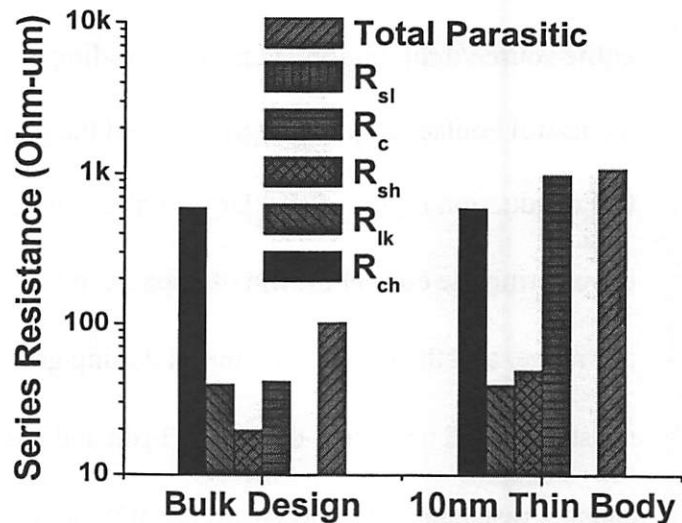
- R_{sl} Silicide Sheet Res.
- R_c Specific Contact Res.
- R_{sh} Silicon Sheet Res.
- R_{lk} Link-up Res.

Reference:

- R_{ch} Equivalent Channel Res.

Fig. 2.8 shows the parasitic resistances for a bulk transistor[5], and similar values for a fully silicided thin body structure, shown in Fig. 2.7, with a body thickness of 10nm. The figure also shows the typical equivalent channel resistance for a short channel device with one gate. As long as the parasitic resistances add up to a value significantly less than the equivalent channel resistance the parasitics will not degrade transistor behavior.

Figure 2.8: Comparison of series resistance components for thin-body and bulk designs. The thickness of the thin body is 10nm. Channel resistance is given for a single gate, for a double gate device R_{ch} would be half of the value shown. These resistances are calculated for an NMOS device.



As can be seen from Fig. 2.8 parasitic resistance for thin-body devices is significantly larger than bulk transistors. The largest component of the difference is the contact

resistance, a direct result of the smaller contact area in the thin-body design. This area difference is significant, even for devices of the same gate-length. Therefore a thin-body device with a source/drain silicided by a mid-gap silicide will have a contact resistance of $1\text{k}\Omega\text{-}\mu\text{m}$ per contact assuming a specific contact resistance of $10\Omega/\mu\text{m}^2$. This value is larger than the equivalent channel resistance and therefore the thin-body design will be dominated by parasitic resistance, degrading the on-current significantly.

Experimental data confirms the extrapolation of these resistance values. Studies of relatively large fully-depleted SOI structures with silicon thickness of 50nm and a gate length of 250nm have found serious series resistance issues[12]. Experimentally measured specific contact resistance of these structures was found to be $\sim 30\Omega\text{-}\mu\text{m}^2$. This value corresponds to a series resistance of $300\Omega\text{-}\mu\text{m}$ for a body thickness of 50nm, a resistance that already reduces the transconductance of the NMOS transistor by 15-20%. If the same contact technology is extrapolated to a thin-body geometry with silicon thickness of only 5nm the series resistance is $3\text{k}\Omega\text{-}\mu\text{m}$, reducing the performance of the transistor by as much as 80%. So there appears to be a source/drain thickness limit, of approximately 50nm (per gate), below which traditional source/drain silicidation does not yield sufficiently low series resistance.

As discussed in the introduction, researchers have attempted to solve the contact resistance issues of thin-body designs by growing the thin source/drain regions to a 50nm thickness where mid-gap silicidation is still viable. The regrowth technique can be a selective silicon epitaxy[13] or selective germanium deposition[14]. However all selective deposition techniques developed thus far have a high defect density. This is because once a selective deposition nucleates on a defect the nucleated crystal will grow

at the same rate as the desired film. This process has the ability to magnify atomic scale defects or chance nucleations to silicon islands up to 50nm in radius. VLSI yield requirements have very stringent defect density limits, it has not been demonstrated that selective regrowth techniques can be made to reach these limits.

2.3 Synopsis of previous silicide source/drain device research

In a radical departure from traditional source/drain engineering two independent groups developed the idea of fabricating the source/drain regions of a transistor from a low-barrier silicide without the use of any doping[15,16]. This structure, shown in Fig. 2.9, consists of a bulk design in which the doped silicon regions are replaced with a low barrier silicide. The advantage of this structure is that it has an atomically sharp junction, eliminating the link-up resistance, as well as numerous difficulties associated with fabricating shallow highly doped junctions.

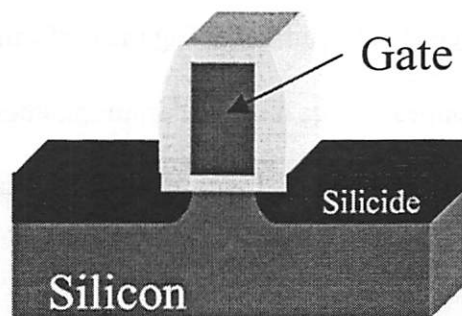
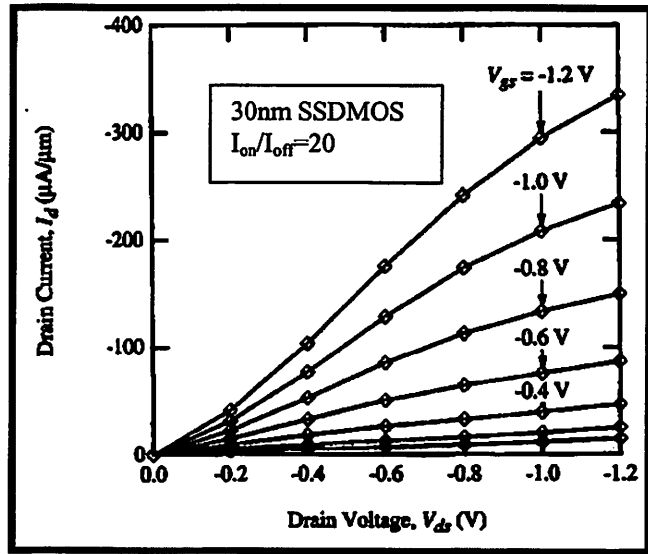


Figure 2.9: Cross section of a bulk silicide source drain structure. Body is undoped in the standard design, although some variants add a uniform doping to control leakage.

So far only bulk PMOS silicide source/drain (SSD) structures have been demonstrated[17], with source/drains made from PtSi ($\Phi_{bp0}=0.24\text{eV}$). Fig. 2.10 shows the turn on characteristics of a PMOS bulk SSD device, scaled to 30nm gate length.

Figure 2.10: I_d - V_d plot of a bulk PtSi PMOS SSD device. I_{on} is similar to traditional bulk transistors. However I_{on}/I_{off} ratio is several orders of magnitude smaller than expected in a traditional bulk device. Figure reproduced from [17].



This device shows an on-current level comparable to that of a traditional bulk pMOSFET. The off-characteristics however are very different, in the traditional device $I_{on}/I_{off}=10E5$ while in the SSD device $I_{on}/I_{off}=20$. The high leakage of the SSD bulk device is caused by thermal emission over the low-barrier, and travels through a similar area as the punch-through current would in a normal transistor. Increasing substrate doping to limit this current also decreases the ability of the gate to lower the barrier at the source, decreasing transistor on-current.

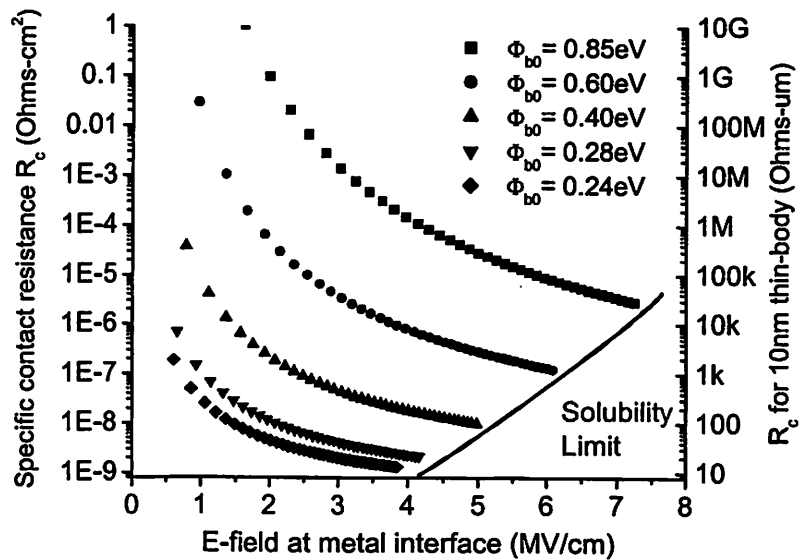
Other groups have fabricated silicide source/drain devices on SOI[18,19,20], in order to decrease the leakage current. PMOS devices that show good short channel characteristics have been fabricated down to 77nm gate length, using a 350Å thick silicon layer[18]. NMOS silicide source/drain devices on SOI using $ErSi_2$ as the silicide material have been attempted. However devices fabricated thus far show a high electron barrier due to contamination of the silicide silicon barrier[19]. The SOI silicide source/drain devices show considerably better turn off characteristics, because they are in effect thin-

body devices and eliminate punch through leakage paths. An I_{on}/I_{off} of $1E3$ has been demonstrated for long channel PMOS devices.

2.3.1 Design space for the silicide source/drain device

The barrier height design space for the undoped silicide source/drain device is well demonstrated by considering the electric field at the source-body interface. This field is responsible for lowering the barrier, and thus is critical in determining the current drive of the transistor. Maximizing the electric field leads to lower contact resistance and better on-current. Fig. 2.11 shows the specific contact resistance as a function of the electric field at the silicide interface, as predicted by the conventional model. The figure also shows, on the right axis, the corresponding contact resistance of the device, assuming the area contributing to the current flow along the silicide boundary is 10nm deep in the direction perpendicular to the gate oxide.

Figure 2.11: Contact resistance plot as a function of e-field at the metal interface. Left axis shows specific contact resistance, right axis shows R_c for a 10nm thin-body SOI silicide source/drain device.



Considering that the channel resistance is $700\Omega\text{-}\mu\text{m}$, and noting that bulk PMOS SSD devices show the same on-current as their traditional counterparts, it can be concluded

that the contact resistance is below $300\Omega\text{-}\mu\text{m}$. Assuming the barrier height of 0.24eV for PtSi, Fig. 2.11 indicates that such a resistance would correspond to an electric-field strength of 1MV/cm or greater. When designing the SSD device it is important to remember that it is the structure geometry that determines the field strength given the bias, and that the resistance increases exponentially with decreasing field strength or increasing barrier height.

References:

- [1] C. Y. Chang, Y. K. Fang, and S. M. Sze, "Specific contact resistance of metal-semiconductor barriers," *Solid-State Electronics*, Vol. 14, pp. 541-550, 1971
- [2] S. G. Louie and M. L. Cohen, "Electronic structure of a metal-semiconductor interface," *Physical Review B*, Vol. 13, pp. 2461, 1976
- [3] S. M. Sze, C.R. Crowell, and D. Kahng, "Photoelectric determination of the image force dielectric constant for electrons in Schottky barriers," *Journal of Applied Physics*, Vol. 35, p. 2534, 1964
- [4] D. A. Neamen, "Semiconductor physics and devices," Boston MA: Irwin, 1992, p. 349
- [5] Keys, P.; Gossmann, H.-J.; Ng, K.K.; Rafferty, C.S. "Series resistance limits for 0.05 μ m MOSFETs," *Superlattices and Microstructures*, vol.27, p.125-36. 26-27 Aug. 1999
- [6] S. M. Sze, "Physics of semiconductor devices," 2nd ed. New York: Wiley, 1981, pp. 250-254
- [7] S. M. Sze, "Physics of semiconductor devices," 2nd ed. New York: Wiley, 1981, pp. 274-275
- [8] G. Ottaviani, K. N. Tu, J. W. Mayer, "Barrier heights and silicide formation for Ni, Pd, and Pt on silicon," *Physical Review B*, Vol. 34 (no.6), p.3354, 15 Sept. 1981
- [9] M. Liehr, P. E. Schmid, F. K. Le Gones, P.S. Ho, "Correlation of Schottky-barrier height and microstructure in the epitaxial Ni silicide on Si(111)," *Physical Review Letters*, Vol. 54 (no.19), p. 2139-2142, 13 May 1985
- [10] S. Wolf, "Silicon Processing for the VLSI era, volume 2 – process integration," Sunset Beach: Lattice Press, 1990, pp. 144-152
- [11] Kouji Matsuo, Tomohiro Saito, Atsushi Yagishita, Toshihiko Inuma, Atsushi Murakoshi, Kazuaki Nakajima, Seiichi Omoto, Kyoichi Suguro, "Damascene metal gate MOSFETs with Co silicided source/drain and high-k gate dielectrics," 2000 Symposium on VLSI Technology, pp. 70-71, June 13-15 2000
- [12] Harvey I. Liu, James A. Burns, Craig L. Keast, and Peter W. Wyatt, "Thin silicide development for fully-depleted SOI CMOS technology," *IEEE Transactions on electron devices*, Vol. 45, no. 5, pp. 1099-1104, May 1998

- [13] Srivastava, A.; Sun, J.; Bellur, K.; Bartholomew, R.F.; O'Neil, P.; Celik, S.M.; Osburn, C.M.; Masnari, N.A.; Ozturk, M.C.; Westhoff, R.; Fowler, B. (Edited by: Massoud, H.Z.; Iwai, H.; Claeys, C.; Fair, R.B.) "A 0.18 μm CMOS technology for elevated source/drain MOSFETs using selective silicon epitaxy," Proceedings of the Sixth International Symposium on Ultralarge Scale Integration Science and Technology. ULSI Science and Technology 1997, p.571-85, 1997
- [14] Yang-Kyu Choi; Yoo-Chan Jeon; Ranade, P.; Takenuchi, H.; Tsu-Jae King; Bokor, J.; Chenming Hu, "30 nm ultra-thin-body SOI MOSFET with selectively deposited Ge raised S/D," 58th DRC. Device Research Conference, p.23-4, 19-21 June 2000
- [15] J. R. Tucker, "Schottky barrier MOSFETs for silicon nanoelectronics," Advanced Workshop on Frontiers in Elec. 1997 Proceedings, p. 97-100, 1997
- [16] Snyder, J.P.; Helms, C.R.; Nishi, Y. "Experimental investigation of a PtSi source and drain field emission transistor," Applied Physics Letters, vol.67, (no.10), p.1420-24, 4 Sept. 1995
- [17] C. Wang, John P. Snyder, and J. R. Tucker, "Sub-40 PtSi Schottky source/drain metal-oxide-semiconductor field-effect transistors," Applied Physics Letters, Vol. 74, Num. 8, pp. 1174-1176, 22 Feb. 1999
- [18] Watary Saitoh, Shigeharu Yamagami, Atsushi Itoh, and Masahiro Asada, "35nm metal gate p-type metal oxide semiconductor field-effect transistor with PtSi Schottky source/drain on separation by implanted oxygen substrate, Japanese Journal of Applied Physics, Vol. 38, Part 2, No. 6A/B, pp. 629-631, 1999
- [19] Wataru Soitoh, Atsushi Itoh, Shigeharu Yamagami, and Masahiro Asada, "Analysis of short-channel Schottky source/drain metal-oxide-semiconductor field-effect transistor on silicon-on-insulator substrate and demonstration of sub-50-nm n-type devices with metal gate, Japanese Journal of Applied Physics, Vol. 38, Part 1, No. 11, pp. 6226-6231, November 1999
- [20] Atsuchi Itoh, Masayuki Saitoh, and Masahiro Asada, "A 24nm-long channel metal-gate p-type Schottky source/drain metal-oxide-semiconductor field effect transistor on separation-by-implanted-oxygen substrate," Japanese Journal of Applied Physics, Vol. 39 Part 1, No. 8, pp. 4757-4758, August 2000
- [21] van Schilfgaarde, M. "Scattering from ionized dopants in Schottky barriers," Journal of Vacuum Science & Technology B, vol.8 (no. 4), p.990-4, July-Aug. 1990

3.0 Thin-body complementary silicide source/drain devices

As outlined in the previous chapters, thin-body devices show superior short-channel effects to bulk transistors but suffer from serious series resistance and threshold voltage control problems. The most significant component the series resistance of a silicided thin-body structure is the contact resistance between the silicide and the silicon extension region, see section 2.3.

A novel way to reduce this resistance is to use a silicide with a barrier that is lower than the mid-gap value of $\sim 0.6\text{eV}$. In CMOS this implies the use of two silicides since the electron and hole barriers for each particular silicide add up to E_g , or 1.1eV . Each of the silicides has to be biased toward a particular carrier, one, with a low electron barrier, must be used in NMOS devices, the other, with a low hole barrier, must be used in PMOS devices. Just as NMOS and PMOS are thought of as complementary to each other, these new silicides can be also be thought of as complementary. Fig. 3.1 shows a thin-body CMOS technology implemented with complementary silicides (CS).

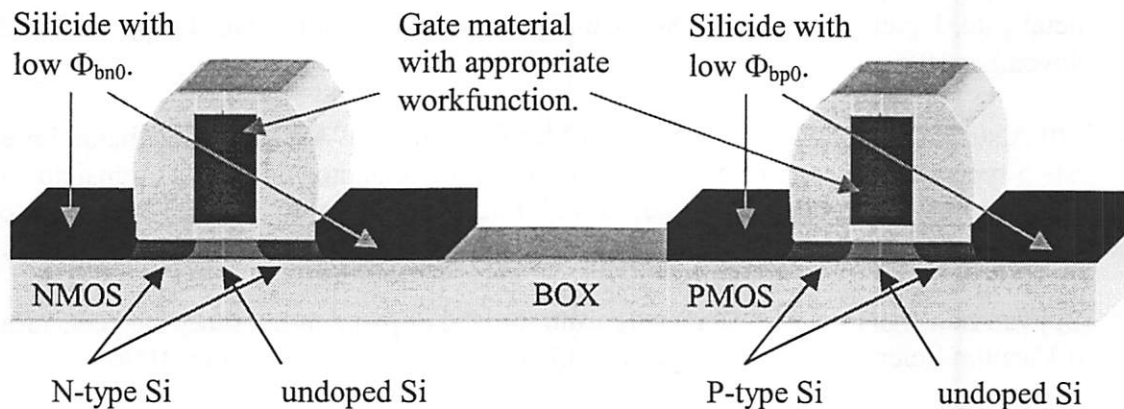


Figure 3.1: CMOS implemented with thin-body structures using complementary silicide source/drains with doped extensions(DCS). In addition to two silicides two gate materials with appropriate workfunctions may be needed in order to control threshold voltage.

How low do the barriers have to be to satisfy series resistance requirements of a thin-body device? The conventional contact resistance model presented in section 2.1.1 is used in Fig. 3.2 to map out the contours of acceptable and unacceptable resistance, as a function of extension region doping concentration and barrier height, for a 10nm thin-body transistor:

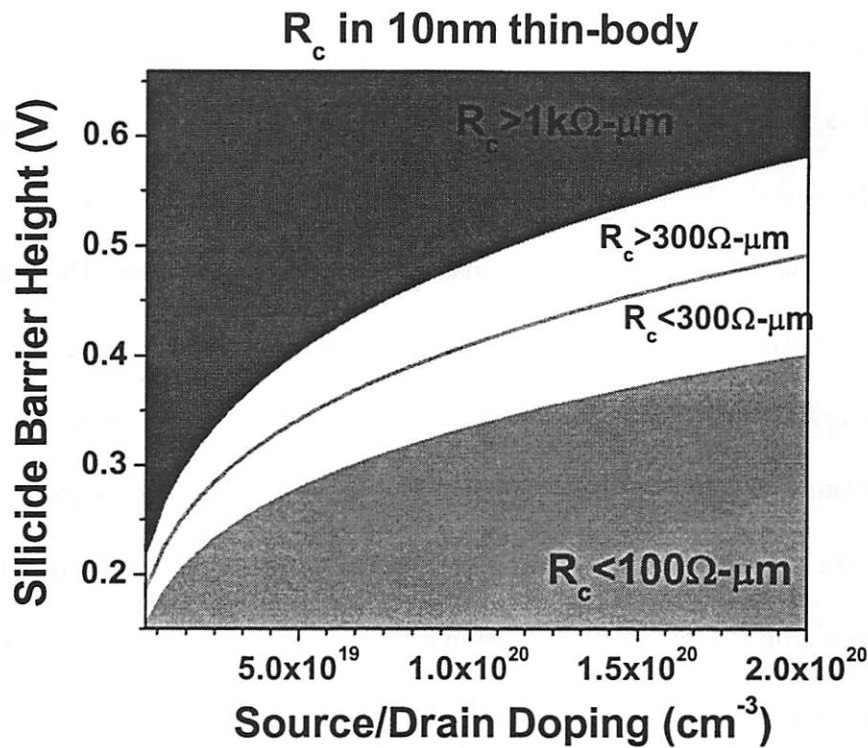


Figure 3.2: Design space for the DCS structure. In the green(gray) region parasitic resistance does not affect device behavior, in the red(dark) region parasitic resistance cripples device performance.

The highest acceptable contact resistance is assumed to be $300\Omega\text{-}\mu\text{m}$, but to realize as little performance loss as possible R_c should be $<100\Omega\text{-}\mu\text{m}$. These contact resistance values, shown in green, can not be achieved with a mid-gap silicide, which has a barrier

of 0.6eV. For a doping level of around $1E20\text{cm}^{-3}$ a barrier height of 0.38eV or lower is required. The design space for the doped complementary silicide thin-body (DCS) structure is fairly large, with any material with a barrier of less than 0.38eV and any doping larger than $1E20\text{cm}^{-3}$. Considering that barrier heights vary from 0.24eV to 0.88eV, and the electrically active doping limit is approximately $3E20\text{cm}^{-3}$ for NMOS, and $2E20\text{cm}^{-3}$ for PMOS, the design space for the DCS structure and fabrication constraints have a large intersection.

It is also possible to design a complementary silicide thin-body structure without the use of doping. As will be discussed in following chapters, there are numerous advantages and disadvantages to eliminating doping in a transistor. The design of such a structure follows the approach of the silicide source/drain device presented in section 2.3. Fabricating a silicide source/drain structure on a SOI layer improves its short channel characteristics. When the SOI thickness approaches the typical thickness of the thin-body design the devices start resembling the DCS design with one important difference, the doped region is replaced by extension of the silicide itself, as shown in Fig. 3.3.

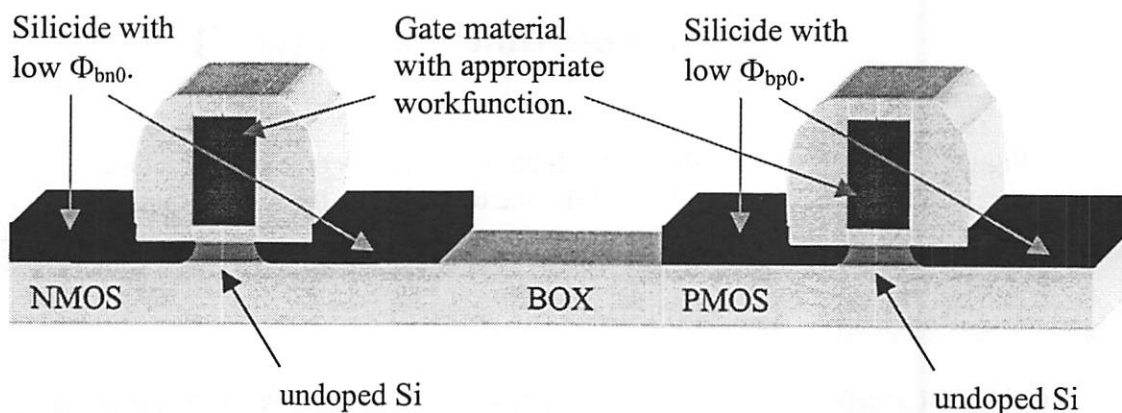


Figure 3.3: CMOS implemented with thin-body structures using undoped complementary silicide(UCS) source/drains. In addition to two silicides two gate materials with appropriate work-functions may be needed in order to control threshold voltage.

Doped and undoped complementary silicide thin-body devices(DCS and UCS), are both investigated in this work as a general solution to the resistance problems of thin-body transistors.

To see the barrier height design space for the undoped complementary silicide thin-body (UCS) structure it is necessary to examine the electric field at the silicide-silicon interface of the DCS structure. A contour plot of the maximum electric field at the silicide-silicon interface of a DCS structure superimposed on the resistance requirements is shown in Fig. 3.4.

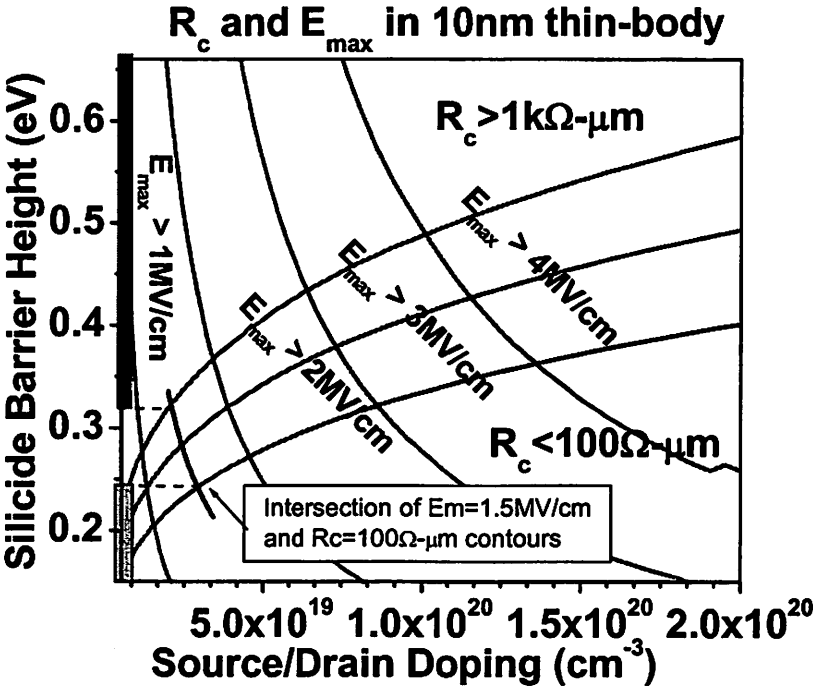


Figure 3.4: Silicon-silicide boundary e-field contours superimposed on the contact resistance contour of a DCS structure. On y-axis: The design space for the UCS structure, assuming that an effective e-field of 1.5MV/cm can be generated by the gate without the use of doping. In green(gray) region contact resistance does not limit current, in red(dark) region contact resistance dominates.

The main function of the doping in the DCS structure is to provide an electric field at the silicon-silicide interface, this field lowers the contact resistance via methods discussed in section 2.1.1-2.1.2. In order to remove the doping and maintain a low contact resistance in a UCS structure the electric field must be induced by other means. The way that the UCS structure induces the electric field is with the gate-to-source bias. The magnitude of the field that can be induced at the source by the gate can be calculated given the device geometry and applied biases. Assuming a 1.5nm gate oxide and a 0.85V gate bias, the simulated value for E_{\max} in an UCS structure is 1.5MV/cm (simulation performed on SILVACO a 2d device simulator). This electric field is sufficient to lower the contact resistance to the desired value if the barrier height is lower than 0.25eV, see Fig. 3.4. This barrier height design space is much more demanding than that of the DCS structure, and unfortunately has little intersection with fabrication limits. There are only two metals that achieve such a low silicide barrier for holes: Pt and Ir[1], and there are no known solutions for electrons although Eu and Yb are possible candidates[2].

There are some advantages to the UCS structure that make solving its fabrication difficulties a worthy pursuit. First, by eliminating doping in the source/drain regions the fabrication temperature of the entire device can be decreased from $\sim 1000^{\circ}\text{C}$, needed for dopant activation, to $\sim 750^{\circ}\text{C}$, needed for silicide anneal, gate oxide, and CVD steps. The abruptness of the source/drains in the UCS structure could effectively be infinite, since the junctions are potentially atomically sharp. Such junction definition is impossible to achieve with doped source/drains with currently known methods. Also, UCS devices have a Schottky barrier that helps to keep the leakage current low in the off state, and thus are expected to have better short channel characteristics than the DCS structure.

When comparing the DCS and UCS structures it is useful to remember that the DCS structure will behave like a traditional MOSFET in the limit that the source/drain doping is high. So any advantages the UCS structure has over the DCS structure it will also have over the traditional MOSFET.

Both the UCS and the DCS suffer from difficulties in integrating the C-silicides into the fabrication process. The fact that two silicides are used will necessitate the patterning of the silicidation steps. The exact integration methods will be discussed in another chapter, but conservative approaches require two extra lithography steps, a P-silicide lithography and a N-silicide lithography, similar to the NMOS and PMOS lithography steps used for dopant definition.

There is a fundamental challenge in working with low barrier silicide materials that is important to understand. Low-barrier NMOS materials have a low workfunction, making them chemically reactive. Europium for example is so reactive that it ignites in air[3]. As expected, this leads to processing difficulties that can not be solved by simply changing silicidation material. Once the silicide is formed it is in general more stable, but reactivity problems can still make back of the line processing difficult. Low-barrier PMOS materials suffer from the opposite problem. High workfunction metals have low chemical reactivity. For example, Iridium is so stable that it does not react with any known acid[3]. Platinum reacts only with a very strong acid mixture, that also etches PtSi[4]. Integrating such extreme materials is difficult, and the processing requirements stringent. The problems caused by low and high workfunction values get more severe the further the metal workfunction is from the silicon midgap value. Unfortunately, a metal with a workfunction value around the silicon midgap gives a silicide with a high barrier

to electrons and holes. This fact suggests that a trade off exists between the height of the barrier and the silicide processing difficulty. A low barrier can only be obtained by overcoming the difficulties associated with integrating low and high reactivity materials.

References:

[1] S. P. Murarka, "Refractory silicides for integrated circuits," *Journal of Vacuum Science and Technology*, Vol 17 (no. 4), p. 775, July-Aug. 1980

[2] J. A. Knapp, and S. T. Picraux, "Epitaxial growth of rare-earth silicides on (111) Si," *Applied Physics Letters*, Vol. 48 (no. 7), p. 466, 17 Feb. 1986

[3] Mark Winter, WebElements; <http://www.webelements.com/webelements/index.html>

[4] Myron J. Rand and James F. Roberts, "Observations on the formation and etching of platinum silicide," *Applied Physics Letters*, Vol. 24, No. 2, pp. 49-51, 15 Jan. 1974

4.0 Description of Device Fabrication

This chapter will describe the process used to fabricate thin-body complementary silicide(CS) source/drain devices. Special attention will be given to the difficult steps of silicide formation and lithography, and steps that required detailed development such as the alignment procedure and metalization.

4.1 Overview of process flow

Fig. 4.1 graphically demonstrates the process used in this experiment, and Appendix A gives detailed process parameters for every step. The fabrication of CS devices started with SOITEC[1] silicon on insulator (SOI) wafers. SOITEC SOI is fabricated by a layer transfer technique that relies on fracturing the silicon at a depth defined by a H₂ implant. SOI wafers fabricated by this method have a high quality

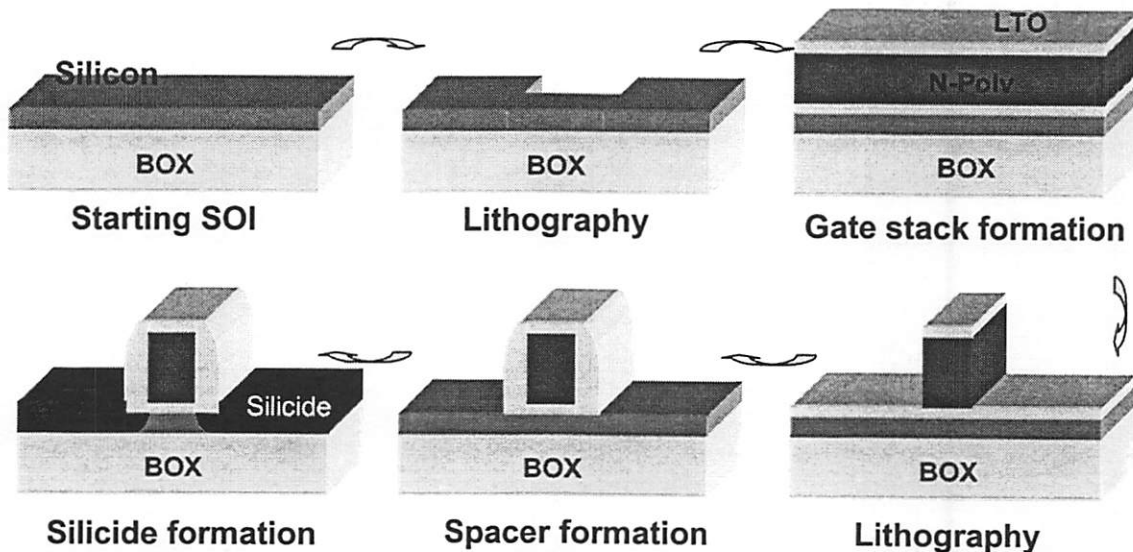


Figure 4.1: Simplified process flow for the fabrication of CS devices. BOX stands for Buried Oxide, N-poly stands for N-type poly-silicon, and LTO stands for Low Temperature Oxide.

silicon-SiO₂ interface since the oxide is generated by thermal growth. The fact that thermal oxidation is used to generate the SiO₂ layer also results in uniform thickness of both the silicon and oxide layers. The alternative SOI, fabricated by the SIMOX technique of implanting a high concentration of oxygen and annealing, was found to show significant variation in silicon thickness by TEM analysis.

The SOI silicon layer was thinned by two oxidation steps from 1000Å to ~160Å. A 40% germanium SiGe layer was then deposited to a thickness of ~8000Å. Alignment features were patterned in the SiGe layer with an optical lithography step, and etched in a HBr dry etch. Alignment marks were placed at the corner of each e-beam field. Alignment mark formation was followed by mesa lithography. Mesa lithography consisted of an optical and electron beam exposure using g-line and calixarene resists respectively. This double resist technique was developed to give a reasonable exposure time and 15nm resolution, and is described in more detail in section 4.3.2. The silicon mesa was patterned with a reactive ion etch(RIE) using both resists as the mask. In order to remove silicon regions damaged by the etch, 65Å of sacrificial oxide was grown and removed with HF. The gate stack consisted of 37Å of thermally grown SiO₂, followed by a n-type poly-silicon layer deposited by chemical vapor deposition(CVD), to a thickness of 550Å. A 160Å SiO₂ hard mask was deposited by CVD to cap the gate layer. The height of the gate stack was kept under 750Å in order to keep the aspect ratio of the gate below 5:1 for gate lengths of 150Å.

Gate lithography was also performed in two exposures, one optical and one electron beam. First the calixarene electron beam lithography step was used to pattern the fine features in the oxide hard mask. Then the large features were patterned with G-line

lithography. The gate etch used both the hard mask and the optical resist as the etch mask. On select wafers the oxide hard mask was trimmed with a short HF dip to reduce minimum gate length. Following the gate definition, a 200Å thick oxide spacer was deposited by CVD. Due to the porous nature of CVD oxide the film was densified with a 950°C 10m anneal. The spacer was etched with a timed RIE etch. Following spacer etch, wafers were diced up into 2mm wide strips for metal processing. The strips were cleaned with a O₂ plasma and an 'HF-last' dip just prior to metal deposition. For NMOS strips, Er was deposited at 5E-9 torr (UHV), and ErSi_{1.7} was formed with a 400°C anneal. For PMOS, Pt was deposited at 1E-6 torr, and PtSi was formed with a 400°C anneal. The stoichiometry of final silicide phases has been established in the literature[23,24], and was not verified experimentally in this work. Unreacted metal was etched with a wet etch. The silicidation process is described in detail in sections 4.3.4.1-2.

4.2 Critical process windows

Several of the standard processing steps require special scrutiny because of their tight controllability requirements. These steps require careful monitoring and measurement of the process variability with test wafers. Gate oxidation is the first such step, due to the difficulty of measuring thin oxides and the variability of the oxidation rates, growing a 20Å gate oxide is challenging. Measuring the test oxidations with spectroscopic ellipsometry helps to decrease measurement uncertainty, also performing the test and run oxidations in tandem helps to reduce uncertainty. Due to process variability the gate oxide in this experiment was ~17Å thicker than desired, at 37Å.

Another difficult step is the sacrificial oxide removal. HF must be used to remove the sacrificial oxide prior to gate oxidation. If this HF dip is too short any sacrificial

oxide that remains increases the thickness of the gate oxide. If the HF dip is too long the mesa can be severely undercut due to the etching of the buried oxide (BOX). The undercut is filled with the gate material during the gate deposition, and is masked during the gate etch by the mesa itself. Precise timing and exact control of the HF concentration is needed to perform this step. Critical HF solutions should be prepared just prior to use and stirred for at least two minutes to ensure uniformity. Undermixed or unmixed HF solutions can show as much as an order of magnitude variation in wet-etch rates. The etching of the spacer oxide is another difficult step. The dry etch selectivity of the oxide etch to silicon is only ~2:1, in the UCB Microlab dry etcher. So when etching 200Å of oxide spacer only a 10% overetch window is available given the constraint that only 10Å of the mesa is to be etched. This requires a careful calibration of the HTO etch rate and a control of the etch time to within 1~2 seconds.

Another critical HF dip is performed just prior to the metal deposition. An 'HF-last' clean is required to passivate the silicon surface and prevent native oxide formation. However this clean also etches the spacer. Since the spacer profile is critical to the amount of underlap that the source will have with the gate it is important to control this etch precisely. Careful test calibrations of this etch rate are needed especially since the HF wet etch rate of densified HTO is different from both thermal oxide and normal HTO.

4.3 Specific processing modules

The lithography and silicidation modules of this experiment had requirements that are quite different from that of the standard process. Thus, the lithography and silicidation steps required careful research and development before they were

incorporated into the process flow. This section will discuss the background knowledge needed for the development of these steps and the development itself.

4.3.1 Lithography - Electron beam lithography fundamentals

When optimizing the resolution of an electron beam lithography step it is important to understand the interaction between an e-beam resist and the electron beam itself. Especially since it has been determined that electron beam lithography resolution is primarily limited by these interactions, and not by the size of the beam. The actual focused beam spot size on the sample surface is roughly equal to the state of the art SEM resolution, about 2nm[2]. This can be further decreased with improvements in lens quality and the correspondingly higher beam convergence angles. However the minimum pattern sizes typically obtained with EBL are on the order of 10-100nm and depend heavily on the resist type, even when the exposure dose is optimized[3,4]. So to understand the fundamental resolution limits of electron beam lithography the details of electron - resist interactions have to be investigated.

In most EBL systems, the beam consists of high-energy electrons, typically in the 30-100keV range. Higher energy electrons have several advantages; first they penetrate further into the resist. Second, increased electron speed reduces the interactions between the electrons in the beam. As an e-beam enters a solid, it interacts with both the nuclei and the other electrons. Any single electron-nuclei collision is likely not to change the electron trajectory drastically[5]. However, once in a while the electron will get scattered though a large angle. The likelihood of e-n interactions decreases with increasing electron velocity, because at high velocities the electron spends less time in the vicinity of any particular nuclei. Due to this dependence, the mean free path between collisions is

a strong function of electron velocity, for example the mean free path in gold varies from 8-0.6nm for 30-1keV electrons[5]. Of course, since the average angle of scattering is small, the actual penetration depth is much larger than the mean free path. The large mass difference between electrons and protons makes their collisions elastic, with the electron losing less than 1eV per interaction. So high-energy electrons do not lose energy by directly generating atomic vibrations, they do it by interacting with other electrons. Due to the multitude of possible electronic transitions, these inelastic e-e collisions are poorly understood. However some general formulae have been worked out to calculate the average electronic stopping that an electron sees when traveling through a solid[6]. SEM studies of the secondary electrons generated by these collisions indicate that the energy lost in each inelastic collision is small, typically under 50eV[5]. So an electron from the beam, called a primary electron, directly generates thousands of secondary electrons, with a ~50eV typical energy, as it travels through the solid. Because the primary electron's trajectory is virtually unchanged by the inelastic collisions, a friction like force, which continually slows the electron down, can model these interactions.

The combination of these two interactions gives us a picture of how an electron beam behaves in matter. Monte Carlo simulations incorporating these effects show that electrons from the primary beam adopt a stochastic distribution in a tear drop shape, with the size of the tear drop dependent on beam energy[7]. For a 10keV electron the average penetration depth is 0.3 μm , but it grows quickly with beam energy.

The two effects that are thought to limit EBL resolution are: the delocalization of inelastic electron-electron interaction due to the range of electromagnetic potentials, and the lateral penetration of the resist by energetic secondary electrons. In other words the

resist in the vicinity of the electron beam can be exposed by the electromagnetic interaction with the beam, or by collisions with secondaries generated from such interactions. The electromagnetic interaction exposure mechanism has been characterized by an interaction length, λ , that is inversely dependent on a resist parameter, ΔE [8], the bond energy. For the PMMA resist, the bond energy, ΔE , is about 5eV[10]. Assuming a beam energy of 40keV yields, $\lambda=5.7\text{nm}$ (from relation in [10]), which is comparable to the minimum spot size radius in PMMA of about 10nm. This theory also correctly predicts the relation between the resist transition energy, ΔE , and the resist resolution. Resists with lower ΔE , like most optical resists, have lower EBL resolution[9], while most resists with higher ΔE , have higher resolution. For example, LiF, with $\Delta E=15\text{eV}$, has a resolution of 3nm[10].

The second effect which could explain resist resolution is the straggling secondary electrons which expose the resist[11,12]. Although no quantitative model of this theory exists, due to the complexity of e-e interactions, a qualitative model is easy to understand. Secondary electrons generated by electromagnetic interactions penetrate the resist at random angles. The energies from these secondaries vary from 0-70eV; the more energetic ones penetrate further and generate even more free electrons. Finally the low energy electrons interact with the resist and expose it. The exposure profile resulting from such a model is a ring with a radius of about 10nm, the penetration depth of the high-energy secondary electrons. The dependence of the resolution on the resist energy threshold can be explained by this model as well, closer to the primary beam the secondary electrons are more energetic so they expose higher energy resist. As the resist

energy threshold decreases the less energetic electrons with a broader distribution can participate in resist exposure.

Probably both of these effects contribute to the resist resolution limitations, although recent studies tend to favor the electromagnetic interaction model[13]. The clear way to increase resist resolution, regardless of which of these two modes is dominant, is to use higher energy resist. Inorganic materials are a good candidate for high resolution resists due to their strong bonding. NaCl, which has very strong ionic bonds, has been used to pattern very thin structures, with dimensions on the order of several nm[14]. However since the exposure process relies on drilling the material from the crystal, it contaminates the previously exposed areas, and leaves residue at the bottom of deep structures. LiF does not have this problem. When hit with an electron beam, F desorbs from the film, and Li diffuses into the unexposed areas, making development unnecessary. Structures down to 3nm have been fabricated using LiF as a resist. Unfortunately, inorganic resists are difficult to use in microelectronic fabrication and require a very high electron dose.

Experimentally, the relationship between resist sensitivity, resolution, and bond energy is clear; higher bond energy virtually guarantees that a resist will have a lower sensitivity and a higher resolution. In general, for every order of magnitude increase in resolution, the resist sensitivity decreases by a factor of 100. In Fig. 4.2 is a plot showing this relation for some common resists[14].

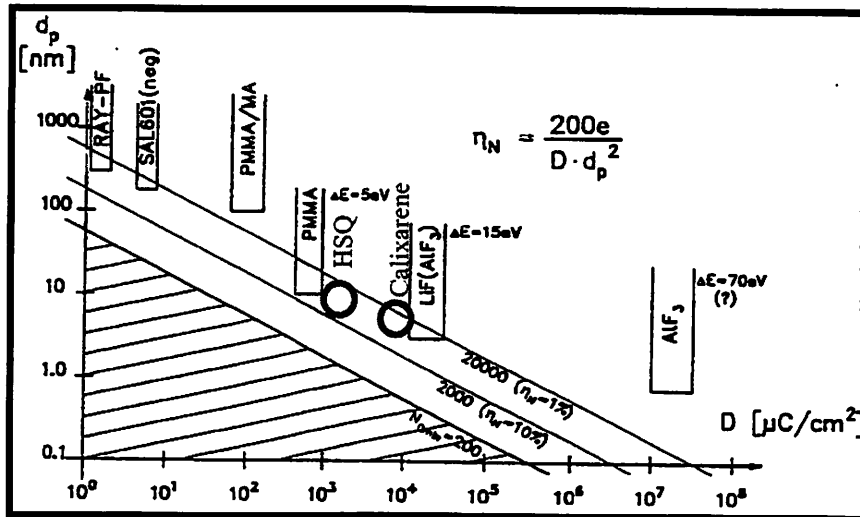


Figure 4.2: The relationship between resist resolution and exposure dose. The resist bond energy is also indicated where it is known. Two high-resolution process compatible resists are shown in red. Adapted from [14].

Of special note are calixarene[15] and HSQ[16], relatively new negative high-resolution e-beam resists. These resists have the resolution to define device features down to 10nm. Unfortunately, due to their high exposure energy their dose is high. Process methods had to be developed to successfully integrate them into a timely fabrication plan.

Chemically amplified resists such as SAL and UVHS-II have significantly lower dose levels. Their resolution is reduced due to the diffusion of the active acid during the post exposure bake(PEB). Resolution of these resists can be increased, at the expense of sensitivity, by limiting the PEB time. However even with an optimized PEB, such resists have resolutions of 40-100nm.

4.3.2 Lithography process

As can be summarized from the previous section, high energy resists will give high resolution, unfortunately the same resists will also have a very high dose. Since with the available lithography tool the electron energy is fixed at a high 50-100keV, the dose for high-resolution resists will necessarily be large.

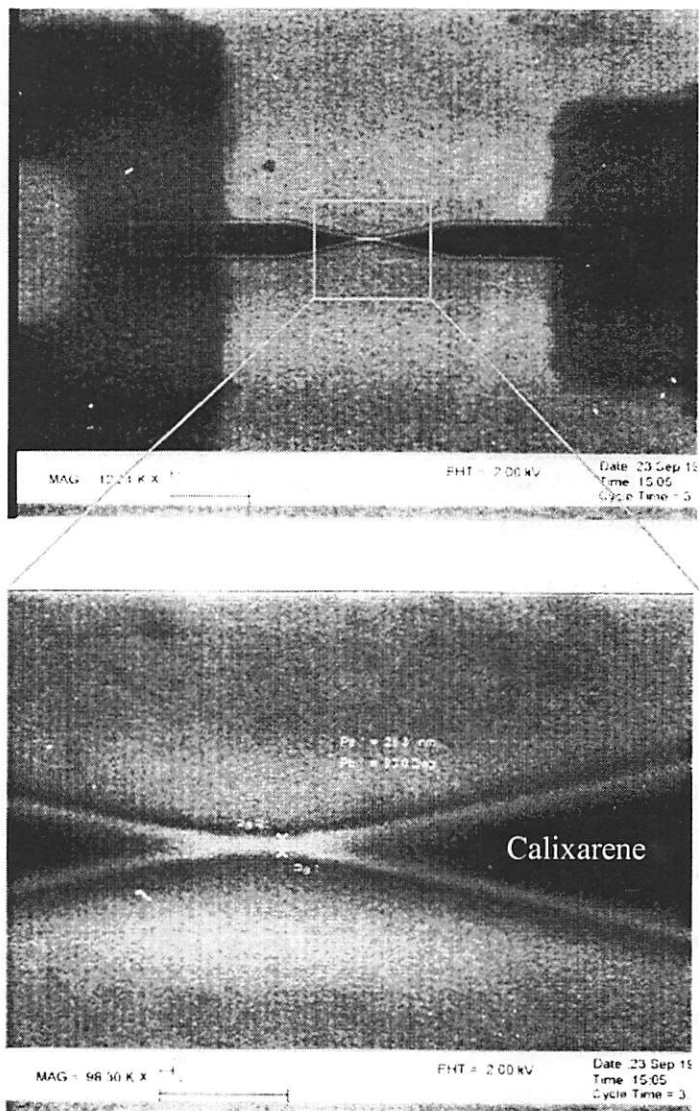
In order to maximize the resolution, the highest resolution resist, calixarene [15], was used to define the critical device dimensions. Methods had to be developed to reduce the calixarene write time. These methods rely on patterning a single lithography layer with two exposures, one optical, exposing coarse features, and one e-beam, exposing fine features. Although for arbitrary patterns such a division might not result in a drastic reduction in e-beam write time, for many patterns of interest in the research community this division can reduce the write time by many orders of magnitude. The first method, calixarene-last, consists of an optical G-line lithography step followed by a calixarene exposure and modified development. This method was used to define the mesa layer. The second method, calixarene-first, initially transfers the calixarene pattern to a silicon dioxide hard mask with a dry etch and then defines the large features with an optical lithography step. This method was used to define the gate layer.

4.3.2.1 Resist Processing

Calixarene has several chemical forms, in this study the 4-methyl-1-acetoxycalix[6]arene was used. The 1g of calixarene powder was dissolved in 25g of o-chlorobenzene, and 5g of dichloromethane. The addition of dichloromethane helped the calixarene powder to dissolve faster. The 3% mixture of calixarene can be spun at 3000rpm to produce an approximately 500Å thick resist layer.

In the calixarene-last process a standard G-line optical lithography step was completed up to and including development. The G-line process used 1000Å of G-line resist, and the standard soft and hard bakes of 1m 90C, 1m 120C respectively. Next, calixarene was spun on the wafer at 2000 rpm, to a thickness of 500Å. Keeping the ratio of the G-line thickness to calixarene thickness small is critical. If the G-line resist is significantly thicker than the calixarene thickness, surface tension causes the accumulation of calixarene at the boundary of the G-line resist. This accumulation can totally deplete the field regions around the G-line pattern of calixarene. The calixarene resist was exposed with a dose of 20mC/cm² at 100keV using the Nanowriter[17], a direct write electron beam tool. Development was done in xylene for 30s, followed by a rinse with running DI water. The standard xylene rinse of IPA was not used since it dissolves G-line resist. Both patterns were then transferred to a silicon layer using a dry etch. Fig. 4.3 shows both resists prior to etching, and Fig. 4.4 shows the 15nm silicon pattern fabricated with this process. If required the line width can be further reduced by ashing the calixarene resist in an oxygen plasma.

Figure 4.3: The combination of G-line and calixarene resist in the calixarene-last process, prior to etching. Inset: A 15nm pattern in calixarene.



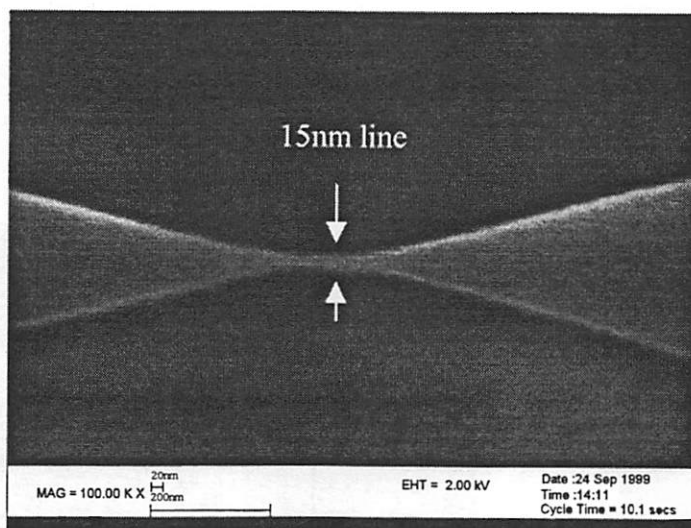


Figure 4.4: A 15nm feature in 100Å of silicon following an RIE etch and resist removal. The narrow section of the mesa will serve as the channel for the thin-body complementary silicide device.

Several other resists were tested for use in the calixarene-last process. I-line resist dissolves in xylene making it incompatible with calixarene development. However SAL-601, a negative electron beam resist, is chemically compatible with calixarene. The standard SAL-601 resist is unaffected by xylene after development. A calixarene-last process using SAL-601 also results in 15nm resolution but requires two electron beam exposures.

The simplest way to do a calixarene-first double exposure process is to expose and develop the calixarene resist, then spin, expose, and develop the low-resolution resist. When this was done, with G-line, I-line, or SAL-601 the calixarene swelled from a minimum resolution of 15nm to 40nm. The wet steps of the second lithography step

most likely cause this swelling. In order to preserve the 15nm resolution the calixarene had to be transferred to a non-organic hard mask.

The Calixarene-first process was done by spinning 500Å of calixarene onto a 200Å silicon dioxide hard mask. After the electron beam exposure and development, the hard mask was patterned with a dry etch and the calixarene was removed in an oxygen plasma. After the calixarene is removed, any optical lithography step can be performed to expose the low resolution patterns. During the etch both the oxide hard mask and the optical resist are used to pattern the underlying silicon layer. This particular process was developed to etch a narrow polysilicon gate; the high selectivity of silicon to oxide dry etching makes fabrication of 1000Å thick silicon gates possible with a thin layer of calixarene. The calixarene-first process resulted in a resolution of 20nm, although the resolution can be increased by 'trimming' the hard mask with a timed dilute HF etch. Results are shown in Fig. 4.5, where this process is used to pattern a 500Å thick polysilicon gate. The polysilicon lines fabricated with the calixarene-first process are wider and have considerably more line width variation than lines fabricated with the calixarene-last process. The resist process does not cause this roughness. It is caused by uneven etching of the polysilicon grains at these narrow dimensions. Fig. 4.6 shows the Calixarene resist line, in the calixarene-first process, prior to the hard-mask etch, this line is narrower and significantly straighter than the etched polysilicon line.

Figure 4.5: Combination of the G-line resist and oxide hard mask etched with calixarene. Inset: 20nm poly-gate pattern etched using calixarene-first process. Alignment to previous layer is better than 5nm. The SEM shows the device structure just after gate etch.

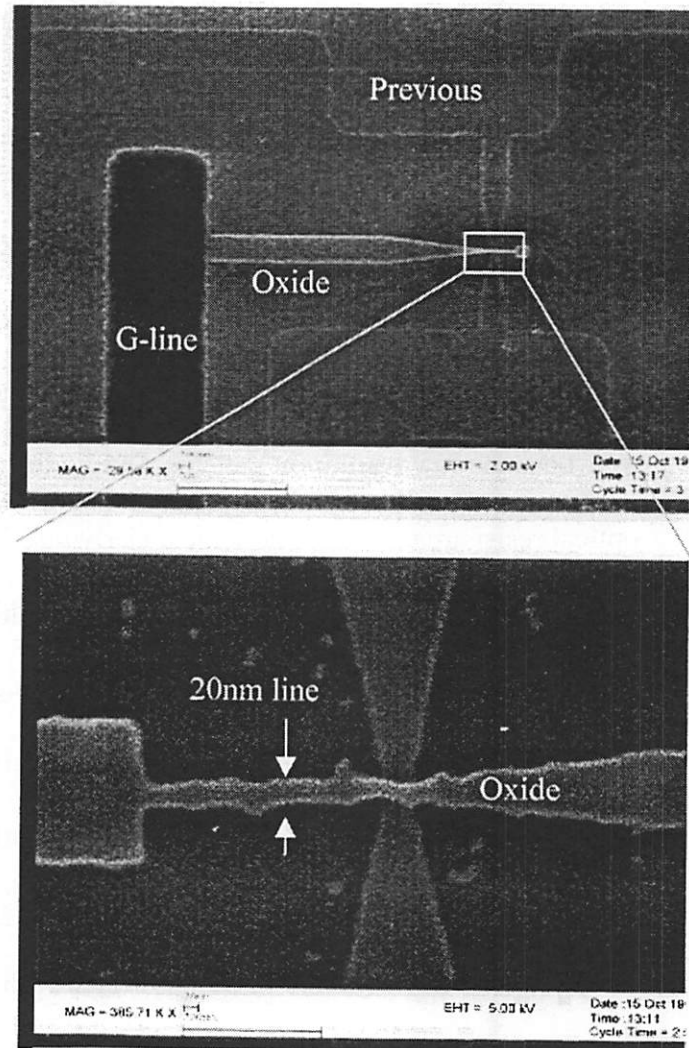
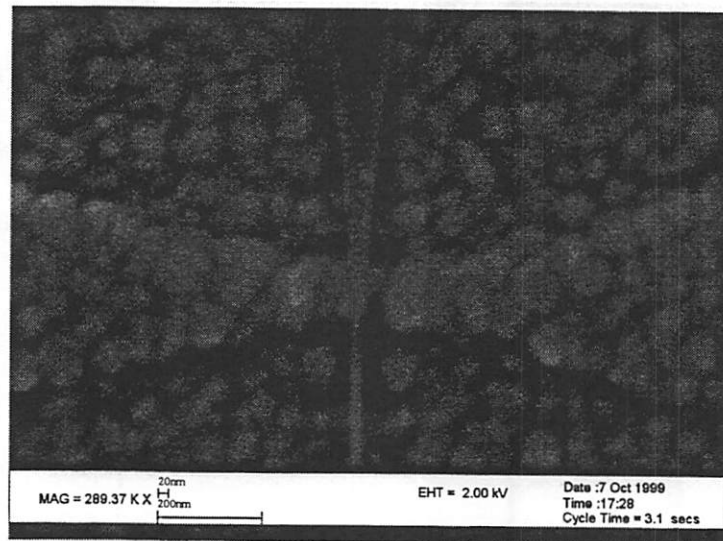


Figure 4.6: A 15nm calixarene line exposed in the calixarene-first process on 550Å of polysilicon. The resist line shows significantly less line width variation than the polysilicon line after etching. This resist feature will define the gate of the complementary silicide device.



4.3.2.2 Alignment Process

The methods presented above require careful alignment of the electron beam exposure to the optical exposure. Such alignment is not trivial since the contrast mechanisms for electron beams and photons are drastically different. Most material and topography variations are easily seen optically, provided they are not significantly smaller than the wavelength of the illuminating light. Optical alignment marks are typically made from the same material as the previous layer, often silicon dioxide, silicon, or silicon nitride, for MOSFET applications. Unfortunately these optical alignment marks tend to be difficult to detect under the electron beam. Most e-beam writing tools use a backscattered electron detector to be able to detect material differences from below the photoresist layer, the signal received by the detector is proportional to the Z^2 density of the material. Unfortunately silicon dioxide, silicon, and silicon nitride have very similar Z^2 densities making alignment mark contrast poor. In addition, substances that have a high Z^2 density, like gold are often contaminants in silicon nanoelectronics.

Silicon germanium is a good alignment mark material between electron and photon lithography steps. It is a material that is compatible with most silicon nanoelectronics processing. It etches easily in silicon dry etches, can be deposited by LPCVD, stands up to standard cleans, and is not a contaminant. Marks made from SiGe 6000Å thick had excellent contrast under a 100keV, 300pA beam. The 1:1 ratio of silicon to germanium is determined by the fact that silicon germanium marks with a significantly higher concentration of germanium are damaged by sulfuric acid, while marks with lower concentrations have lower contrast.

4.3.3 Silicide formation – Physical vapor deposition fundamentals

The formation of a silicide consists of a careful preparation of the silicon surface, the deposition of the metal to be reacted, and the reaction anneal. After the silicide is formed an etch must be performed to preferentially remove the unreacted metal. A second anneal may be performed after the etch to alter the silicide microstructure. The successful formation of CS devices requires a detailed understanding of all the steps in this process.

4.3.3.1 Surface preparation and cleaning

Prior to deposition, the surface of the silicon must be carefully cleaned in order to insure that the metal and silicon react uniformly. An oxide or any residue can totally prevent the reaction or cause it to occur at varying speeds that are a strong function of the local surface contamination. Such a partial reaction is very difficult to control especially on thin films, where a uniform silicidation front is critical to obtaining correct stoichiometry. Removing the oxide from a silicon film seems to be a difficult problem since silicon grows a monolayer of oxide at room temperature in seconds, fortunately it is easy to passivate the surface silicon bonds with hydrogen. This is accomplished with a dilute HF dip, followed by an N₂ drying, in what is commonly referred to as the HF-last clean. The Hydrogen passivation lasts for at least an hour at room temperature and pressure, giving sufficient time to load the sample into a vacuum system. At higher substrate temperatures the passivation will last a shorter time; with any significant partial pressure of oxygen silicon will depassivate in one second at 400C[18]. In UHV, hydrogen will desorb from the surface by itself at temperatures above 600°C. Water accelerates the depassivation process and reactive ions such as O₂⁺ will remove it instantly.

4.3.3.2 Overview of metal deposition techniques

The deposition of thin metal films is a key technology in developing a silicidation process. The three types of processes commonly used in the microelectronics industry for deposition are: evaporation, sputtering, and chemical vapor deposition(CVD). Evaporation is perhaps the simplest deposition method; a source material is heated up in high vacuum to a temperature where its vapor partial pressure is significant. Vapor thermally desorbed from the source travels in a straight line toward the cooler sample where it deposits. The exact deposition process is complex and will be described in the next section. In sputter deposition the source material forms one electrode in a plasma while the sample forms the other one. By applying the correct bias on the plasma, the ions remove atoms from the source material by physical collision. Once the source material is in the plasma it can deposit itself on the sample. CVD requires that precursor gas(es) containing the material to be deposited be prepared. When such gas(es) are introduced into a chamber, at a certain temperature and pressure, they will undergo a surface assisted decomposition, depositing some material in the process. CVD is not possible for all materials since the right chemistry must exist for such a reaction to occur.

Each of these techniques has advantages and disadvantages. Evaporation is the most versatile, in that it allows one to quickly experiment with new materials. In evaporation, the source material is typically just a few grams of high purity element. Covalently bonded compounds can be evaporated in the same manner as elements, by thermally increasing the vapor pressure, however alloys pose a challenge since they often evaporate incongruently. For alloy evaporation, two or more independent source materials must be prepared, each with its own temperature control. Then the ratio of the

flux of the two species can be varied to give alloys of a certain composition. Such an evaporation is difficult to control and shadowing effects can play a significant role on non-planar substrates since the flux of the two species are necessarily non-parallel.

Sputtering requires that an electrode be fabricated from the material to be sputtered, and alloys can be sputtered from one target making stoichiometry control considerably easier.

CVD is not often used in the deposition of novel materials in the microelectronics field, since the chemistry for CVD has to be developed for each deposition material, and the synthesized gases are often highly toxic and expensive.

4.3.3.3 Film coverage during deposition

The profile of the film after deposition is strongly influenced by the deposition process. The typical profiles of films deposited by the methods described in this chapter are shown in Fig. 4.7. Evaporation tends to be line-of sight, since the mean free path at pressures required to get a pure film at any reasonable evaporation rate are much larger than the typical distance between the source and sample. This implies that source atoms arrive at the sample with a relatively high degree of spatial coherence, and shadowing effects are serious. The spatial coherence of the molecular flux in CVD is very low since molecules land from every direction. Also for CVD, the surface mobility of the precursor molecules is normally very high leading to a highly conformal step coverage that is mostly independent of process pressure. Sputtering has coverage that is between these two extremes. Typically, spatial coherence is low, and surface mobility is low, but each can be controlled to some degree by plasma bias and the substrate temperature respectively. The presence of the plasma complicates the deposition process by causing resputtering of the deposited material and increased surface mobilities. As a result the

step coverage in sputtering can be controlled to a large degree by the plasma energy and bias.

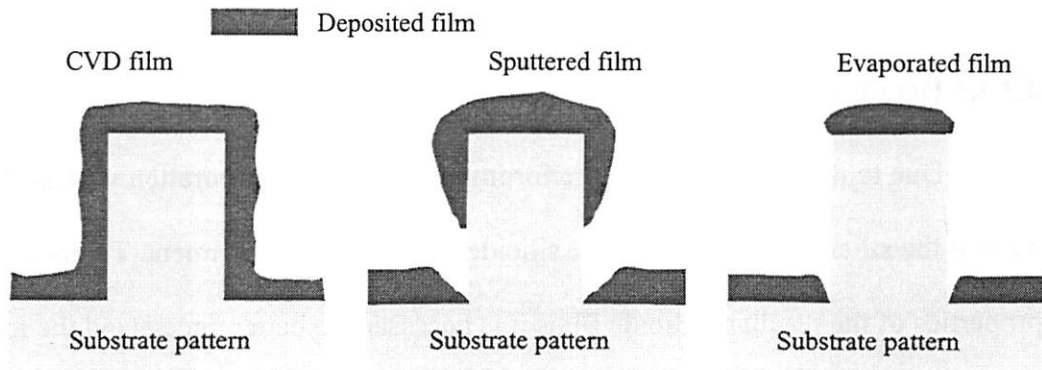


Figure 4.7: Deposition film patterns in red(dark) on the gray substrate for different deposition processes. On left is a conformally deposited CVD film, in the middle is a film deposited by sputtering at low power, and on the right is a film deposited by evaporation at high vacuum.

4.3.3.4 Deposition vacuum requirements

The vacuum requirements in any deposition system are dictated by the desired material purity. The flux of any gas in the vacuum system is given by the equation below[19]:

$$\Phi_{flux} = \frac{P}{\sqrt{2\pi mkT}}$$

Where Φ_{flux} is in $\#/cm^2\text{-s}$, and P is the pressure. Assuming an average radius of 4\AA and a sticking coefficient of 1, results in the standard rule of thumb that at $1\text{E-}6$ torr partial pressure there is one monolayer(ML) of deposition per second. For example, oxygen partial pressure needs to be $1.7\text{e-}6$ torr for 1 ML/s deposition. The sticking coefficients for most metals are close to 1, but for gas phase molecules such as H_2O and O_2 , the sticking coefficient is a strong function of the substrate, with reactive substrates having higher sticking coefficients. The presence of plasma can radically affect the sticking

coefficient of a species, since ions tend to be much more reactive. Since the sticking coefficients are uncertain, it is a good policy to keep the partial pressures of impurities significantly below $1e-6$ torr.

4.3.3.5 Details of evaporation

Due to its versatility and superior impurity control, evaporation was used to deposit the silicidation metals, in the silicide source/drain experiment. To understand the properties of the resultant silicide films it is necessary to better understand the formation of thin films during evaporation[20]. When an atom from the vapor phase comes into the vicinity of the substrate the Van der Waals force can trap the atom at the surface. Two distinct potential minima exist for most surface absorbed atoms. The first is the potential well caused by the induced Van der Waals attraction, and the repulsion of the electron shells, and is a few tens of meV deep. When an atom is trapped in this potential it is termed physisorbed. The second potential well, typically a few eV deep, is caused by the chemical bonding of the adatom to the substrate, an atom in this position is termed chemisorbed. Due to the large energy difference in the two wells it is not surprising that physisorbed adatoms have a much easier time migrating on the surface than chemisorbed adatoms. Since evaporation is a low energy process, (kT at $2000K = 0.17eV$) adatoms tend to be first physisorbed, then undergo some surface migration before they are chemisorbed. The thin film growth during evaporation is often divided into four distinct stages:

Nucleation – Atoms initially deposited on the substrate gather into small grain crystals
Crystal Growth – Nucleated grains grow by the migration of additional atoms
Coalescence – when adjoining crystals meet they form a grain boundary
Film growth – Adjacent grains compete for adatoms, as some grow and some shrink as the film is deposited, normally resulting in a columnar microstructure.

In nucleation, adatoms diffuse on the surface to form groups that typically have one crystal axis determined by the substrate but typically have semi-random radial orientation. The density of the nuclei will in part determine the column diameter at the later stages of growth. Higher substrate temperature and lower adatom energy (to suppress chemisorption) tend to maximize the surface diffusion, and increase grain size. As additional atoms land and diffuse the crystals grow; this growth can be interrupted by the accumulation of surface impurities, or when crystals start to meet each other. Once the growing crystals meet the film starts to grow vertically. If the temperature of the growing film is sufficiently high, the grains will compete for atoms at the grain boundaries and some grains may grow or shrink as the film is deposited. The exact microstructure of the film is dependent on the surface migration of the adatoms, which for evaporation is determined by the substrate temperature. The model that is often used to describe the possible microstructures in a deposited film is the structure zone model (SZM) [21].

4.3.3.6 Structure Zone Model

A simple structure zone model is shown in Fig. 4.8. Typically the SZM is divided into four different zones as a function of the ratio of the substrate temperature (T_d) to the melting point of the deposition material (T_m). The reason that T_d/T_m is a good indicator of

film morphology is because the surface mobility of the adatoms for most substances is similar given the same T_d/T_m value.

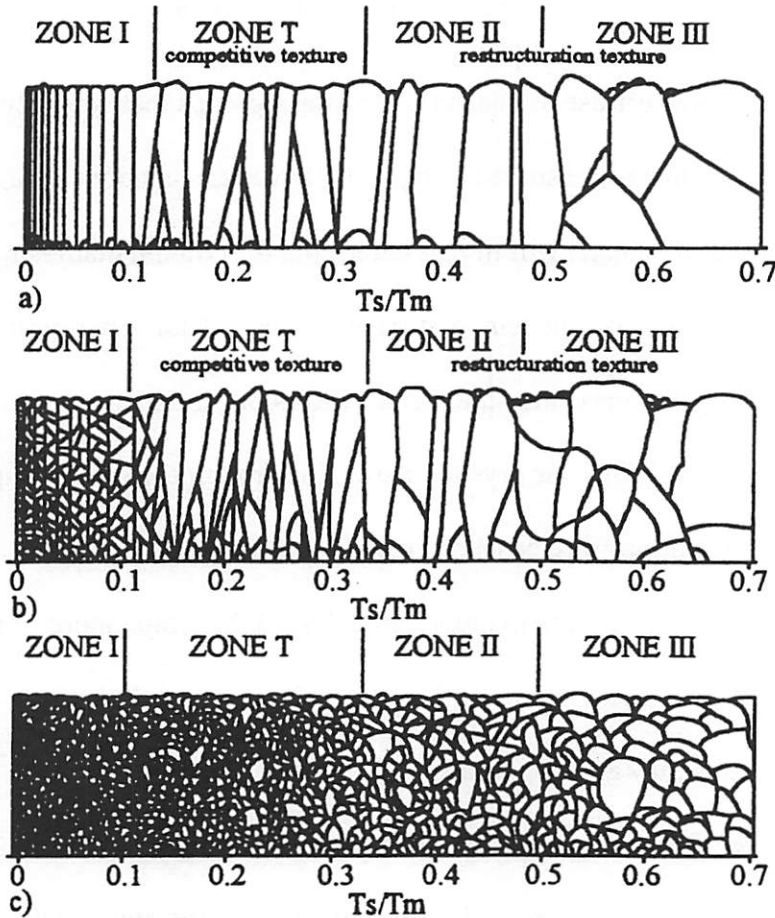


Figure 4.8: Structure zone models from [21]. $T_s=T_d$, lines indicate grain boundaries in film. SZM shown in sub-figure a assumes low impurity density, SZM in b assumes medium impurity density and SZM in c assumes high impurity density. There are 4 Zones in this SZM, Z1, ZT, Z2 and Z3. For simplicity ZT is considered as part of Z1 in this work.

Zone 1 indicates a region where surface mobility is not significant; atoms largely stick where they land. Expectedly this leads to a short range ordered amorphous structure of the film. It also leads to the film being very porous. Typical film densities are 0.5-0.8 of the bulk material density. The porosity is caused by stochastic shadowing of voids in the film that, due to the low surface diffusion, can not be filled with material. Film porosity can have large effects on resistance, as well as wet etch rates. The substrate temperature zone where Z1 microstructure predominates is $0 < T_d/T_m < 0.3$.

Zone 2 is a region where higher temperature causes adatom surface diffusivities to be significant on the time-scale of the evaporation. In this region crystals of various size nucleate at the onset of the deposition. Once the film is continuous, each crystal grain grows in its own orientation as the film thickens producing a columnar microstructure. The columns are typically capped with a dome due to the free surface energy of the grain boundaries. As expected, a higher substrate temperature results in larger columns and higher crystal quality in the columns themselves. The Z2 microstructure occurs for $0.3 < T_d/T_m < 0.5$.

In Zone 3 both the surface diffusivity and the grain boundary diffusivity is significant, columns that develop may grow and shrink due to the diffusion of atoms across the grain boundary. The surface free energy of the columns causes larger columns to grow at the expense of smaller ones as the film grows in thickness. The Z3 microstructure occurs for $T_d/T_m > 0.5$.

The presence of impurities can have a drastic effect on the SZM. Impurities may segregate to grain boundaries and inhibit grain boundary diffusion. They may also segregate to the top of a growing column and terminate its growth forcing continuous nucleation steps. Although the exact role of the impurity will depend on its nature, in general impurities cause grains to grow to a smaller size than they would otherwise.

4.3.3.7 Silicidation reaction

After the metal deposition is complete, the silicide must be formed by reacting the silicon film with the metal film. The reaction occurs at a certain characteristic temperature, which is dependent on the metal and to a smaller degree on the metal film microstructure. For most metals the silicidation reaction occurs in several steps as the

temperature is increased. First, provided there is no barrier layer between the silicon and the metal, the metal rich silicide phase is formed. The exact stoichiometry of this phase depends on the metal, some examples are Pt_2Si , Ti_5Si_3 [22]. As the temperature is increased, several other intermediate phases may form before the final high temperature phase is reached. Most phase changes result in a stoichiometrically different materials, but there are exceptions, for example the C49-to-C52 change in TiSi_2 only involves the reordering of the crystal structure. The description here assumes that the system is close to the thermodynamic equilibrium as the temperature is increased, if the temperature is raised sufficiently fast that reaction kinetics become a limiting factor, the intermediate phases may form only on the boundary between silicon and metal or may not form at all.

The silicon-metal system will typically have a dominant diffuser, that is the diffusion of the metal atoms is typically quite different from the diffusion of silicon atoms. Which species is the dominant diffuser is of special interest in two dimensional problems such as the formation of CS devices. If silicon diffusion dominates voids may be injected at the silicide silicon boundary [23].

4.3.4 Silicide formation process

This section will discuss development of the low-barrier silicide process. The formation of low-barrier silicides is more difficult than the formation of traditional mid-gap silicides due to the extreme workfunction values of the metals involved. As mentioned before metals that have a high workfunction and form a low hole barrier are chemically unreactive, while metals which have a low workfunction and form a low electron barrier are very reactive.

4.3.4.1 Erbium silicide process

Erbium, Er, is a good candidate for an NMOS silicide, since it is known to have a relatively low resistivity[24] and form well on silicon. Previous studies of erbium silicide have shown that the high temperature phase is $\text{ErSi}_{1.7}$, with a hexagonal crystal structure[25]. The barrier height has been measured electronically and optically to be 0.28~0.32eV[26-27] for well-formed $\text{ErSi}_{1.7}$. Erbium silicide has a small, 1.4%, lattice mismatch to silicon[28].

4.3.4.1.1 Vacuum requirements and erbium evaporation

The introduction of oxygen into the film, in as little as 5% concentration, is known to increase the erbium silicide electron barrier to as much as 0.46eV[29]. This translates to a requirement of less than $1\text{E-}8$ torr of background oxygen pressure during a nominal deposition at $1\text{\AA}/\text{s}$. Such background pressures can only be achieved in a UHV system. The choice of deposition technique was forced by the fact that the only UHV system available was an evaporation chamber. While sputtering is commonly used in microelectronic fabrication it is more difficult to apply to materials research. This is because large targets are commonly required, and the concentration of impurities in the plasma during deposition is difficult to determine.

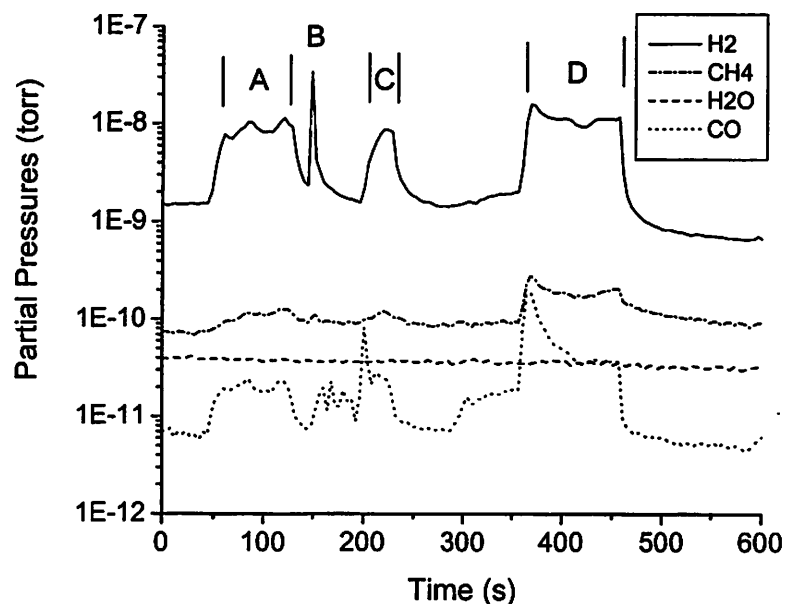
The UHV evaporation system, (known as 'imetal'), has a base pressure of $5\text{E-}10$ torr, and is equipped with an in-situ heater. It is therefore well suited for erbium evaporation and silicide anneal. The source metal is shipped in argon from the supplier (CERAC), in an airtight container. It is loaded into the evaporation system with the minimum air contact. Erbium should be evaporated from a glassy carbon, or coated carbon crucible. Glassy carbon is a densified form of amorphous carbon, and it does not react with erbium.

Reactive metals will necessarily be contaminated with impurities from purification and storage. Following pump down the source erbium has to be baked in order to remove as many volatile impurities as possible. Subsequent bakes starting at a low temperature ($\sim 300^{\circ}\text{C}$) and progressing to a high temperature at which a low evaporation rate is reached ($\sim 1100^{\circ}\text{C}$) must be performed on each new Er source. This process takes ~ 25 - 50 hours for 1cc of erbium; trying to bake out the source any faster will cause the vacuum system to overload and the gun to short. The volatile impurities are mostly molecular hydrogen ($\sim 97\%$), and traces of water, methane, and carbon monoxide.

Once the erbium source has been baked, titanium can be evaporated on the walls of the vacuum chamber. Titanium helps to decrease the base pressure, from $2\text{E}-9$ to $5\text{E}-10$ torr, and absorbs some of the hydrogen released during the evaporation of erbium. At UHV pressures the porous titanium film can take several days to saturate. During actual deposition, the shroud around the source is cooled to liquid nitrogen temperatures in order to condense a majority of the volatile substances, except hydrogen, that escape from the source. The vacuum is monitored by an ion gauge and a residual gas analyzer(RGA). An RGA plot is shown for a typical evaporation in Fig. 4.9.

Figure 4.9: Residual gas analyzer plot of the relevant gas pressures during the erbium deposition.

Evaporation details:
 A: Last pre-evap bake (dep rate $0.1\text{\AA}/\text{s}$)
 B: Hydrogen bubble burst from a crack in source during cooling
 C: Pressure test for actual evaporation (dep rate $1\text{\AA}/\text{s}$)
 D: Actual evaporation (dep rate $1\text{\AA}/\text{s}$)



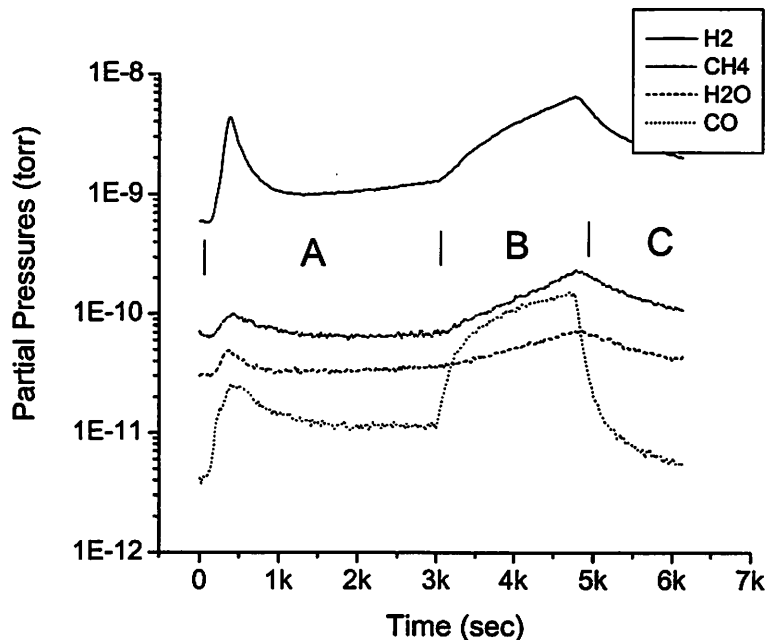
This plot omits a complex series of pre-evaporations done in order to bake the source. The erbium source needs $\sim 10\text{m}$ of bakeout for each day without an evaporation, indicating that it is absorbing impurities from the UHV system.

During the evaporation the sample can be heated to a certain temperature, T_d , with the substrate heater. The same heater is used to perform the silicidation anneal, at one or several temperatures. In Fig. 4.10 is a plot of the RGA during a typical silicidation

anneal:

Figure 4.10: Residual gas analyzer plot of the relevant gas pressures during the erbium silicide anneal.

Evaporation details:
 A: Metal rich phase formation anneal $\sim 300^\circ\text{C}$
 B: $\text{ErSi}_{1.7}$ formation anneal $\sim 425^\circ\text{C}$
 C: cooldown



It is important to keep the shroud cold during the silicidation anneal since warming of the shroud releases CO and CH₄ that can react with the forming silicide.

4.3.4.1.2 Erbium silicide thermal processing

The correct thermal processing is important in obtaining good quality silicide. The temperature of the substrate during deposition, T_d , is critical to the microstructure of the deposited metal layer, and the thermal steps used in the silicidation anneal are critical in determining the type of silicide formed.

There are three regions of T_d observed in this experiment. For low $T_d < 150^\circ\text{C}$ the microstructure of the metal film is porous, or has very small grain size. This results in high resistivity metal films that etch quickly. The porosity of the metal film seems to transfer to some degree to the silicide. Even for high silicide anneal temperature, films deposited in this temperature regime show wildly variable wet etch rates, variable resistivity, and a sensitivity to oxygen, all expected symptoms of highly porous films. As discussed previously, this type of microstructure corresponds to the first zone in the SZM. It is interesting to note that the boundary temperature between zone 1 and zone 2 is expected to be $0.3 \cdot T_m = 250^\circ\text{C}$, considerably higher than the 150°C observed. This difference might be caused by the fact the evaporation rate is very low, at $1 \text{ \AA}/\text{s}$, or the presence of hydrogen in the system.

At temperatures $150^\circ\text{C} < T_d < 225^\circ\text{C}$ the metal film does not exhibit chemical signs of porosity, and an etch resistant silicide forms easily. This is the preferred sample temperature range for erbium evaporation. At the higher temperatures in this range, $200 \sim 225^\circ\text{C}$ the metal film starts to get non-uniform. One explanation for the non-uniformity is the nucleation and subsequent growth of islands during the evaporation.

With a higher temperature, larger islands are expected, and when the islands get to a size comparable to the film thickness, the metal film will be non-uniform. The problem with this explanation is that the non-uniformity exhibits itself as widely (100nm~5 μ m) spaced bumps on the metal film. Simple nucleation and growth should result in much higher frequency non-uniformity. The responsible effect may be a combination of nucleation with initial silicide formation, or some other rare reaction, like the spotty atomic desorption of the hydrogen passivation prior to evaporation, and the subsequent growth of erbium islands on the defect spots.

As the temperature of the sample is increased to above 225°C, a faceted growth structure is observed. The spotty bumps transform themselves into a dense stochastically distributed population of pyramids of various sizes. The pyramids are aligned to the crystal orientation of the underlying silicon layer, and can grow to a width as large as 1000 times the film thickness. This film structure produces non-uniform silicide that is not useful for thin-body CS device applications. However the self-assembly process observed in forming the pyramids may be useful for more exotic structures, especially if it can be controlled through the selective removal of the hydrogen passivation layer. Notably the pyramids etch quickly in erbium metal etchant, suggesting that they are tall and do not form a significant amount of ErSi_{1.7} with the underlying silicon layer.

The second critical thermal step is the silicidation anneal. The experimental results indicate that erbium forms two phases of silicide. The metal rich phase, ErSi forms at a lower temperature ~300°C (Typical condition: 60m, 100Å metal thickness), and etches quickly in nitric acid, the wet etch used to remove erbium. The second phase is the stable form ErSi_{1.7}, which forms at ~420°C (Typical condition: 60m, 100Å metal

thickness). Due to this fact most of the silicidation anneals consisted of two 60m steps, one at $\sim 300^{\circ}\text{C}$, to form ErSi , and another at $\sim 420^{\circ}\text{C}$, to form $\text{ErSi}_{1.7}$.

The exact nature of the film after the silicidation anneal is difficult to determine, since it can consist of several non-uniform thin layers. Optical characterization is of little use since accurate models for erbium silicide do not exist. The resistivity and etch properties of the film are the most relevant method of characterizing the formed film. In Fig. 4.11 below is the resistivity data from typical set of anneals done at different temperature as a function of etch time. The etchant, dilute nitric acid at room temperature, etches metallic erbium, ErSi , but does not etch $\text{ErSi}_{1.7}$.

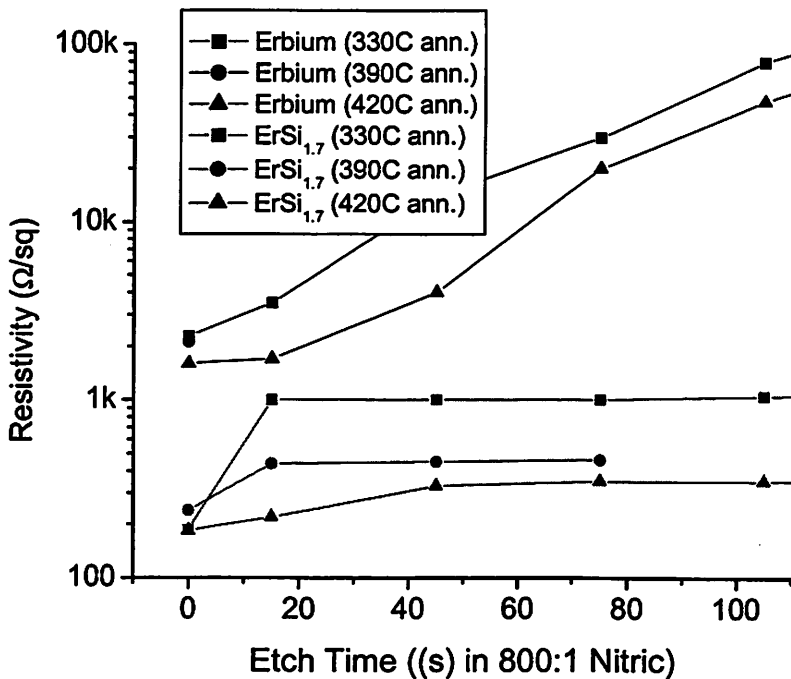


Figure 4.11: Resistances of the erbium and erbium silicide deposited and annealed at different temperatures and etched in nitric acid. Note that the initial 20s of etch significantly increase the resistivity of the silicide. The deposition thickness of the erbium was 75\AA on 100\AA of silicon.

The etch time of zero represents resistivity of the film right after silicidation anneal. Interestingly erbium silicide films have an order of magnitude lower resistance than erbium metal; the established bulk resistivity ratio is only 3. Most likely the

columnar microstructure of the metal and the difference in the thickness of the two films causes the discrepancy. From the Fig. 4.11 we can see that while the silicide does not exhibit a constant etch rate. It appears that some silicide is etched quickly in the first 20s of the etch, and the remainder of the silicide does not etch even upon a prolonged exposure to the etchant. The post etch resistivity is also a strong function of the anneal temperature, while the resistivity prior to the etch is constant in the silicide formation temperature range 320–450°C.

There are two effects that could be responsible for this behavior. At lower temperatures the silicide layer could consist of $\text{ErSi}_{1.7}$ underlayer and an ErSi overlayer, the ErSi would etch away during the etch explaining the uneven etch rate. A higher temperature formation anneal would increase the thickness of the $\text{ErSi}_{1.7}$ at the expense of ErSi , explaining the temperature trends. This model does not explain why the post etch resistivity tends to a value 35% higher than the resistivity prior to the etch when the temperature is increased significantly beyond the point where all the ErSi should be converted to $\text{ErSi}_{1.7}$. Even for anneal temperatures as high as 450°C the silicide loses 35% of its conductivity in the first few seconds of the wet etch. A second effect is probably responsible for this behavior. It is possible that the wet etch penetrates the columnar grain boundaries and increases the grain-to-grain resistivity.

4.3.4.1.3 Erbium wet etching

While the wet etching of erbium was discussed as an analysis technique in the previous section it is necessary to remove the unreacted metal after the silicide formation anneal. Dilute nitric acid was found to preferentially etch erbium and not $\text{ErSi}_{1.7}$, and a 3m 800:1 nitric acid etch was used to remove $\sim 75\text{\AA}$ erbium films in the silicidation

process. The porosity of the metal film has a large influence on the wet etch rate. Erbium films that were evaporated at $\sim 175^{\circ}\text{C}$ show $1\text{\AA}/\text{s}$ etch rate, while erbium films evaporated at 20°C have etch rates as high as $\sim 10\text{\AA}/\text{s}$, in 800:1 nitric acid. It is also interesting that the wet etch rate slows down as the wet etch progresses for samples evaporated at 175°C . This would seem to indicate that the wet etch preferentially attacks defects such as grain boundaries.

Warm water was also found to preferentially etch porous erbium, although it is possible that the metal film was just oxidized, since only a resistivity measurement was used to determine the etch rate. A water etch at an optimized temperature may show superior selectivity than dilute nitric acid.

4.3.4.1.4 Erbium thickness considerations

The control of the thickness of the evaporated erbium is critical to the thin-body silicidation process. If too little erbium is deposited it will not consume the entire silicon film, and the silicide film will be more resistive. If too much erbium is deposited the formation of ErSi will consume too much silicon, making the full conversion of ErSi to $\text{ErSi}_{1.7}$ impossible due to the lack of available silicon. Since ErSi etches in nitric acid it will be removed in the wet etch.

The plot in Fig. 4.12 shows the resistivity of a silicide film after wet etch as a function of starting silicon thickness:

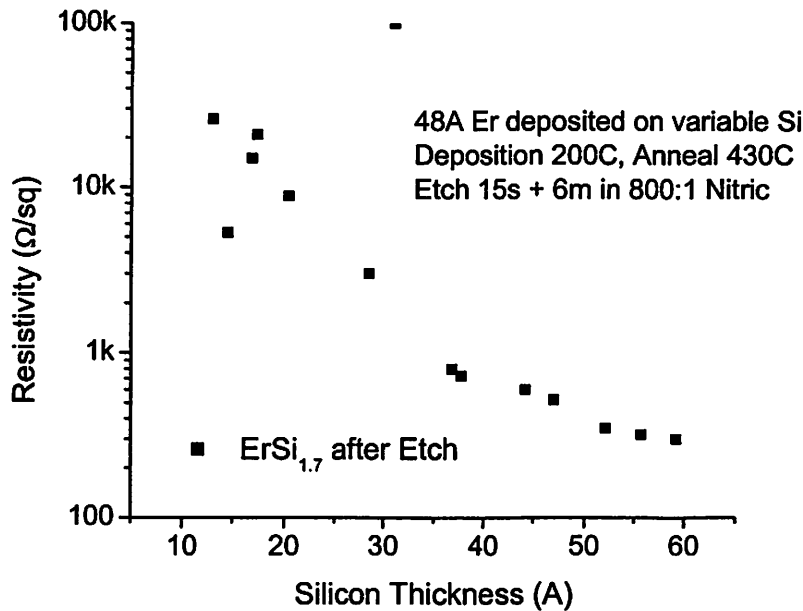


Figure 4.12: Resistances of the erbium silicide after anneal and etch. 48Å of erbium was deposited on a variable thickness of silicon. The resistance doesn't decrease after silicon thickness of 55Å is exceeded indicating an optimum Er:Si ration of 1.1:1.0.

According to this plot a ratio of Er:Si (1:1.1) is optimal. This is a larger thickness of erbium that would be expected from bulk atomic density comparison. The reason for the discrepancy is that the film measurement equipment was calibrated at $T_d=20^\circ\text{C}$. At such a temperature the metal film is highly porous and has a lower than expected density.

4.3.4.2 Platinum silicide process

Platinum silicide has been used in the microelectronics industry before, so little development was needed for this experiment. Platinum can be deposited either by evaporation or sputtering. Evaporation was used in this experiment due to the poor quality of the available sputtering system. The sputtering system was found to contain a sufficient quantity of oxygen radicals in the plasma to remove the hydrogen passivation and oxidize the surface prior to deposition. Platinum was evaporated with the substrate at 20°C , creating a porous film. Fortunately platinum is unreactive, so the increased surface area due to porosity probably does not increase the amount of impurities in the film. The

silicide anneal was done in a nitrogen purged oven at atmospheric pressure, it consisted of 250°C 1hour, 325°C 1h, 400°C 1h, steps. Platinum is a dominant diffuser during silicide formation, so it is likely that the porosity of the metal does not transform into porosity of the silicide. Following the formation anneal, the sample was annealed in air for 5min at 400°C, this step is necessary to form a thin silicon dioxide layer on top of the silicide. It turns out that this is the most critical step in platinum silicide formation, since the platinum wet etch, dilute aqua regia (3 HCl: 1 HNO₃: 4 H₂O) at 85°C 30s, etches platinum silicide as well as platinum. If no protective layer would exist on top of the silicide to protect it from the wet etch it would all be removed. Unfortunately, platinum also oxidizes, albeit very slowly, certainly more slowly than platinum silicide. More work may be necessary to optimize this protective oxidation step. The typical resistivities obtained from platinum silicide formation are ~80Ω/sq for 75Å Pt deposited and annealed on 100Å of silicon. After the etch, the silicide resistivity goes up, probably due to local etching through the oxide barrier. The silicide resistivity after the etch for the thicknesses given above is 100~170Ω/sq. Probably an optimization of the protective oxidation step would result in a consistent 100Ω/sq resistivity.

4.4 Devices at the end of processing

After the silicidation process was finished the devices were tested electrically; the data from those measurements is presented in the following chapter. Below is a SEM image of a finished NMOS UCS device with a 15nm gate-length.

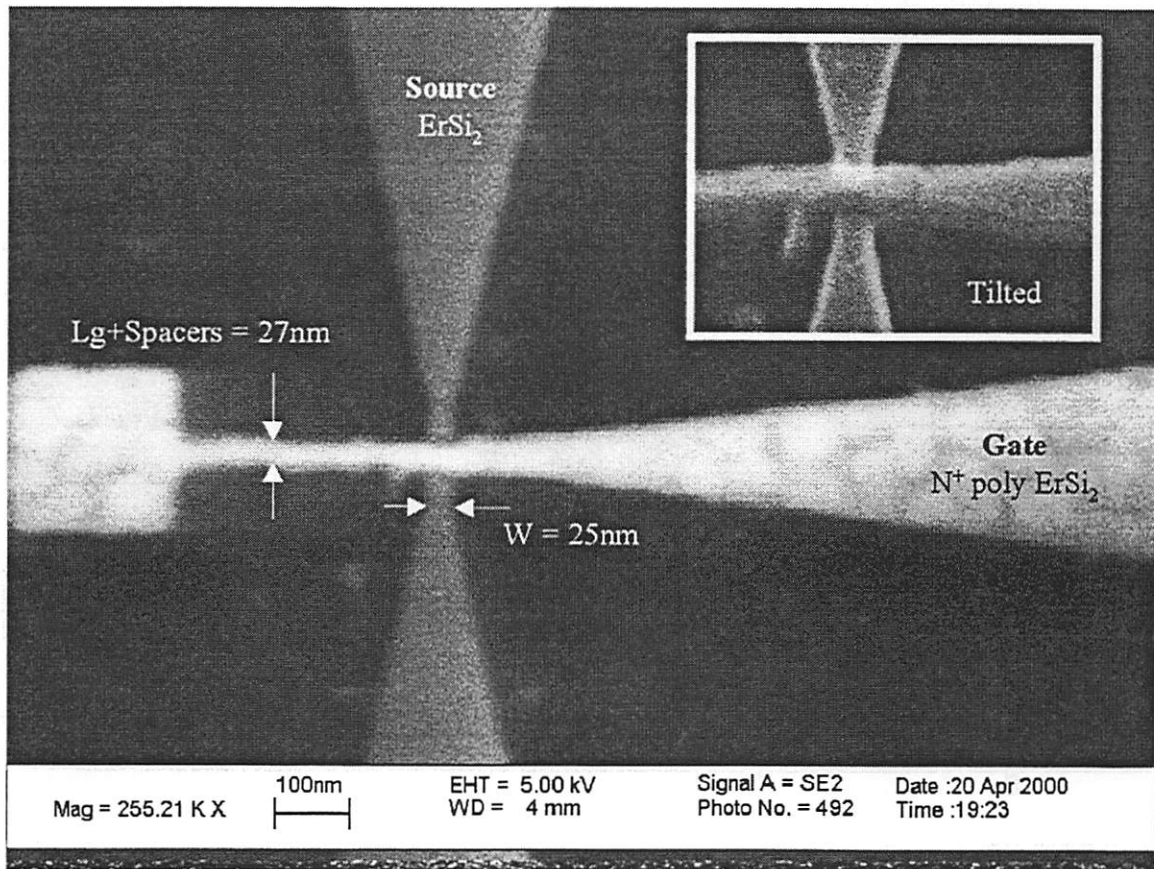


Figure 4.13: SEM of a finished NMOS CS thin-body device. No metalization past the silicide was performed on this lot.

The width of this device is 25nm. The silicide grains on top of the gate have a $\sim 60\text{nm}$ diameter, and they are likely a direct result of the grain structure of the underlying polysilicon. The silicide of the source and drain regions shows no visible grain formation. Either the grains are small or the formation of the source/drain silicide was

epitaxial. The large square to the left of the channel region is the gate anchor. This feature helps to give structural stability to the thin gate photoresist line during the gate definition steps. The gate dimension as seen on the figure is actually the gate length plus the width of both the spacers. The approximate spacer width of $2 \times 70 \text{ \AA}$ results in a calculated gate-length of 130 \AA . When reporting the gate-length this indirectly measured value was rounded up to 15 nm to give a conservative gate-length estimate.

References:

- [1] Auberton-Herve, A.J.; Bruel, M.; Aspar, B.; Maleville, C.; Moriceau, H. "Smart-Cut(R): the basic fabrication process for Unibond(R) SOI wafers," IEICE Transactions on Electronics, Vol. E80-C (no.3), p.358-63, March 1997.
- [2] Hoshino Y.; Toyonura, K.; Oho, H., "Precise measurement of resolution in SEM," Research Reports of Kogakuin University, (no. 80), p.67-70, April 1998
- [3] Broers, A.N.; Hoole, A.C.F.; Ryan, J.M., "Electron beam lithography-resolution limits," Microelectronic Engineering, vol.32 (no.1-4), p.131-42, Sept. 1996
- [4] Zandbergen, P.; Dijkstra, H.J. "Evaluation of a positive tone chemically amplified deep UV resist for E-beam applications," Microelectronic Engineering, Vol. 23 (no.1-4), p.299-302, Jan. 1994
- [5] Ludwig Reimer, Scanning Electron Microscopy, Bryant W. Rossiter and John F. Hamilton, eds., *Physical Methods of Chemistry Vol. IV, Microscopy*, (John Wiley & Sons, New York, 1991) 285-341
- [6] H. Bethe, *Ann. Phys. (Leipzig)*, Vol. 5, p. 325, (1930)
- [7] Kuan, S.W.J.; Frank, C.W.; YenLee, Y.H.; Eimori, T.; Allee, D.R.; Pease, R.F.W.; Browning, R. "Ultrathin poly(methylmethacrylate) resist films for microlithography," *Journal of Vacuum Science & Technology B*, Vol. 7 (no.6), p.1745-50, Nov.-Dec. 1989
- [8] Lutwyche, M.I. "The resolution of electron beam lithography," *Microelectronic Engineering*, vol. 17 (no.1-4), p.17-20, March 1992
- [9] Hintermaier, M.; Anzinger, H.; Knapek, E. "CAMP-6, a deep-UV positive tone resist applied to e-beam exposure," *Microelectronic Engineering*, vol.23, (no.1-4), p. 295-298, Jan. 1994
- [10] Langheinrich, W.; Beneking, H. "The resolution of the inorganic electron beam resist LiF(AlF/sub 3/)," *Microelectronic Engineering*, vol.23, (no.1-4), p. 287-290, Jan. 1994.
- [11] Fujita, J.; Ohnishi, Y.; Ochiai, Y.; Nomura, E.; Matsui, S. "Nanometer-scale resolution of calixarene negative resist in electron beam lithography," *Journal of Vacuum Science & Technology B*, Vol.14 (no.6), p. 4272-4276, Nov.-Dec. 1996
- [12] Rishton, S.A.; Beaumont, S.P.; Wilkinson, C.D.W. "Measurement of the profile of finely focused electron beams in a scanning electron microscope," *Journal of Physics E (Scientific Instruments)*, vol. 17 (no.4), p.296-303, April 1984

- [13] Tanenbaum, D.M.; Lo, C.W.; Isaacson, M.; Craighead, H.G.; Rooks, M.J.; Lee, K.Y.; Huang, W.S.; Chang, T.H.P. "High resolution electron beam lithography using ZEP-520 and KRS resists at low voltage," *Journal of Vacuum Science & Technology B*, vol.14 (no.6), p.3829-33, Nov.-Dec. 1996
- [14] Broers, A.N.; Cuomo, J.; Harper, J.; Molzen, W.; Laibowitz, R.; Pomerants, M. (Edited by: Sturgess, J.M.) "High resolution electron beam fabrication using STEM," *Electron Microscopy 1978*, p.343-54, 1978
- [15] Fujita, J.; Ohnishi, Y.; Manako, S.; Ochiai, Y.; Nomura, E.; Sakamoto, T.; Matsui, S. "Calixarene electron beam resist for nano-lithography," *Japanese Journal of Applied Physics, Part 1*, Vol.36 (no.12B), p.7769-72, Dec. 1997
- [16] Van Delft, F.C.M.J.M.; Weterings, J.P.; van Langen-Suurling, A.K.; Romijn, H. "Hydrogen silsesquioxane/novolak bilayer resist for high aspect ratio nanoscale electron-beam lithography," *Journal of Vacuum Science & Technology B*, vol.18 (no.6), p.3419-23, Nov 2000
- [17] Muray, L.P.; Anderson, E.H.; Boegli, V. "Design and implementation of a real-time hierarchical parallel postprocessor for 100 keV electron beam lithography," *Journal of Vacuum Science & Technology B*, Vol.15 (no.6), p. 2204-8, Nov.-Dec. 1997
- [18] Masahito Kodama, Hirofumi Funabashi, Yasuichi Mitsushima, and Yasunori Taga, "A solid-phase epitaxy with clean surface formation using SiH₄ and evaluation of SOI layer," *Electronics and Communications in Japan, Part 2*, Vol. 79 (no. 12), p. 71-77, 1996
- [19] Richard C. Jeager, "Introduction to microelectronic fabrication," (Addison-Wesley Publishing: Reading, MA, 1988), p. 109
- [20] F. C. Matarotta, G. Ottaviani "Science and technology of thin film deposition," (World Scientific Publishing: Singapore, 1995)
- [21] Barna, P.B.; Adamik, M. "Fundamental structure forming phenomena of polycrystalline films and the structure zone models," *Thin Solid Films*, vol. 317 (no.1-2), p.27-33, 1 April 1998
- [22] Canali, C.; Catellani, F.; Ottaviani, G.; Prudenziati, M. "On the formation of Ni and Pt silicide first phase: The dominant role of reaction kinetics," *Applied Physics Letters*, Vol.33 (no.2), p. 187-90, 15 July 1978
- [23] S. P. Murarka, "Silicides for VLSI applications," (Academic Press, New York, 1983)
- [24] Travlos, A.; Salamouras, N.; Flouda, E. "Epitaxial erbium silicide films on (100) silicon: growth, structure and electrical properties," *Applied Surface Science*, Vol. 120 (no.3-4), p.355-64, Dec. 1997

- [25] K. A. Gshneider, L.-R. Eyring (Eds), "Handbook on the physics and chemistry of rare earths," (North-Holland Publishing Company, 1978), Vol. 2; Alloy and Intermetallics
- [26] M. H. Unewisse, and J. W. V. Storey, "Conduction mechanisms in erbium silicide Schottky diodes," *Journal of Applied Physics*, Vol. 73 (no. 8), p. 3873-3879, 15 April 1993
- [27] S. Vandre, T. Kalka, C. Preinesberger, and M. Dahne-Prietsch, "Flatband conditions observed for Lanthanide-silicide monolayers on n-type Si(111)," *Physical Review Letters*, Vol. 82 (no.9), p. 1927-1930, 1 March 1999
- [28] Mantl, S., "Materials aspects of ion beam synthesis of epitaxial silicides," *Nuclear Instruments & Methods in Physics Research, Section B*, vol. B84 (no.2), p. 127-34, Feb. 1994
- [29] P. Muret, T. A. Nguyen Tan, N. Frangis, J. Van Landuyt, "Unpinning of the Fermi level at erbium silicide/silicon interfaces," *Physical Review B*, Vol. 56 (no. 15), p. 9286-9289, 15 October 1997

5.0 Thin-body complementary silicide device results and modeling

This chapter will present the electronic device performance of the thin-body CS devices fabricated in this experiment. A transmission model of device behavior will be constructed and fit to the data. A second point-to-point tunneling model will be used as a part of a 2D simulator to examine the design space of the CS devices.

5.1 UCS CMOS device performance

UCS devices fabricated in the manner described in chapter 4 were tested at room temperature shortly after fabrication. Device yield was approximately 30-50% for successfully completed fields. By far the most common failure mechanism was a source/drain open circuit; some source/gate shorts were also seen. It is likely that the first failure mode is a result of stress fractures in the silicide or particle contamination, the second mode is caused by local wet overetching of the spacer prior to metal deposition.

Fig. 5.1 show the I_d - V_d characteristics of the minimum gate-length transistors, and Fig. 5.2 shows their I_d - V_g characteristics[1].

Figure 5.1: I_d - V_d plot for the minimum gate-length CS devices. Two different devices, the erbium silicide NMOS, and the platinum silicide PMOS, are shown on this plot.

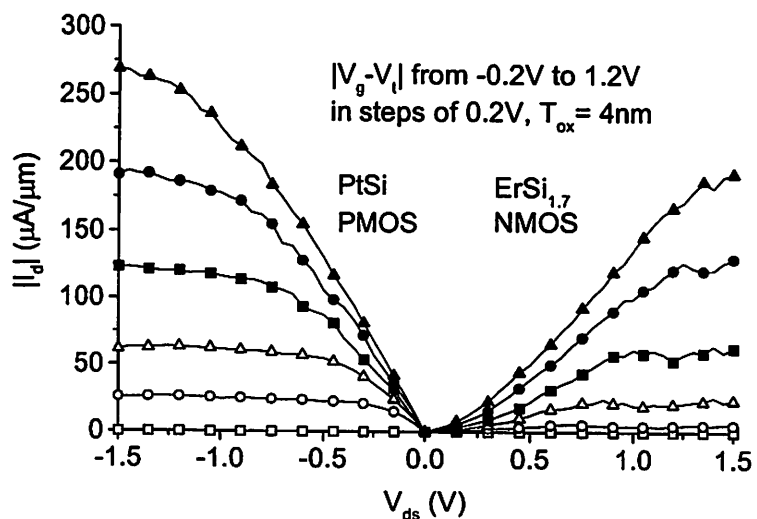
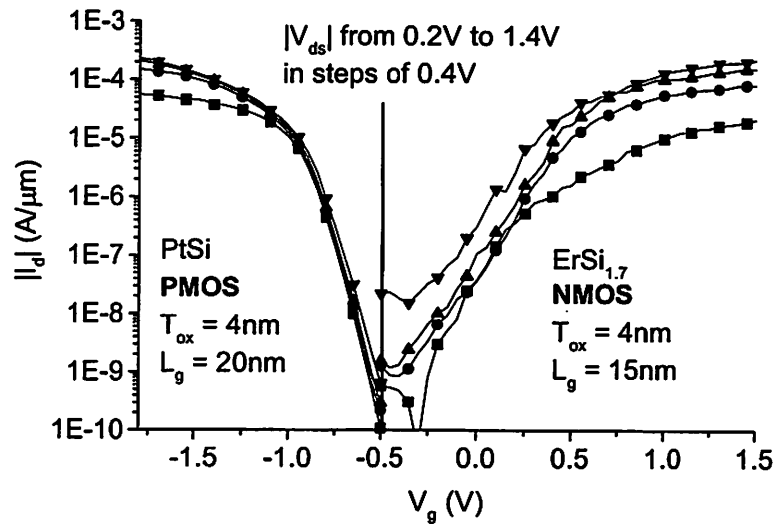


Figure 5.2: I_d - V_g plot for the minimum gate-length CS devices. Two different devices, the erbium silicide NMOS, and the platinum silicide PMOS, are shown on this plot.



The PtSi PMOS transistor has a 20nm gate length, and $|I_{dsat}|=270\mu A/\mu m$ with $T_{ox}=40\text{\AA}$ at $|V_{ds}|=1.5V$. Sub-threshold characteristics show excellent short-channel effects, and a V_t of $-0.7V$. The abnormally high V_t is due to the N^+ poly gate, it can be reduced using a mid-gap gate material such as P^+SiGe [2]. The $ErSi_{1.7}$ NMOS transistor has a 15nm gate length, $|I_{dsat}|=190\mu A/\mu m$ at $|V_{ds}|=1.5V$. The lower NMOS I_{dsat} is due to the higher $ErSi_{1.7}$ barrier of 0.28V. Still, this is the lowest flat-band NMOS barrier achieved for a silicide. NMOS sub-threshold characteristics show a swing of 150mV/dec, and a V_t of $-0.1V$, which can be adjusted to a higher value with the use of a mid-gap gate. The presence of the Schottky barrier can be seen in the exponential behavior of the I_d - V_d plot for the NMOS devices at low V_{ds} . The minimum gate-length devices shown in this section are also very narrow, between 20~40nm. The narrow width helps to control the short channel effect due to the side fringing fields. Also narrow devices have a higher effective channel resistance, and are less degraded by the significant parasitic resistance of the source-to-pad silicide.

Figure 5.3: Linear and saturation threshold voltages as a function of gate length for both NMOS and PMOS.

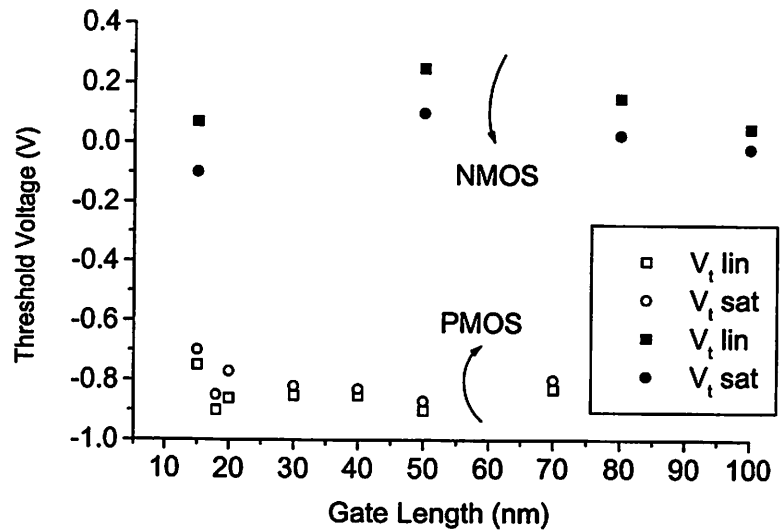
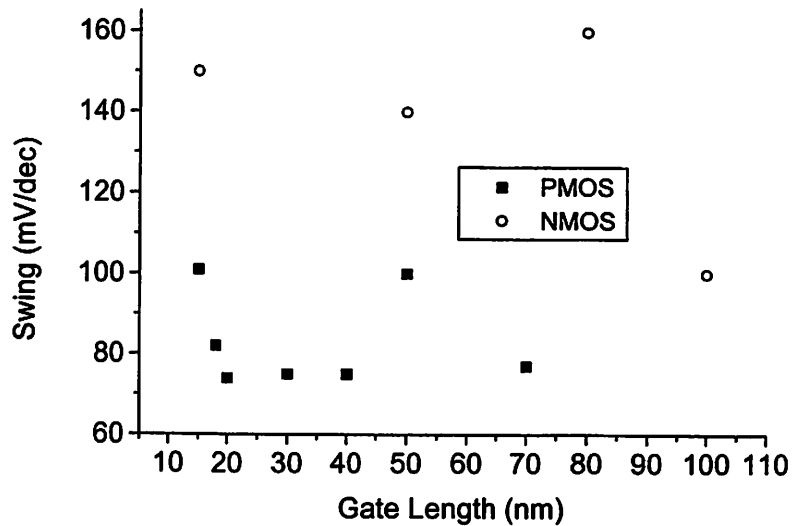


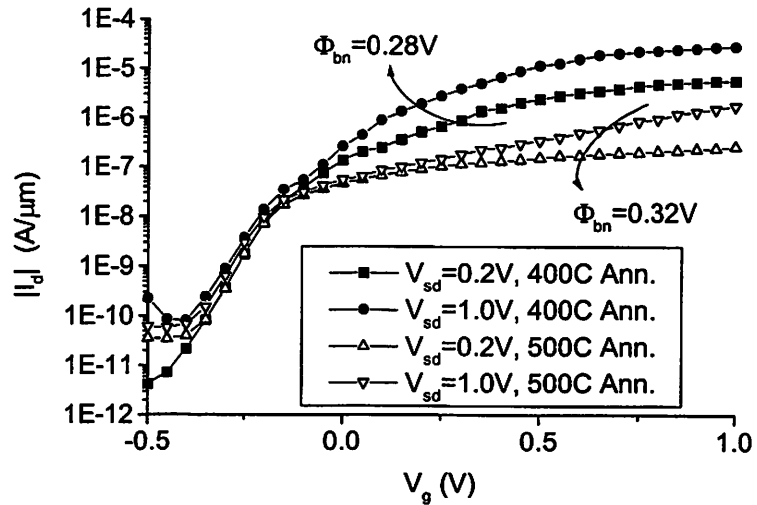
Figure 5.4: Swing dependence on gate length. For PMOS devices $S=75\text{mV/dec}$ down to 20nm gate lengths. NMOS devices show a larger variation in swing with



The gate-length dependencies of the threshold voltages and swing characteristics are shown in Figs 5.3 and 5.4 respectively. PMOS devices show excellent short channel characteristics down to 15nm gate-length, with $\Delta V_t = 0.2\text{V}$ and $S=100\text{mV/dec}$. NMOS devices show a similar $\Delta V_t = 0.2\text{V}$, but worse DIBL and $S=150\text{mV/dec}$. Since the swing and DIBL don't show a strong dependence on gate-length, they are probably determined

by the silicon/silicide interface trap density. Annealing devices at 500°C decreases interface trap density but increases the barrier to 0.32V, Fig. 5.5.

Figure 5.5: I_d - V_g of a long channel ErSi_{1.7} NMOS device before and after a 500°C anneal. Silicidation temperature is 400°C.



Long channel V_t values are $-0.85V$ for PMOS, and $0.05V$ for NMOS. These V_t values are similar to what would be expected for conventional thin-body transistors with N^+ poly gate, indicating that the low Schottky barrier doesn't inhibit current flow at threshold. The threshold data also shows that with the use of a single gate of appropriate mid-gap work-function the $|V_t|$ can be adjusted to $0.45V$ for both NMOS and PMOS. While this value may be acceptable for low-power applications, it is too high for high-performance logic.

5.2 Transmission model

Ballistic short channel devices, in which channel carrier scattering does not play a significant role, can be modeled by assuming that any carriers that are injected into the channel from the source are quickly swept into the drain. This approach is often called the transmission model[3,4] since the device current consists of the fraction of incident

carriers that are transmitted by the source potential barrier. The transmission approach to modeling MOSFETs is ideal for short channel thin-body CS devices.

The method for constructing a transmission model in a thin-body CS device consists of first assuming a certain electron distribution in the source, and then assuming a certain source barrier that blocks the carriers from entering the channel. Therefore the behavior of the source barrier as a function of device parameters and biases is a crucial part of the transmission model.

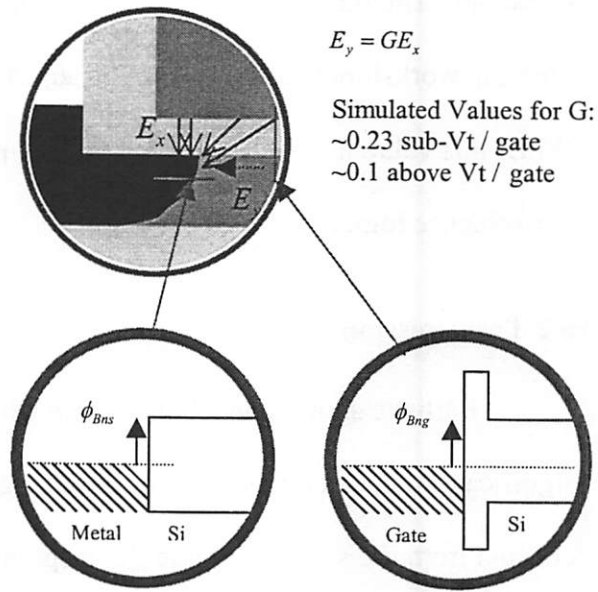
The exact shape of the barrier at the source is a strong function of the lateral electric field E_y . In an undoped thin-body device the channel potential is directly tied to the gate potential, until a significant amount of inversion charge is present. This fact can be used to make the assumption that the change in E_y is proportional to a change in E_x , the vertical field between the source and the gate, through a constant G . Translating the electric fields into voltages and considering the relevant flat-band conditions gives the following relations, Fig. 5.6:

$$V_g^* = V_g - (\phi_{Bng} - \phi_{Bns}) = V_g - V_{sbf\bar{b}}$$

$$E_x = V_g^* / t_{ox} \quad E_y = GE_x$$

$$E_y = \frac{GV_g^*}{t_{ox}}$$

Figure 5.6: Definition of relevant barrier heights, electric fields, and voltages in the derivation of the lateral electric field dependence on gate voltage. Band diagrams show the source-body and gate-body flat-band conditions.



As can be seen from the relations E_y can be positive or negative. The gate voltage at which it is zero is the source-body flat band voltage, V_{sfb} . The following discussion will assume NMOS devices for simplicity, but similar arguments can be made for PMOS. The V_{sfb} can be defined as:

$$V_{sfb} = (\phi_{Bng} - \phi_{Bns})$$

With Φ_{bng} and Φ_{bns} being the gate-to-body and source-to-body electron barriers respectively. Carriers at the source see different barriers with the potential above and below the V_{sfb} . Fig. 5.7 shows the band diagram for an NMOS device in both conditions. With the gate potential above V_{sfb} , electrons are blocked by the Schottky barrier; with the gate potential below V_{sfb} , the barrier is formed by the conduction band. V_g^* is just the gate voltage modified by V_{sfb} , Fig 5.6.

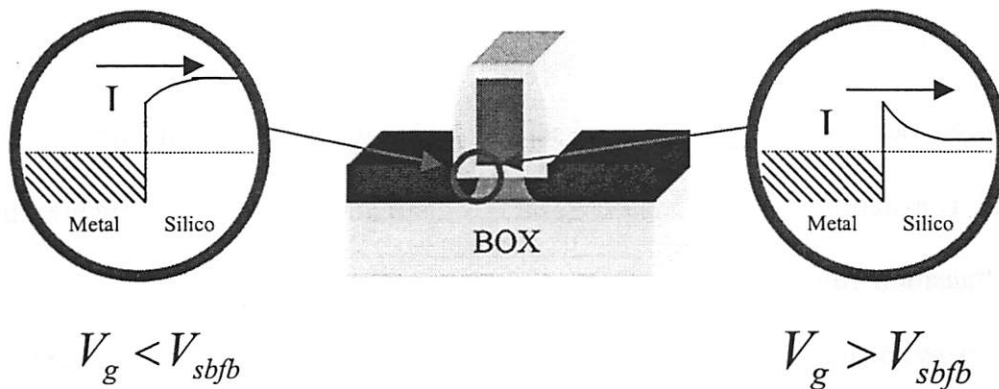
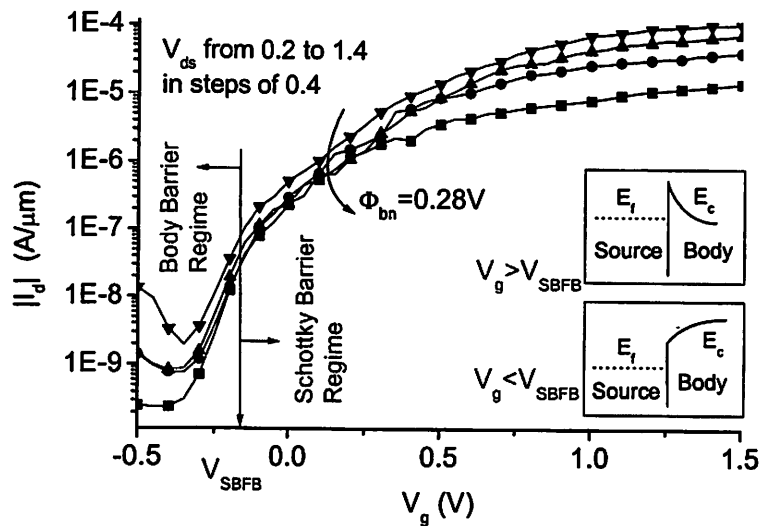


Figure 5.7: Band diagram of source with gate bias below and above the source-body flat-band voltage.

Since the conduction band potential is directly tied to the gate potential, for $V_g < V_{sfb}$ the long channel device swing is expected to be close to the ideal 60mV/decade at 300K.

This is due to the fact that at 300K a barrier lowering of 60mV produces an order of magnitude more transmission current assuming no tunneling and Boltzmann particle statistics. For $V_g > V_{sfb}$ the swing will be different than 60mV/dec, the exact value will depend on how much the Schottky barrier is lowered per increase in gate voltage. In theory this value could be smaller or larger than 60mV/dec, but for single gate thin-body UCS structures with 15-40Å T_{ox} it is larger, somewhere ~100-300mV/dec. The transition

Figure 5.8: A 100nm NMOS device showing the two sub-threshold barrier mechanisms. Current barrier is controlled by the silicon body potential below the V_{sfb} , and by the Schottky barrier above V_{sfb} .



between these swing regimes can be seen clearly in the longer channel devices, when DIBL is insignificant. Fig. 5.8 shows an I_d - V_g plot and the two swing regimes, including the transition voltage, V_{sfb} .

The experimental V_{sfb} value of -0.24V is close to the expected value of -0.28V.

5.2.1 Schottky barrier lowering model

In order to obtain a quantitative model for the barrier at $V_g > V_{sfb}$ it is important to derive the dependence of the Schottky barrier on E_y , and ultimately V_g . The two important barrier lowering mechanisms are image charge, and tunneling. Although

technically tunneling is not a barrier lowering mechanism it can be modeled as such, see section 2.1.2. The relations and diagrams for the two mechanisms are shown in Fig. 2.3-

2.4. Combining these relations with the dependence of E_y on V_g gives:

$$\Delta\phi_i = \left(\frac{e}{4\pi\epsilon_s t_{ox}} \right)^{1/2} (GV_g^*)^{1/2} \quad \Delta\phi_t = \frac{1}{e} \left(\frac{3e\hbar(\ln 2)}{4\sqrt{2m^* t_{ox}}} \right)^{2/3} (GV_g^*)^{2/3}$$

Where $\Delta\phi_i$ and $\Delta\phi_t$ are the barrier lowering terms for image charge and tunneling respectively. In order to obtain the total barrier to current flow, Φ_{Bn} (in case of NMOS), as a function of V_g it is necessary to subtract these two barrier-lowering terms from the flatband barrier height, Φ_{B0} . Fig. 5.9 shows the relevant voltages and the electron barrier, Φ_{Bn} , as a function of V_g .

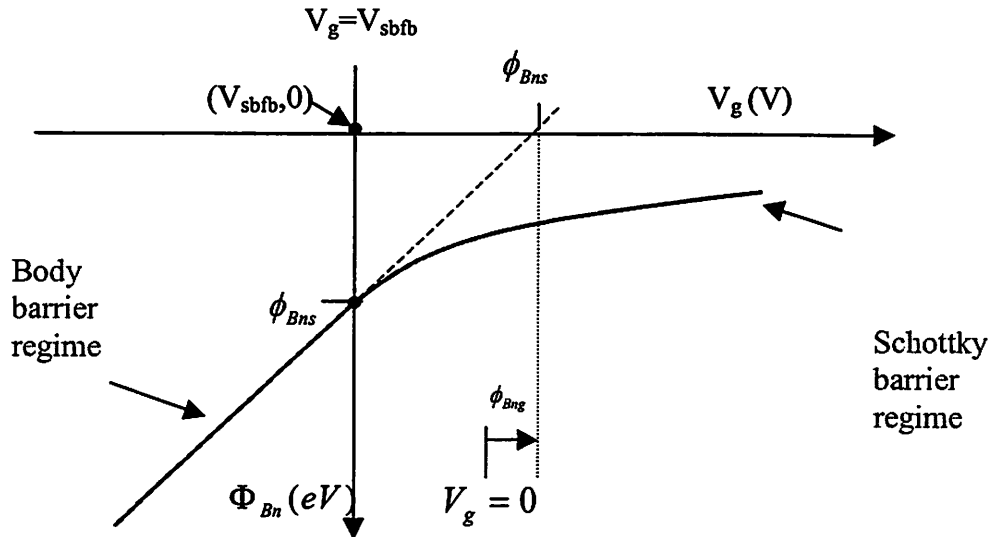


Figure 5.9: The electron barrier Φ_{Bn} , on y-axis, as a function of V_g , on x-axis. Barrier values increase toward the bottom of the page. The origin of this plot occurs at $\Phi_B=0\text{eV}$, $V_g=V_{\text{sfb}}$. To the right of the Y-axis the Schottky barrier is larger than the body barrier, the converse is true on the left.

It is important to note that the minority carrier can also be emitted over its respective barrier. Minority holes flow from drain to source, and must overcome a significantly larger barrier Φ_{bp} , however this barrier is lowered by the gate-to-drain

voltage. This voltage can be significant when V_g is low and V_d is high. The addition of the hole barrier gives the following Fig. 5.10.

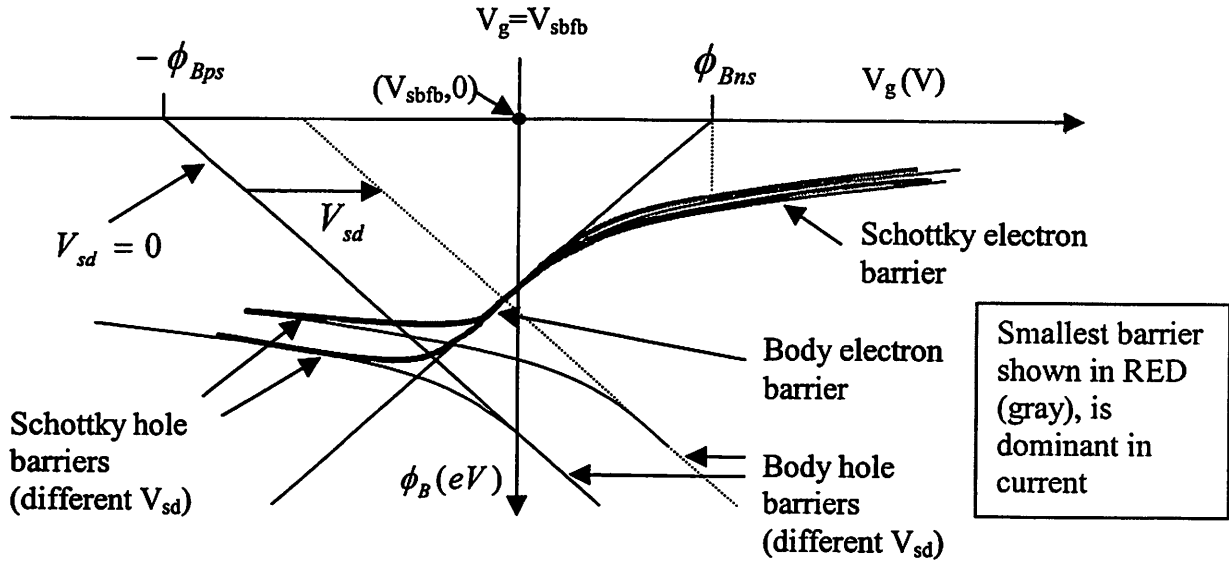


Figure 5.10: Both hole and electron barriers on y-axis, as a function of V_g . Since the barriers are responsible for current they must be 'added' only after conversion to current. However since the dependence of current on barrier height is exponential it is a good assumption to take the smallest barrier and treat it as dominant.

Now the barrier height must be converted to a drain current. This conversion can be done assuming that the current will be limited at the ballistic current limit, I_{dBAL} , when the barrier height is zero. The ballistic current values are 3.5mA/um for NMOS, and 1.5mA/um for PMOS with 1V V_{ds} [5]. To obtain I_d for a certain barrier height we use the assumption that the transmission current falls off by one decade for each 60mV of additional barrier height, this assumption is equivalent to assuming a Boltzmann electron distribution. Fig. 5.11 shows the full transmission model construction for both PMOS and NMOS using realistic fitting parameters G , Φ_{b0n} , and Φ_{b0p} . Minority carrier currents were neglected for simplicity.

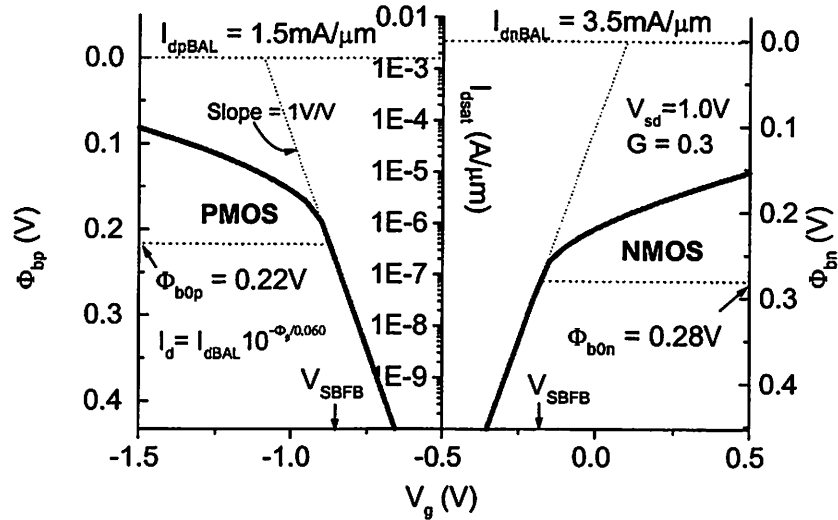


Figure 5.11: Construction of transmission model. Model includes barrier contributions from the silicon body, Φ_{b0} , $\Delta\Phi_i$, and $\Delta\Phi_t$. Current is derived from the barrier assuming a transmission probability of 1.

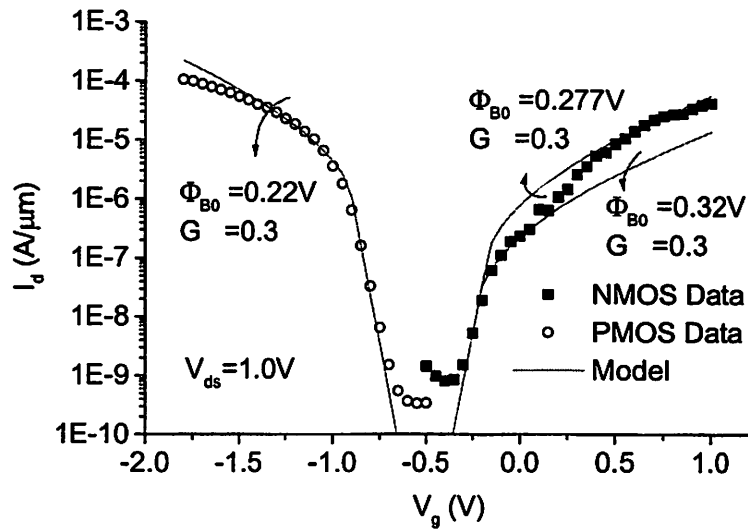


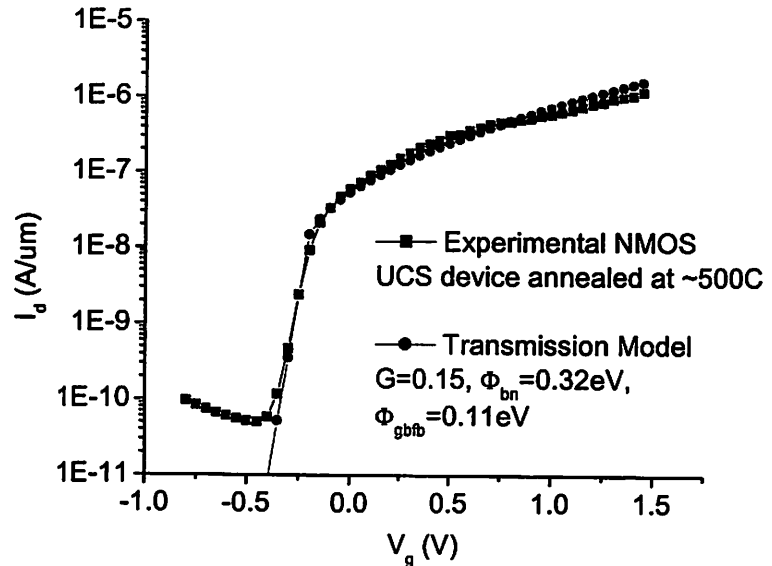
Figure 5.12: Barrier model fitting of 30nm wide PMOS and NMOS transistors. The G factor drops in strong inversion due to screening by the channel charge, leading to a deviation from the model at high gate biases.

5.2.2 Transmission model fitting

Fig. 5.12 shows the transmission model fit to the data for both the NMOS and PMOS devices. In PMOS devices $\Phi_{b0p}=0.22V$, slightly lower than the expected 0.24V; in NMOS devices $\Phi_{b0n}=0.28V$, the reported value for ErSi_{1.7}[6]. Fig. 5.12 also indicates that for ErSi_{1.7}, Φ_{b0n} changes from around 0.32V to 0.28V as the electric field at the

interface is increased. This suggests the presence of donor states, or donor-like interface traps, at the silicide interface. With the states occupied by an electron the $\Phi_{b0n} = 0.32V$, and as the electrons are removed by the gate field Φ_{b0n} changes to 0.28V. This picture is consistent with the fact that Φ_{b0p} for $ErSi_{1.7}$ is 0.80V[7], indicating that the true Φ_{b0n} should be $E_g - \Phi_{b0p} = 0.32V$. At 500°C these interface states are annealed out and the Φ_{b0n} becomes 0.32V, Fig. 5.5. The presence of interface states is the likely cause of the higher variability in the NMOS device current, as well as the higher NMOS swing and DIBL values. The annealed $ErSi_{1.7}$ device, due to its lower density of traps, and lower current, also fits the model with a higher degree of precision. Fig. 5.13 shows the annealed erbium device data overlaid on top of a transmission model fit.

Figure 5.13: NMOS device characteristics and fit to transmission model, after a 500C anneal. Better fit with model is obtained probably due to a decrease in interface charge, barrier height is increased.

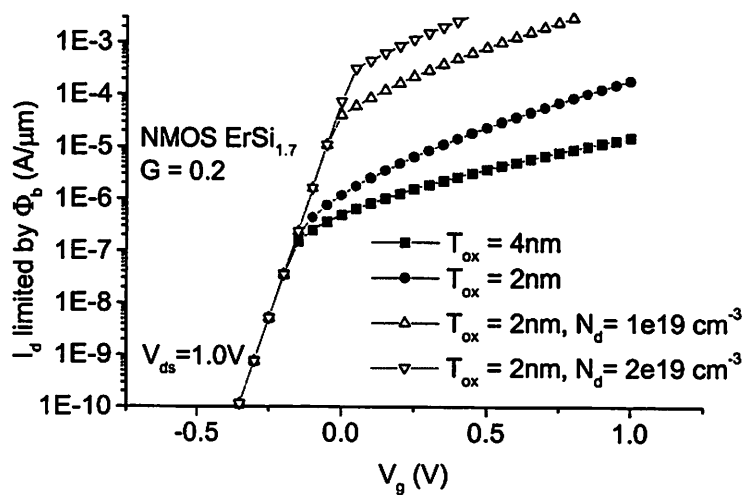


5.2.3 Methods to improve UCS device performance

The I_{dsat} of the thin-body UCS devices needs to be improved to meet ITRS specifications, especially for the NMOSFET. The methods for improving device performance are: use a silicide with a lower barrier, reduce the oxide thickness, or add an

extension doping to increase the e-field at the silicide boundary. Utilizing the transmission model described in the previous section, Fig. 5.14 shows the effect of oxide scaling and the use of modest source/drain extension doping on the Schottky limit for I_{dsat} in NMOSFETs, assuming $G=0.2$. Oxide scaling improves I_{dsat} for $V_g > V_{sfb}$, since it increases the electric field at the source/body interface. Adding an extension doping in the silicon drastically improves I_{dsat} by providing a depletion layer at the source/body junction with an electric field largely independent of gate voltage. The disadvantage of this approach is that it requires high temperature annealing, while silicide source/drains without doping can be made below 400°C . The DCS devices are expected to show worse short channel effects, since the Schottky barrier will not be present to limit leakage current. Fig. 5.15 shows the extension-doping level required to reach ITRS roadmap I_{dsat} specifications, with T_{ox} scaled to 2nm, as predicted by the transmission model. Devices with the modest extension doping concentrations of $3E19\text{cm}^{-3}$ are projected to reach ballistic performance.

Figure 5.14: Investigation of the effect of oxide thickness scaling and extension doping on transmission current limit for NMOS. A conservative value of 0.2 is used for G . Reducing the oxide thickness to 20\AA increases current for large V_g . A 40\AA long extension doping (N-type) reduces the barrier height independently of V_g .



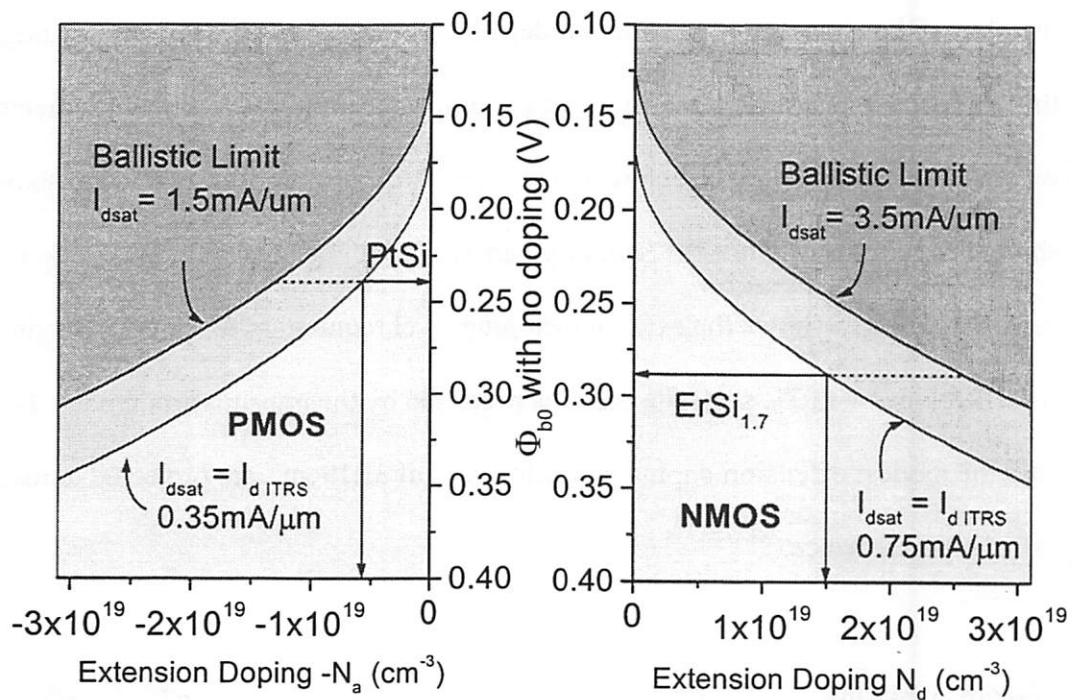


Figure 5.15: Contour plot of saturation current levels with $V_{dd}=1.0\text{V}$, for thin-body silicide source/drain transistors as a function of extension doping and flat-band barrier height. Strong screening is assumed, with $G=0.1$. Oxide thickness is 2nm. Ballistic current limits for both NMOS and PMOS can be reached with doping levels below $3 \times 10^{19} \text{ cm}^{-3}$.

5.3 TiSi₂ source/drain devices

Non-complementary silicide source/drain devices using a mid-gap silicide, TiSi₂, were also fabricated in this experiment for comparison[8]. The standard titanium silicide process was used to form source/drains. Curiously, since TiSi₂ has a similar hole and electron barrier, these devices function in NMOS and PMOS mode equally well. The only device feature that breaks the carrier symmetry is the N-poly gate, which makes the NMOS V_t lower than the PMOS V_t . Fig. 5.16a,b shows a 100nm gate-length TiSi₂ device biased in both NMOS and PMOS modes. The performance is poor for both cases since the barriers are high, $\Phi_{bn}=0.6\text{eV}$, $\Phi_{bp}=0.55\text{eV}$ [9].

The NP ambiguity of these devices has some interesting effects. For example when the gate voltage decreases the electron current drops, however if it decreases sufficiently far the hole current starts to increase. This ON-OFF-ON behavior is shown in Fig. 5.16c, in the I_d - V_g plot of a narrow TiSi₂ device.

Figure 5.16a: Characteristics of a single TiSi_2 device, $W=L=100\text{nm}$, device can be biased in N-mode. Figures a,b,c show the same device with different biases.

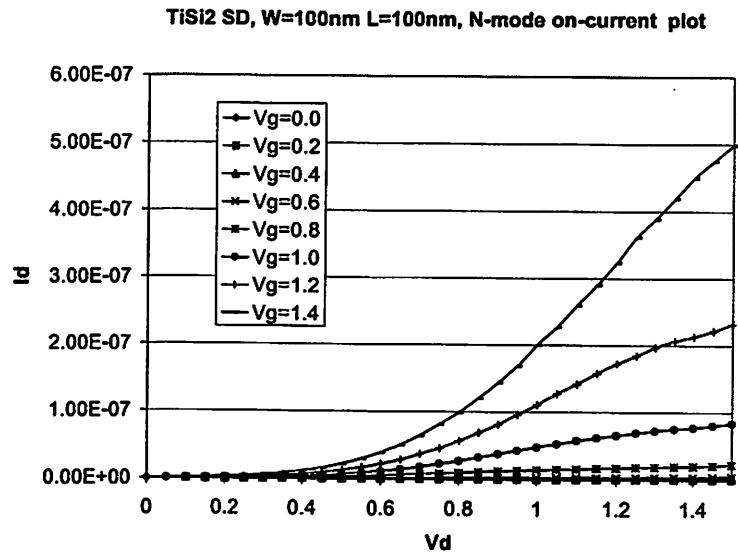


Figure 5.16b: Characteristics of a single TiSi_2 device, $W=L=100\text{nm}$, device can be biased in P-mode. Figures a,b,c show the same device under different bias conditions.

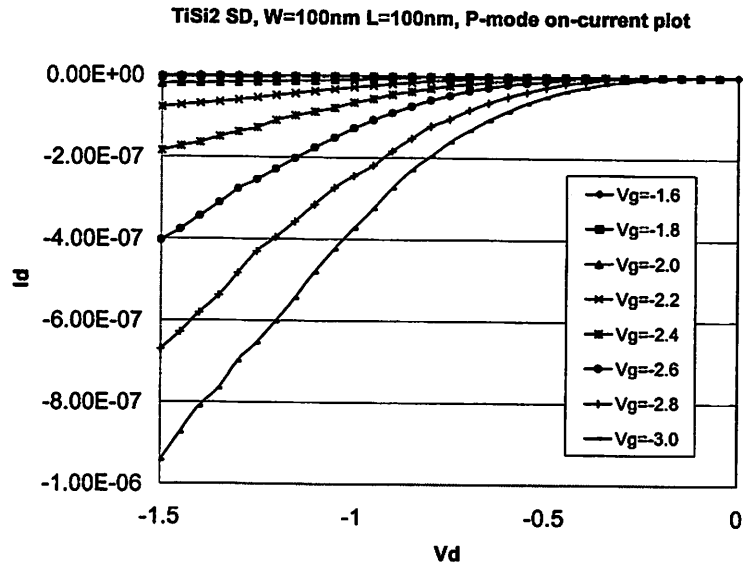
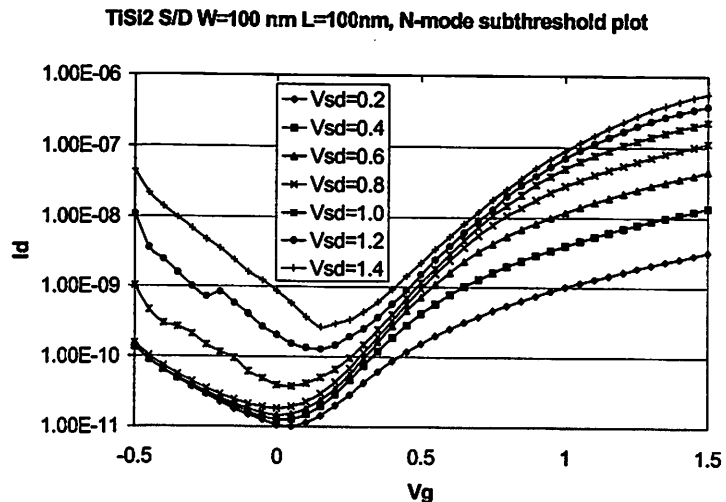


Figure 5.16c: I_d - V_g characteristics of a single TiSi_2 device, $W=L=100\text{nm}$. NMOS biasing, although PMOS parasitic current is visible for $V_g < 0$. Figures a,b,c show the same device under different bias conditions.



5.4 Fielday2D simulation of thin-body CS devices

The transmission model presented in the previous sections is too simple to be used for sophisticated device simulation and design. Traditionally device design is performed on multidimensional finite-state simulation tools that model many aspects of device behavior for each grid point. Unfortunately most such tools do not treat the current transport through a Schottky barrier with a model what can be extended to the silicide source/drain devices.

One exception is the Fielday2D simulator which has a realistic point-to-point tunneling model[10]. The model was upgraded by MeiKei Jeong(IBM SRDC) and myself to include barrier lowering due to image charge induction, and the model fitting parameters were fit to the data obtained for NMOS UCS devices. It is important to note that the considerable series resistance due to silicide leads and pad contact resistance was not corrected for, therefore the simulation most likely underestimates the current that can be obtained from these devices.

The 2D analysis performed on Fielday2D focused on NMOS erbium devices because the erbium's electron barrier is higher than platinum's hole barrier. Scaled relative to their respective CMOS references, the PMOS CS device will show larger I_{on} than the NMOS device, due to the Φ_{b0} difference. Simulating the NMOS devices examines the weaker of the two cases.

5.4.1 Undoped CS device simulation

The dimensions of the UCS structure simulated in Fielday2D are shown in Fig. 5.17. The variables of interest in this structure are the flat band barrier height, Φ_{b0} , the gate oxide thickness, T_{ox} , and the silicon body thickness, T_{si} .

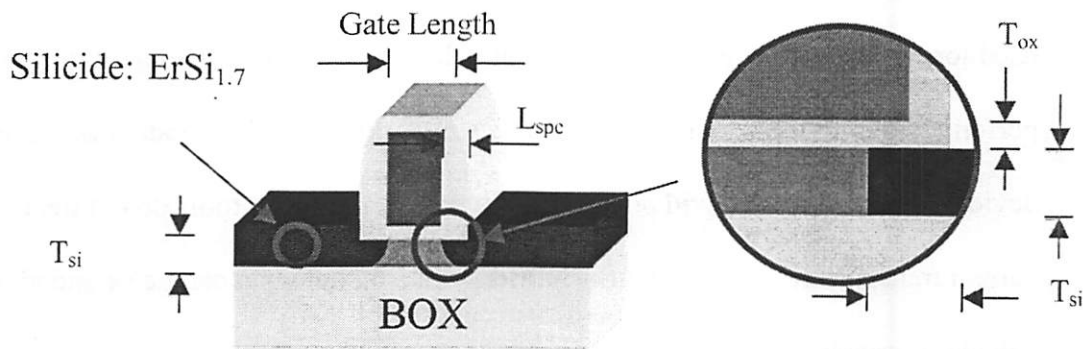
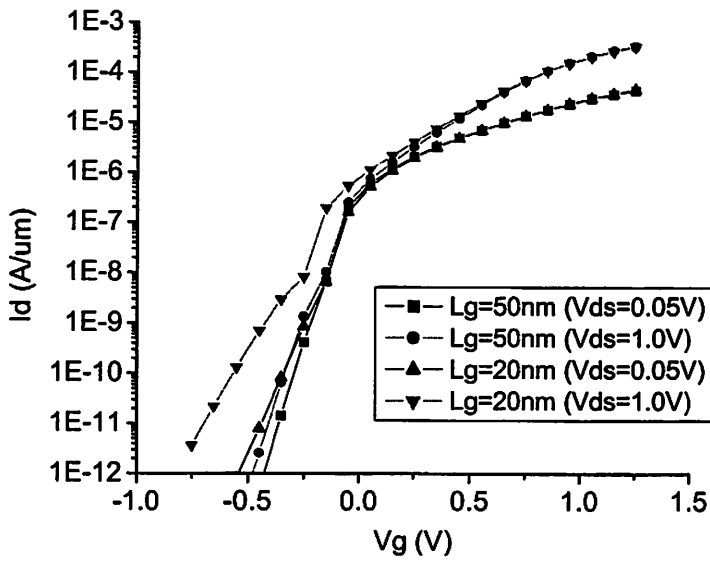


Figure 5.17: 2D UCS structure used for simulations in Fielday2D. Mesh (not shown) was refined near the gate oxide and the silicide junctions. N-type poly was used as the gate, but gate depletion was not modeled.

L_g	= 50nm, 20nm
T_{ox}	= 20Å
T_{si}	= 50Å
L_{spc}	= 40Å
U	= 10Å
Φ_{Bn}	= 0.28eV

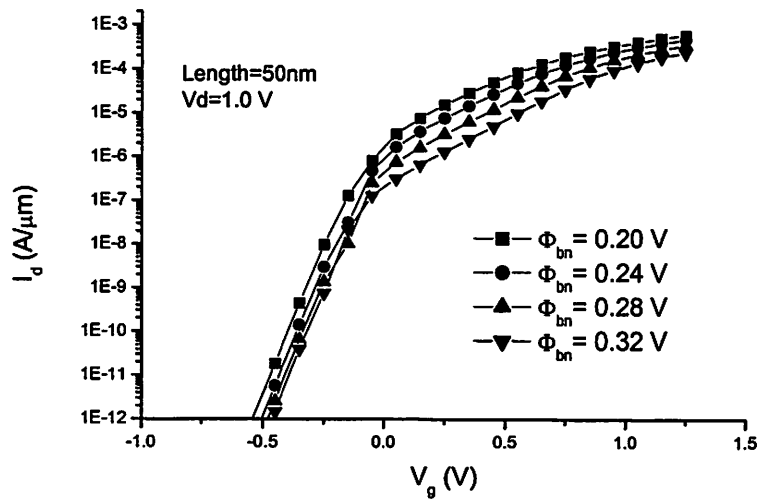
The standard structure dimensions were varied one at a time in order to determine their influence on device behavior. All simulations were done for two representative gate lengths $L_g=50\text{nm}$, and $L_g=20\text{nm}$. These two dimensions were chosen because devices with a gate-length of 50nm or more show very little short channel effects, and thus can be termed as ‘long channel devices’. Conversely devices with the aggressive 20nm gate-length shows significant short channel effects, and thus can be termed as ‘short channel devices’. The biases chosen for the simulations were initially 1.5V since experimental devices performed well up to this bias, however due to convergence problems the maximum V_d had to be lowered to 1.0V. The device characteristics of the standard UCS structure are shown in Fig. 5.18, for both gate lengths.

Figure 5.18: Fielday2D Id-Vg data for the standard UCS structure. Gate-lengths of 50nm and 20nm are shown, with Vds of 0.05V and 1.0V. The kink seen around 1e-8A/um is the result of a singularity in the model when the e-field perpendicular to the silicide interface is zero.



Clearly the barrier height is of critical importance when considering the UCS behavior, Fig. 5.19 shows how the variation in Φ_{b0} affects the I_d - V_g characteristic. As can be seen devices with high barriers will show an exponential I_d dependence on V_g even above threshold, with the I_d considerably lower than the devices with low barriers.

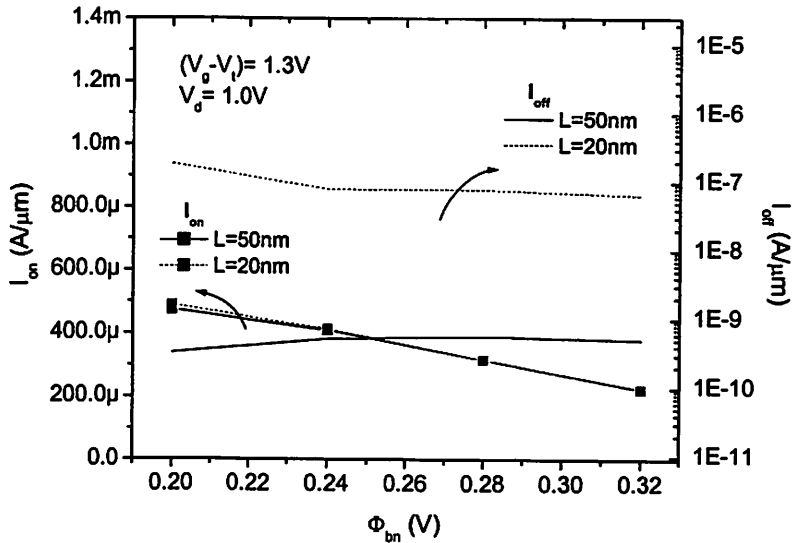
Figure 5.19: Fielday2D Id-Vg data for the UCS structures with different barriers. Gate-lengths of 50nm are shown with Vds of 1.0V.



The performance data can be presented in a more concise fashion by extracting the relevant drain current values I_{on} , and I_{off} , and plotting them against a particular device

parameter. I_{on} was extracted with the condition $V_g - V_t = 1.3V$, $V_d = 1.0V$, and I_{off} was extracted with the condition $V_g - V_t = -0.2V$, $V_d = 1.0V$. Fig. 5.20 shows the I_{on} , I_{off} plot for the Φ_{b0} data presented in the previous figure.

Figure 5.20: Fielday2D I_{on} - I_{off} data for the UCS structures with different barriers. Gate-lengths of 50nm and 20nm are shown with V_{ds} of 1.0V.

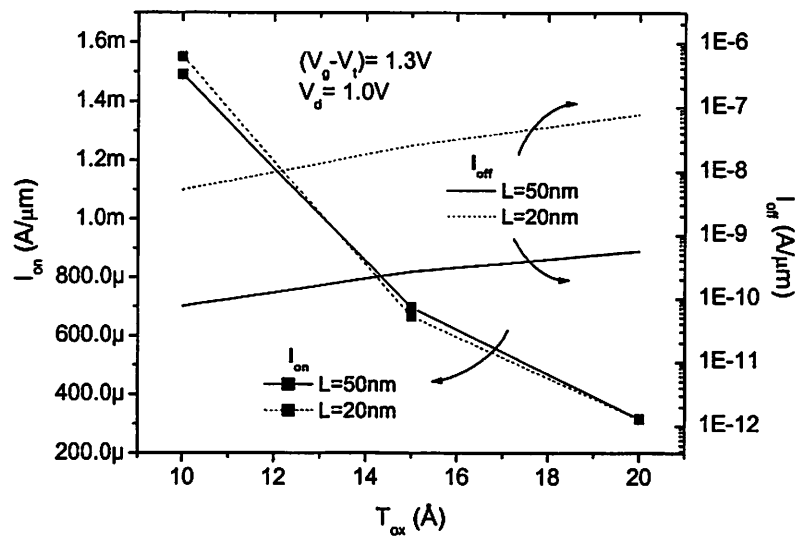


The dependence on the barrier height, although significant, is weaker than expected. Both conventional (section 2.1.1) and transmission (section 5.2) model predict that at a barrier height of 0.2eV the CS device should have current comparable to the MOSFET, 800-1000 μ A/ μ m, not the 500 μ A/ μ m obtained in the simulation. It is possible that inversion charge screening of the electric field is stronger than estimated, or that the simulator incorrectly extrapolates to barrier heights that differ from the experimentally fitted value of 0.28eV.

The dependence of the I_{on}/I_{off} characteristics on the gate oxide thickness is shown in Fig. 5.21. This strong dependence is due to the fact that a reduction in oxide thickness will increase the electric field, lowering the barrier and exponentially increasing the transmission current. High levels of I_{on} are achievable with an oxide thickness of 10 \AA . The simulator doesn't distinguish between the physical and electrical oxide thickness,

ignoring gate depletion and quantum mechanical inversion layer depth. While it is clear that ignoring the former effect overestimates the I_{on} , it is not so clear as to the effect of ignoring the latter. It is possible that ignoring the latter effect actually decreases the predicted current since with a quantum mechanical inversion layer depth the screening effect would be less significant. As expected, I_{off} shows a significant dependence on T_{ox} , falling by an order of magnitude as T_{ox} is reduced from 20Å to 10Å.

Figure 5.21: Fielday2D I_{on} - I_{off} data for the UCS structures with different oxide thickness. Gate-lengths of 50nm and 20nm are shown with V_{ds} of 1.0V.

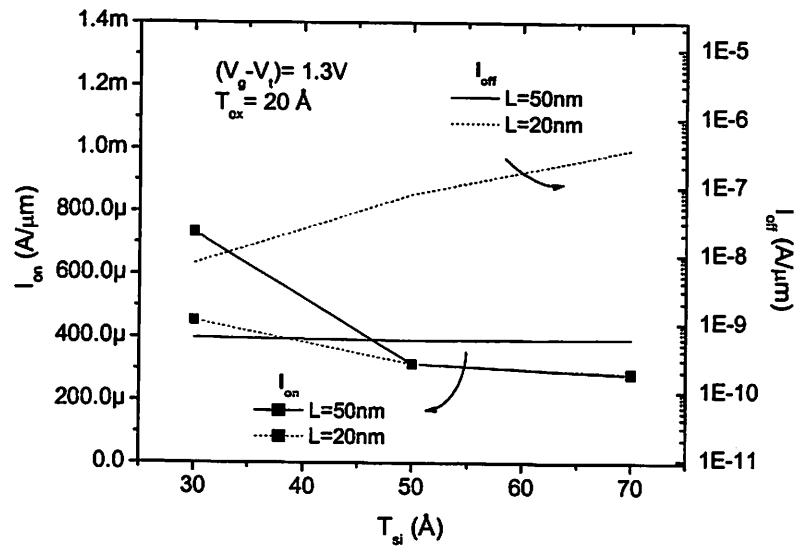


The influence of the silicon thickness is perhaps the most unexpected. An increase in silicon body thickness is expected to increase the off-current, due to the loss in the gate control of the body. This effect is observed, as seen in Fig. 5.22. However a significant dependence of I_{on} on T_{si} is also seen. This unexpected behavior is due to the dependence of the geometric factor G on the body thickness. A thinner body will have a larger G , leading to a larger e-field for the same bias conditions. The analogy to a capacitor is useful to the understanding of this effect. Consider a cylindrical capacitor held at a certain voltage, as the inner plate is shrunk in radius the electric field across it is

$E_m = V / (R_i \cdot \ln(R_o / R_i))$, where R_i and R_o are the inner and outer radius respectively. The behavior is non-monotonic but $E_m \rightarrow \infty$ for as $R_i \rightarrow 0$.

Analogously in the UCS device ($T_{ox} + T_{si}$) corresponds to R_o , while T_{si} corresponds to R_i . Reducing either T_{si} or T_{ox} to zero produces a singularity in E_m . The importance of G in device design will be discussed in the next chapter.

Figure 5.22: Fielday2D I_{on} - I_{off} data for the UCS structures with different body thickness. Gate-lengths of 50nm and 20nm are shown with V_{ds} of 1.0V.



5.4.2 Doped CS device simulation

Dimensions of the DCS structure simulated in Fielday2D are shown in Fig. 5.23. The variables of interest in this structure are the flat band barrier height, Φ_{b0} , the gate oxide thickness, T_{ox} , the silicon body thickness, T_{si} , the doping extension length, L_{ext} , and the doping level, N_{max} .

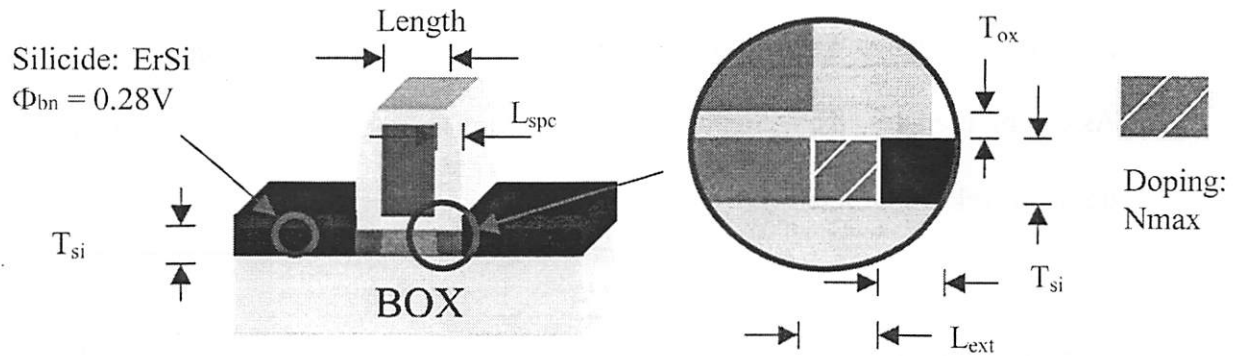
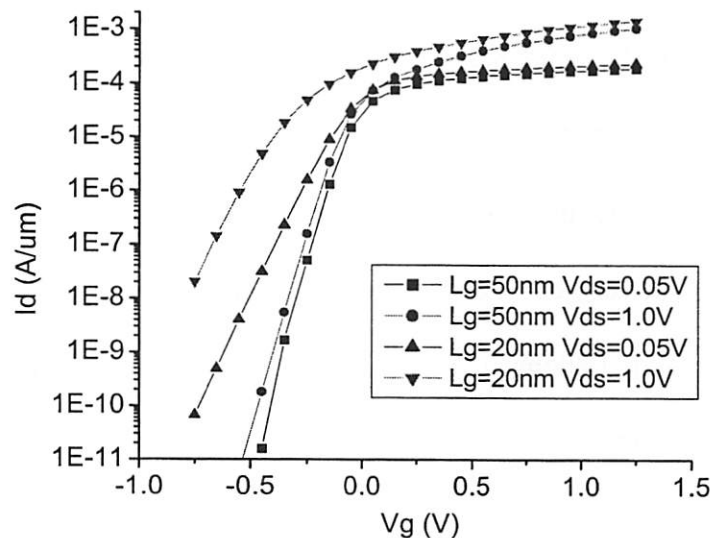


Figure 5.23: 2D DCS structure used for simulations in Fielday2D. Mesh (not shown) was refined near the gate oxide and the silicide junctions. N-type poly was used as the gate, but gate depletion was not modeled.

- $L_g = 50nm, 20nm$
- $T_{ox} = 20\text{\AA}$
- $T_{si} = 50\text{\AA}$
- $L_{spc} = 80\text{\AA}$ $L_{ext} = 30\text{\AA}$
- $U = 0\text{\AA}$
- $\Phi_{Bn} = 0.28eV$
- $N_{max} = 5e19cm^{-3}$

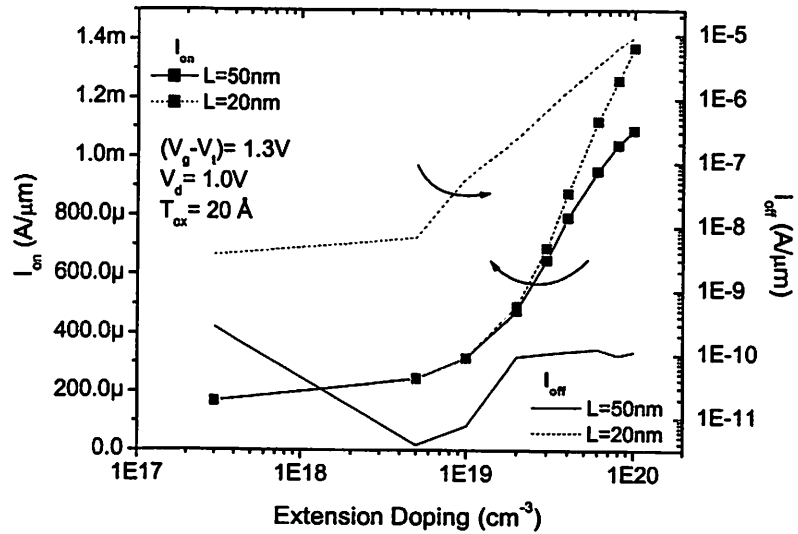
Like the UCS structures the DCS devices were simulated for gate lengths of 20nm and 50nm. The behavior of the DCS device is much more conventional, this can be seen immediately from the I_d-V_g plot of the standard DCS structure, which resembles that of a traditional MOSFET device, Fig. 5.24.

Figure 5.24: Fielday2D I_d-V_g data for the standard DCS structure. Gate-lengths of 50nm and 20nm are shown, with V_{ds} of 0.05V and 1.0V.



The extension doping concentration is of primary importance in DCS devices; its influence is shown in Fig. 5.25. For low doping levels ($3E17cm^{-3}$), the DCS structure performs like the UCS structure with a slightly lower G due to the difference in underlap. As doping increases the I_{on} increases, reaching the ITRS specification of $750\mu A/\mu m$ at a doping level of $3E19cm^{-3}$.

Figure 5.25: Fielday2D Ion-Ioff data for the DCS structures with different extension doping levels. Gate-lengths of 50nm and 20nm are shown with V_{ds} of 1.0V.

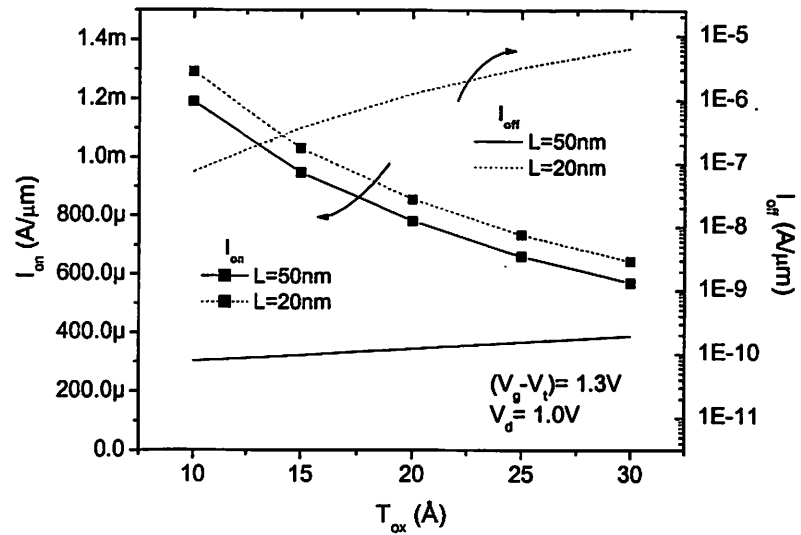


This doping value is just a little higher than what is predicted by the transmission model, and close to the value predicted by the conventional model. At doping values greater than $\sim 6E19cm^{-3}$ the I_{on} in long channel devices starts to saturate indicating that the Schottky barrier is no longer a significant component of resistance. The resistance predicted by the conventional model at $\Phi_{bn}=0.28eV$ and $N_{max}=6E19cm^{-3}$ is $\sim 200\Omega\cdot\mu m$, smaller than the equivalent channel resistance of $700\Omega\cdot\mu m$. Unfortunately, increasing the N_{max} has a detrimental effect on the I_{off} of short channel devices. Between $5E18cm^{-3}$ and $1E20cm^{-3}$ I_{off} increases exponentially up to $10\mu A/\mu m$. This is caused by the increased transparency of the Schottky barrier in the off state, and the increased barrier

lowering due to the lack of the built-in barrier at the source and drain. The role of the Schottky barrier in controlling I_{off} is clearly demonstrated in Fig. 5.25. Both the source and drain barriers contribute to the I_{off} control. By changing the boundary conditions at the end of the silicon region the barriers decrease the amount of DIBL for a fixed drain voltage.

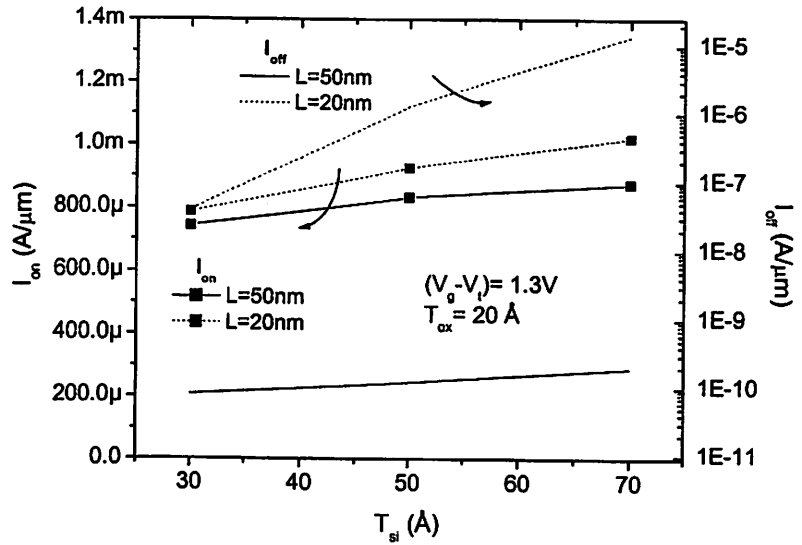
Fortunately decreasing the T_{ox} or T_{si} can control leakage current. Fig. 5.26 shows the influence of T_{ox} on DCS behavior. I_{on} increases steadily with decreasing T_{ox} , while I_{off} decreases by an order of magnitude by going from $T_{ox}=20\text{\AA}$ to $T_{ox}=10\text{\AA}$.

Figure 5.26: Fielday2D I_{on} - I_{off} data for the DCS structures with different gate oxide thickness. Gate-lengths of 50nm and 20nm are shown with V_{ds} of 1.0V.



Decreasing T_{si} has an even greater impact on I_{off} , Fig. 5.27, decreasing it by more than an order of magnitude between $T_{si}=50\text{\AA}$ and $T_{si}=30\text{\AA}$. The influence of body thickness on I_{on} is slight, with I_{on} dropping as T_{si} is decreased. This effect is due to the lower contact area at the source and therefore increased R_c . This is in stark contrast to the UCS structure, which shows a higher I_{on} with decreasing body thickness.

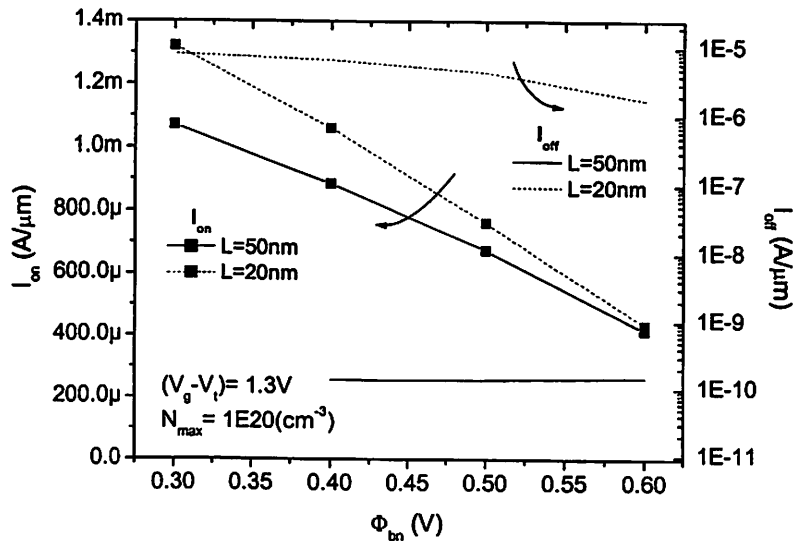
Figure 5.27: Fielday2D I_{on} - I_{off} data for the DCS structures with different body thickness. Gate-lengths of 50nm and 20nm are shown with V_{ds} of 1.0V.



The influence of Φ_{bn0} on the DCS device behavior is also interesting to note.

Specifically if DCS devices with a Φ_{bn0} as high as 0.6eV met I_{on} requirements, then traditional mid-gap silicides could be used to metalize thin-body devices, and there would be no need for the introduction of CS. Fig. 5.28 shows that even with a doping level of $1\text{E}20\text{cm}^{-3}$ increasing the barrier to 0.6eV decreases the I_{on} to $400\mu\text{A}/\mu\text{m}$. This is a smaller performance loss than would be expected from the conventional model, again indicating that the Fielday2d model, is too insensitive to changes in barrier height.

Figure 5.28: Fielday2D I_{on} - I_{off} data for the DCS structures with different barrier heights. Gate-lengths of 50nm and 20nm are shown with V_{ds} of 1.0V. Extension doping is $1\text{E}20\text{cm}^{-3}$.



References:

- [1] Jakub Kedzierski, Peiqi Xuan, Erik H. Anderson, Jeffery Bokor, Tsu-Jae King, Chenming Hu, "Complementary silicide source/drain thin-body MOSFETs for the 20nm gate length regime," International Electron Devices Meeting 2000, p. 57-60, Dec. 10-13 2000
- [2] Hellberg, P.-E.; Zhang, S.-L.; Petersson, C.S. "Work function of boron-doped polycrystalline Si/sub x/Ge/sub 1-x/ films," IEEE Electron Device Letters, vol. 18 (no.9), p. 456-8, Sept. 1997
- [3] Datta, S.; Assad, F.; Lundstrom, M.S. "The silicon MOSFET from a transmission viewpoint," Superlattices and Microstructures, Vol. 23, (no. 3-4), p.771-80, 1998
- [4] Lundstrom, M. "Elementary scattering theory of the Si MOSFET," IEEE Electron Device Letters, vol.18 (no.7), p.361-3, July 1997
- [5] G. Timp, J. Bude, K. K. Bourdelle, J. Garno, A. Ghetti, H. Gossman, M. Green, G. Forsyth, Y. Kim, R. Kleiman, F. Klemens, A. Kornblit, C. Lochstampfer, W. Mansfield, S. Moccio, T. Sorsch, D.M. Tennant, W. Timp, R. Tung, "The ballistic nano-transistor," International Electron Devices Meeting 1999, p. 55-58 Dec. 5-8 1999
- [6] M. H. Unewisse, and J. W. V. Storey, "Conduction mechanisms in erbium silicide Schottky diodes," Journal of Applied Physics, Vol. 73 (no. 8), p. 3873, 15 April 1993
- [7] S. Vandre, T. Kalka, C. Preinesberger, and M. Dahne-Prietsch, "Flatband conditions observed for lanthanide-silicide monolayers on n-type silicon," Physical Review Letters, Vol. 82 (no. 9), p. 1927, 1 March 1999
- [8] Kedzierski, J.; Peiqi Xuan; Subramanian, V.; Bokor, J.; Tsu-Jae King; Chenming Hu; Anderson, E. "A 20 nm gate-length ultra-thin body p-MOSFET with silicide source/drain," Superlattices and Microstructures, vol.28, (no.5-6), p.445-52, 2000
- [9] S. M. Sze, "Physics of semiconductor devices," 2nd ed. New York: Wiley, 1981
- [10] MeiKei Jeong; Solomon, P.M.; Laux, S.E.; Wong, H.-S.P.; Chidambarao, D. "Comparison of raised and Schottky source/drain MOSFETs using a novel tunneling contact model," International Electron Devices Meeting 1998, p.733-6, 6-9 Dec. 1998

6.0 Discussion

This section will discuss the advantages and disadvantages of CS devices, methods that may be used to obtain better performance, and alternative device structures. Emphasis will be placed on electrical and fabrication constraints placed by the microelectronics industry.

6.1 Comparison of DCS and UCS structures

The DCS experimental run did not yield doped CS devices as expected. Due to an implant problem, the e-beam defined regions of the wafers did not receive an implant. Therefore to compare DCS devices to UCS devices the DCS behavior must be obtained indirectly. Simulation on Fielday2d was used for this purpose, and the simulation results presented in the last chapter are discussed below. This discussion focuses on NMOS devices because the $\text{ErSi}_{1.7}$ has a higher barrier making it more difficult for CS NMOS devices to meet I_{on} targets. The DCS structure showed superior I_{on} in simulations presented in section 5.4, however the undoped structure consistently showed significantly lower I_{off} , for a similar L_{eff} . This is due to the fact that in the off-state the Schottky barrier helps to control the leakage current. Figs. 6.1, and 6.2 show the comparison of doped and undoped structures for different oxide thickness. With a 10\AA gate oxide both structures have a similar I_{on} , but the undoped structure has superior short channel characteristics, indicating that in this design space it might be an interesting alternative to the more traditional doped source/drain design. The operating voltage for the comparison is fairly high at $V_g=1.5\text{V}$, when the operating voltage is lowered the UCS device looks less favorable.

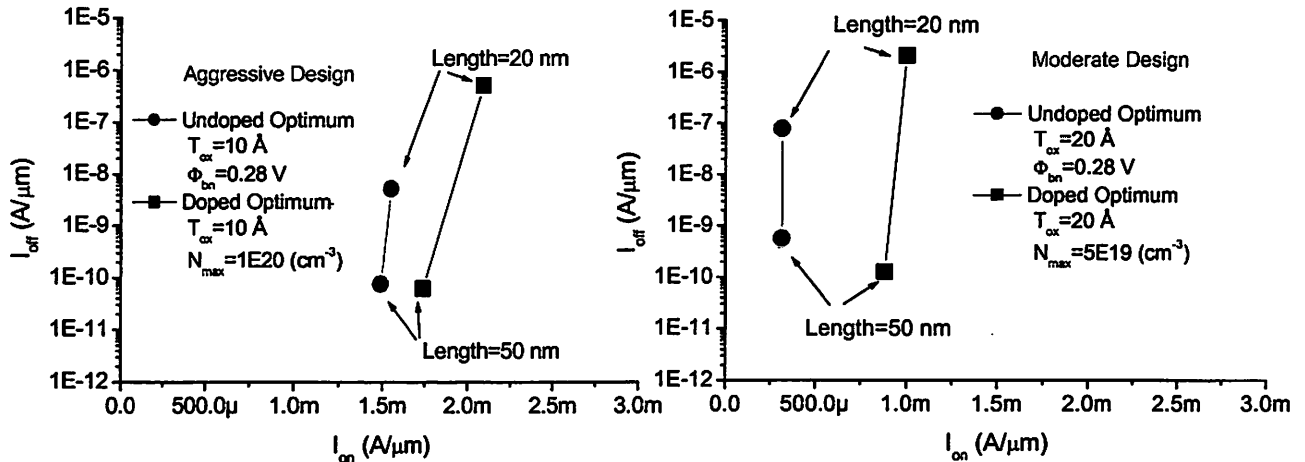
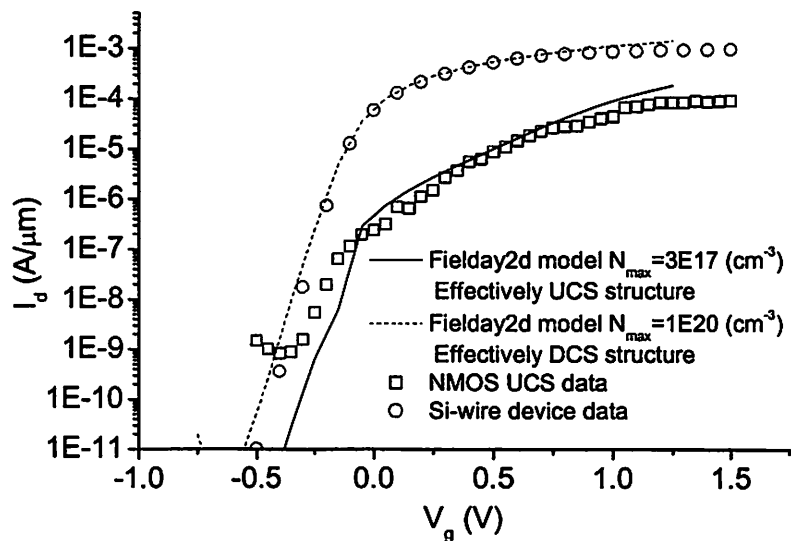


Figure 6.1: Comparison of I_{on} , I_{off} between doped and undoped designs for oxide thickness of 10Å.

Figure 6.2: Comparison of I_{on} , I_{off} between doped and undoped designs for oxide thickness of 20Å.

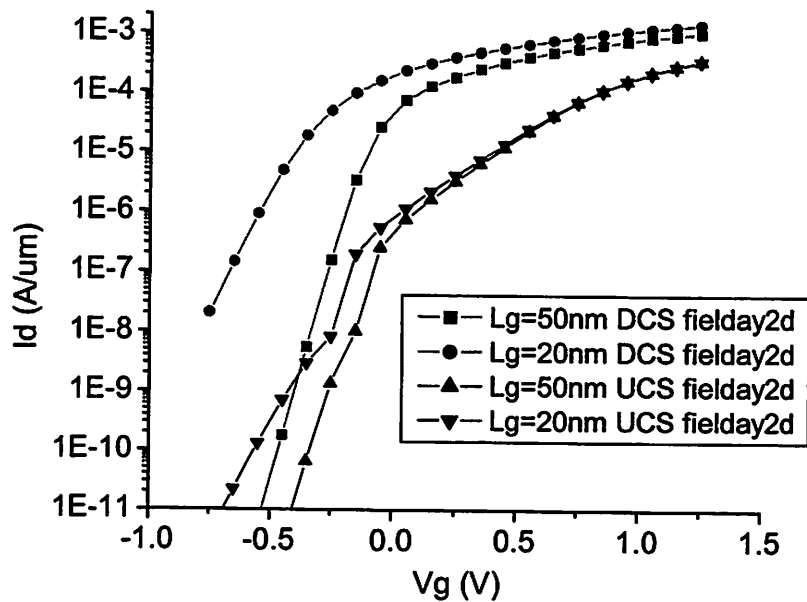
A secondary method of comparing the DCS and UCS structures is to assume the DCS will yield similar characteristics to a traditional thin-body device with no silicide. Then the experimental UCS data can be compared against data from unsilicided thin-body structures. Such a comparison is useful in its own right, since unsilicided thin-body structures such as the FinFET[1], and the RSDFET[2], have shown promising behavior at 20nm gate length dimensions. The similarity between DCS devices and traditional MOSFETs can be seen in Fig. 6.3.

Figure 6.3: The similarity between simulated DCS results and experimental MOSFET results taken from a Si wire device. Also shown is the comparison of fielday2d simulated and experimental UCS results.



Experimental data from a silicon wrap-around gate device[3] with doped silicon source/drain fits the simulated DCS data perfectly. The plot also shows that the experimental and simulated UCS data also match each other closely. The UCS I_d is particularly low due to the 40Å thick gate oxide of experimentally obtained UCS devices. A more fair comparison of the I_d - V_g plots can be made by assuming an equal gate oxide thickness of 20Å and using fielday2d simulation, Fig. 6.4.

Figure 6.4: Comparison of fielday2d simulated UCS and DCS results. Oxide thickness is 20Å. Note the difference in the current levels for low ($V_g - V_t$). Both devices have an N-type poly-silicon gate.



A strong argument in favor of DCS devices, seen in Fig. 6.4, is that at low voltages UCS devices perform poorly when compared with DCS counterparts. Reducing the operating voltage reduces the e-field in UCS devices, and since the drain current is exponentially dependent on the e-field, the reduction in performance is substantially larger than in DCS devices. Industry trends point to a continuing reduction in V_{dd} , to values where acceptable UCS operation may be difficult to obtain.

Single device current characteristics are not the only important parameter in the comparison of the UCS and DCS structures. Low variability between devices is

important to the qualification of the technology for manufacturing. Theoretical considerations indicate that UCS devices will be extremely sensitive to the exact geometry of the source, down to the atomic arrangement, and occupation of interface traps. Such sensitivity is common in field emitters in many other applications. Only when the contact resistance is very small will this sensitivity be reduced. A barrier of $\Phi_{b0} < 0.2\text{eV}$ would probably be sufficient, but materials that yield such a low barrier have not yet been found. Experimentally UCS devices showed large current variability, even when a single device was measured repeatedly. In a single device fluctuations in I_{on} of ~20% for NMOS and ~5% for PMOS were common. This is consistent with the above analysis since PMOS devices have a lower contact resistance component in the overall carrier path. The consistency argument strongly favors the DCS devices, which have an inherently lower R_c and a fixed electric field.

6.2 The design of DCS doped extension regions

The DCS devices seem to come out ahead in this comparison, however a hybrid between the UCS and DCS devices may have the advantages of both. The doping profile of the DCS extensions can be engineered to give just the profile needed to induce the required barrier lowering.

Determining the optimal doping profile is a difficult problem because of the number of degrees of freedom that exist. In order to limit the search space from a huge array of arbitrary profiles a specific profile type is usually chosen. Our analysis will focus on the box profile presented in section 5.23. The box profile has only two degrees of freedom, the doping concentration, N_{max} , and the extension length L_{ext} . The simulation results indicate that increasing N_{max} in a DCS structure increases both the I_{on} and I_{off}

considerably, Fig. 5.25. To understand why it is helpful to note the exact potential in the channel during device operation as a function of extension region doping. Figs. 6.5-8 show the conduction band energy as a function of distance from the source silicide for different N_{\max} , with $L_{\text{ext}}=30\text{\AA}$ and $\Phi_{b0}=0.28\text{eV}$.

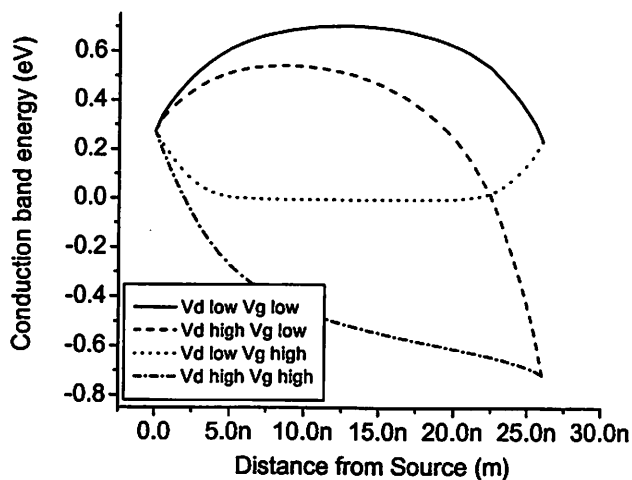


Figure 6.5: E_c profile for Low Doping, $N_{\max}=3\text{e}17\text{cm}^{-3}$. At this doping level the structure behaves like an UCS device.

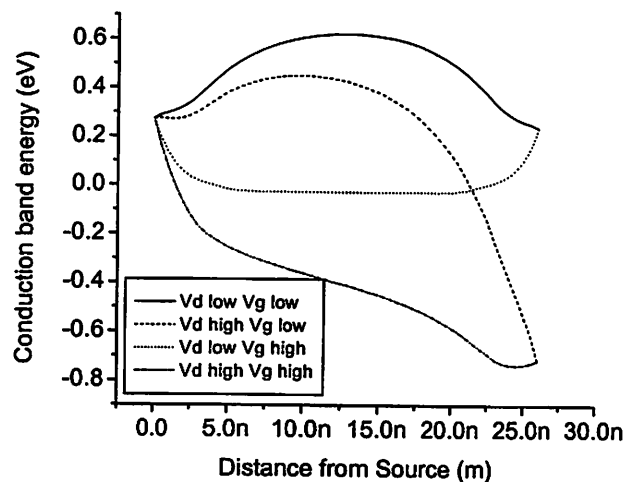


Figure 6.6: E_c profile for Medium Doping, $N_{\max}=3\text{e}19\text{cm}^{-3}$. At this doping level the structure behaves is a DCS device, but not quite like a traditional MOSFET, since the extension region is fully depleted in the off state.

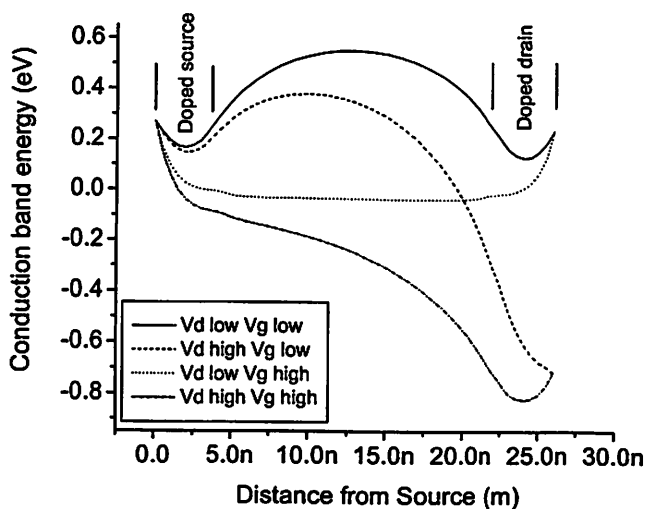


Figure 6.7: E_c profile for Heavy Doping, $N_{\max}=8\text{e}19\text{cm}^{-3}$. At this doping level the structure behaves is a DCS device. L_{ext} for all plots is 30\AA . $L_{\text{gate}} = 20\text{nm}$.

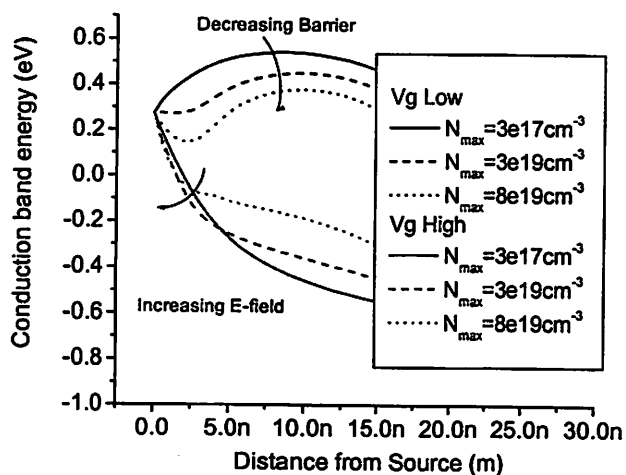


Figure 6.8: Comparison of E_c profiles for different doping levels. As extension doping increases the barrier in the off-state decreases and the electric field at the interface increases.

As expected, increasing the doping in the extension region increases the electric field perpendicular to the silicide during the on-state and therefore increases the on-current. Increasing the doping also forces the conduction band to a lower value near the source and drain, lowering the electron barrier in the off-state and therefore increasing the off-current.

Interestingly the DCS devices with a short L_{ext} do not function exactly like traditional MOSFETs since the extension region can become fully depleted even for fairly heavy doping levels. Depleting the extension region may give the DCS devices superior immunity to short channel effects, when compared with traditional MOSFETs. This is because the Schottky barrier still plays a role in controlling off current as long as the extension can be fully depleted. However a more detailed simulation is needed to compare the DCS structure versus the many available doping profiles used for the traditional MOSFET, to ascertain the extent of this effect. Fig. 6.9 below shows the design space for the extension doping of a DCS structure.

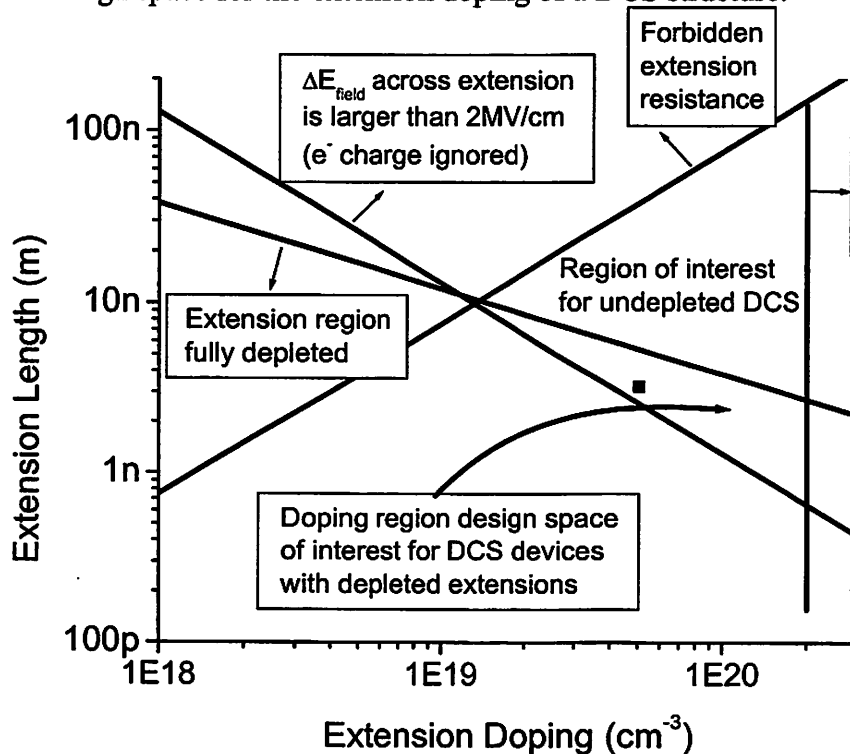


Figure 6.9: Design space for the DCS extension doping region. The boundaries shown are: fully depleted extension, 100 Ω - μm extension resistance (assuming 50 \AA body thickness), E-field change in extension of 2MV/cm, and the solid solubility limit.

This figure shows different design space regions for the two important extension doping variables, N_{\max} and L_{ext} . The first boundary of interest is the $100\Omega\text{-}\mu\text{m}$ extension region resistance, the region above this boundary has too high series resistance even when the R_c is set to zero. Below L_{ext} of $\sim 10\text{nm}$ this boundary is no longer physical since the Debye length reduces effective resistance, this region is however eliminated from the design space by another consideration.

That other consideration is the amount of doping required to reduce the contact resistance to an acceptable value. The 2MV/cm E-field change across the extension region is an approximation that ignores carrier charge in the region and assumes that the gate field will not significantly contribute to the $\sim 2\text{MV/cm}$ lateral field required to reduce R_c of an erbium silicide contact to $\sim 100\Omega\text{-}\mu\text{m}$ for a 50\AA thin-body, see Fig.2.11. The region below this boundary does not have sufficient charge in the depleted source to induce a sufficiently large E-field. This boundary is no longer physical when the extension is not fully depleted or above the next boundary.

This third boundary is the region where the extension region is just long enough to be depleted by the gate field and Schottky barrier. The depletion boundary separates two types of DCS devices. Above the boundary the extension region is never fully depleted, and the device works exactly like a traditional MOSFET with a fixed parasitic resistance caused by R_c , R_{sh} , and R_{link} , see section 2.2. Below the boundary the DCS device shows interesting behavior, as discussed previously. It has a higher on current than an UCS device due to the additional e-field caused by the doping, but the Schottky barrier is not irrelevant, it still has an influence on the channel potential. Due to the

influence of the Schottky barrier, DCS devices in this region may show lower off-current than a similar traditional MOSFET.

It is also important to note that the gate still plays a part in the contact resistance of the fully depleted extension DCS device. Therefore to increase the gate induced E-field at the contact a high geometric factor G , see section 5.2, is required. A high G can be obtained by reducing L_{ext} . The ideal fully depleted extension DCS device doping profile would therefore be a highly doped region, $\sim 5 \times 10^{19} - 1 \times 10^{20} \text{cm}^{-3}$, with a short L_{ext} , $\sim 1 - 3 \text{nm}$. Getting such good alignment between the source doping and the silicide boundary is technologically challenging, and the doping abruptness requirements are higher than what is currently possible.

6.3 Geometry considerations for UCS devices

The importance of the geometric factor, G , in the performance of UCS devices can not be overstated. G defined in section 5.2 as the ratio between the effective perpendicular e-field at the silicide boundary and the vertical field in the gate oxide. This factor controls the slope of the $I_d - V_g$ curve in the Schottky barrier region above the source-body flat-band voltage. The sensitivity of the on-current on G is clearly seen in the transmission model, and in the fielday2d simulation, which shows a high I_{on} sensitivity to oxide thickness. Increasing the e-field through decreasing T_{ox} is the most effective method of obtaining better UCS devices, short of significantly reducing the barrier height.

However changing T_{ox} is not the only method of increasing the electric field. A double gate thin-body device has the potential of doubling G , and thus the e-field without changes in T_{ox} . Therefore double gate UCS devices are expected to perform significantly

better than single gate UCS devices fabricated in this experiment, even when the inherent factor of 2 increase in channel conductance per unit width is accounted for.

Interestingly, the device changes that result in better UCS behavior are all consistent with scaling trends. Reduction in oxide thickness, reduction in thin-body thickness, and the implementation of double gate structure all yield immunity to short channel effects and significantly increase UCS device performance. Therefore if technological problems can be overcome, the UCS structure may still prove interesting in the sub-15nm gate length regime.

Specific double gate structures are more and less amenable to complementary silicides due to fabrication considerations. Fig. 6.10 shows the three common classifications of double gate structures[4]. Structure types II and III are not conducive to silicidation in general because the thin-body region is not parallel to the wafer normal. Another relevant characteristic of structure types II and III is that the thin-body can be made thicker outside the channel region through processing techniques that don't require selective regrowth, making the raised source/drain approach an elegant method of solving the thin-body resistance problems without the use of CSs.

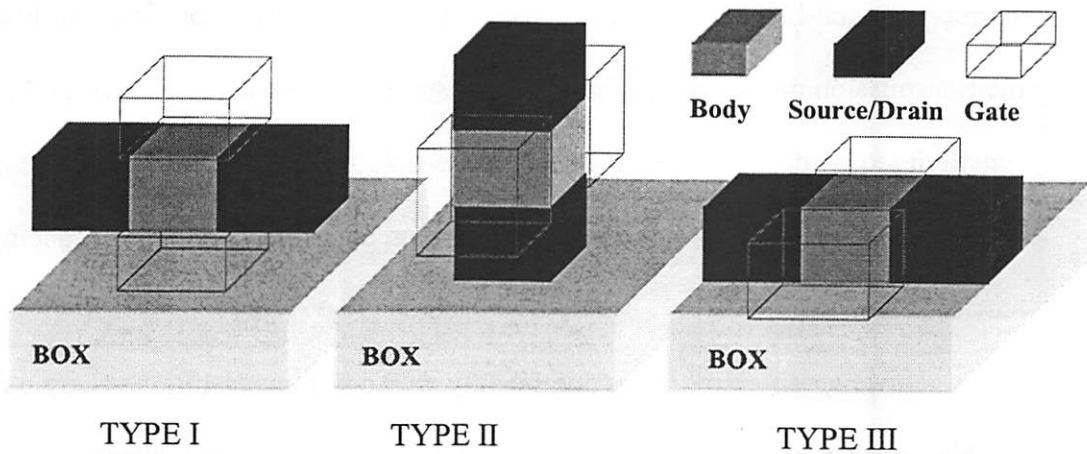
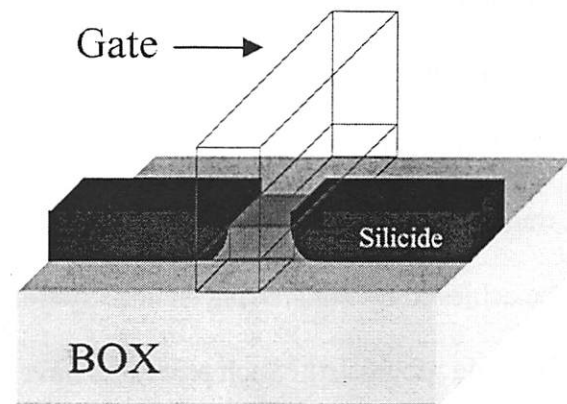


Figure 6.10: Different classifications of double gate structures. Black regions show the source and drain.

However the RSD approach to double gate design assumes that the thin body regions are spaced more than 100nm apart. This is due to the RSD lower thickness bound of 50nm per gate[5]. Placing thin-body regions more than 100nm apart in structure types II and III does not pose a problem except when the device width needs to be small.

An interesting structure that requires a small width is the silicon wire structure shown in Fig. 6.11.

Figure 6.11: The structure of a wire channel device with CS silicide to control thin-body resistance.



This structure, which is a hybrid between a planar thin-body and a type III double gate, has three active gates. It is therefore expected to scale better than a double gate structure. Because it has a small cross section per each 'wire', the wires must be placed close together. This spacing should be on the order of the gate length, and therefore is significantly smaller than the 100nm spacing limit. Such a structure would require a regrowth RSD approach or a CS approach to limit series resistance. Also the G factor of such a wire structure would be even higher than for a double gate device, resulting in better UCS behavior.

6.4 3D integration of UCS structures

UCS devices can be fabricated below 750°C. In the future, advances in metal gate and deposited gate-dielectric technologies may reduce the maximum formation temperature to as low as 450°C. Once formed the UCS structure is stable until ~800°C, without metal layers, and until 450°C with metal layers. This is in contrast to the conventional MOSFET structure which has to be subjected to a ~1000°C anneal during the doping activation step and subsequent ~1000°C anneals are not possible without shorting the devices and destroying the metallization. Therefore fabricating subsequent levels of devices directly on top of previously fabricated device layers is impossible with traditional MOSFET technology. In traditional MOSFET technology, 3D integration can be achieved by transferring an already finished device layer onto a target wafer through a bonding process[6]. Such processes have poor alignment and as such require sparse interconnection between layers.

UCS structures may be integrated directly on top of each other by fabricating a next layer on top of the previous one. Such an integration scheme gives excellent alignment and makes dense interconnection between layers possible. It is also an enabling technology for an interesting CMOS integration technique for UCS devices.

6.5 CMOS integration of CS devices

Fabricating PMOS and NMOS CS devices on one wafer in close proximity is a challenging processing problem. One novel method of achieving this goal is to first fabricate the PMOS devices and then bond a single crystal silicon layer on top of the PMOS devices and fabricate the NMOS devices on it. Such a process is unfortunately only possible with UCS devices that have metal gates and deposited gate-dielectrics, as

discussed in the previous section. For all other structures a method must be found to integrate both of the silicides on one device layer.

A straightforward method for accomplishing this is to use two lithography layers to mask the NMOS silicide (NS) and PMOS silicide (PS) depositions. Masking a metal deposition is not as easy as masking an implant, since photoresist can not be present during the silicidation anneal. An interlayer dielectric material can however serve the same purpose. Therefore the integration scheme could be the following:

- 1) Deposit thin interlayer dielectric (ILD) (after source/drain formation)
- 2) Etch dielectric, only in PMOS areas (first lithography step)
- 3) Deposit platinum and form Psilicide, remove excess platinum
- 4) Deposit thin ILD, etch ILD in NMOS areas (second lithography step)
- 5) Deposit erbium and form Nsilicide, remove excess erbium
- 6) Deposit thick ILD
- 7) Rest of metalization process

An issue of concern is the stability of the Nsilicide during the ILD deposition. Due to its reactivity it may be necessary to protect the Nsilicide with a barrier layer such as TiN or use an ILD that can be deposited at low temperature[7].

6.6 Conclusion

Complementary silicide (CS) devices are a novel direction for sub-30nm gate-length technologies. The use of complementary silicides opens up integration possibilities for thin-body and double gate devices. Specifically the use of doped CS (DCS) devices provides an alternative to the raised source/drain geometry. The use of undoped CS (UCS) devices opens up the intriguing possibility of fabricating transistors at temperatures constrained only by gate dielectric formation.

The challenges that CS devices must overcome to be used in the microelectronic industry are formidable. Since DCS devices have similar electrical characteristics to CMOS devices, integrating them into a technology is significantly easier. The only serious challenges lie in the process integration. UCS devices face far more challenges, most significantly an Nsilicide that has a barrier lower than 0.25eV should be found. Secondly, the device swing in the Schottky barrier regime should not be significantly higher than 100mV/dec; otherwise UCS devices will have poor performance at low voltages. Achieving such swing values may be possible with aggressive oxide-thickness scaling and proper device geometry. Finally, UCS devices suffer from the instability of the metal-silicon junction, and have all the integration challenges of the DCS devices.

References

- [1] Xuejue Huang; Wen-Chin Lee; Charles Kuo; Hisamoto, D.; Leland Chang; Kedzierski, J.; Anderson, E.; Takeuchi, H.; Yang-Kyu Choi; Asano, K.; Subramanian, V.; Tsu-Jae King; Bokor, J.; Chenming Hu "Sub 50-nm FinFET: PMOS," International Electron Devices Meeting 1999, p.67-70, Dec. 1999
- [2] Yang-Kyu Choi; Yoo-Chan Jeon; Ranade, P.; Takenuchi, H.; Tsu-Jae King; Bokor, J.; Chenming Hu "30 nm ultra-thin-body SOI MOSFET with selectively deposited Ge raised S/D," 58th DRC. Device Research Conference. p.23-4, 2000
- [3] Kedzierski, J.; Bokor, J.; Anderson, E. "Novel method for silicon quantum wire transistor fabrication," Journal of Vacuum Science & Technology B, Vol.17 (no.6), p.3244-7, Nov. 1999
- [4] H.-S. Philip Wong, "Novel device options," 1999 IEDM short course: sub-100nm technology, (organizer Mark Bohr), p. 19, Dec. 5 1999
- [5] Harvey I. Liu, James A. Burns, Craig L. Keast, and Peter W. Wyatt, "Thin silicide development for fully-depleted SOI CMOS technology," IEEE Transactions on electron devices, Vol. 45, no. 5, pp. 1099-1104, May 1998
- [6] Auberton-Herve, A.J.; Bruel, M.; Aspar, B.; Maleville, C.; Moriceau, H. "Smart-Cut(R): the basic fabrication process for Unibond(R) SOI wafers," IEICE Transactions on Electronics, Vol. E80-C (no.3), p.358-63, March 1997
- [7] Price, D.T.; Gutmann, R.J.; Murarka, S.P. "Damascene copper interconnects with polymer ILDs," Thin Solid Films, vol.308-309, p.523-8, 31 Oct. 1997

Appendix A

Quan 3 run detailed process information

0.0 Pattern information:

Mask patterns generated by DW2000 – full script in Appendix C

Critical script variables:

E-beam field = 131um

Pad size =

(Alignments)

1.0 Starting wafers

4 SOItech wafers 4"

nominally 1000A si/ 4000A oxide

2.0 Thinning oxidations

Sink6 clean w/ HF

Thinning oxidation 1: 1000C 17m Swetox; 1510Å oxide grown

Sink 6 clean w/ HF

ST1-2

Thinning oxidation 2: 1000C 30m Sgateox; 306A oxide grown

ST3-4

Thinning oxidation 2: 900C 12:15m Swetox; 459A oxide grown

3.0 SiGe alignment layer deposition

Tylan19: SiGe recipe

Nucleation: 550C 1m, 300mT, 200sccm(SiH4)

Deposition: 500C 80m, 300mT, 186sccm(SiH4) 33sccm(GeH4)

4.0 I-line lithography Layer 0 (Alignment Mark)

I-line resist application, primeoven, 1.0um I-line, standard hardbake

I-line exposure

Job F131 Side 2

pass 1: mask, Q3 Al mark

Apertures at minimum

41 columns X 11 rows, spacing of 1.905 mm

Alignment marks: 8r 41c, 8r 1c

pass 2: blank mask

Apertures at 10.0 x 7.5y

10 columns X 3 rows, spacing of 9.9x, 7.4y

drop 8column, 3row

5.0 Alignment mark etch

Lam 5, recipe 5003

Breakthrough: 10s; 13mT press, RF top 200W, RF bottom 40W, 100cc CF4

Main Etch: 60s (auto end point); 12mT press, RF top 300W, RF bottom 150W, 150cc HBr

Over Etch: 40s; 80mT press, RF top 200W, RF bottom 150W, 100cc HBr, 1cc O₂, 100cc He

6.0 Resist strip

technics-c 7m, 300W O₂
sink 8 clean (45s HF)

7.0 G-line Calix lithography Layer 1 (Mesa)

G-line resist application

mix 1 g-line: 4thinner
primeoven
Spin 4000rpm for 1m (800A Gline)
2m 90C bake

G-line exposure

Job Q3L1
pass 1: mask, Q3 Layer 1
Apertures at minimum
41 columns X 11 rows, spacing of 1.905 mm
Alignment marks: 8r 41c, 8r 1c
dropouts: c27-30, c12-14
no soft or hard bake
develop (standard)
Hard bake 120C 1 hour

Calixarene resist application

10m 105C bake
Spin 2000rpm (4% calixarene) (800A Calix)
10m 105C pre-bake
Job name: Q3L1 (dose 21k uC/cm², dose multipliers 4nm=3.3x, 6nm=2.4x
8nm=1.9x,
10nm=1.6x, 12nm=1.4x, 14nm=1.3x, 16nm=1.2x, 18nm=1.1x)
Develop xylene (30s)
Rinse DI (don't use IPA)
Rinse in xylene 30s
Rinse DI

8.0 Mesa Etch

Lam 5, recipe 5963
Breakthrough: 3s; 10mT press, RF top 450W, RF bottom 50W, 50cc CF₄
Main Etch: 2s; 15mT press, RF top 300W, RF bottom 150W, 50cc Cl₂, 150cc HBr
Over Etch: 3s; 35mT press, RF top 250W, RF bottom 120W, 200cc HBr, 5cc O₂

Measured etch rates: BT Si 39A/s, Oxide 40A/s; ME Si 65A/s

9.0 Resist Strip

technics-c 1.5m, 100W O₂
sink 8 clean

10 Sacrificial Oxide

Sink 6 clean

Tylan5, recipe Thin-Ann; 820C 20m dry O2

Measured ox (63A)

HF 25:1 1m (dewet 45s)

Gate stack formation

Sink 8 clean

Gate oxidation

Tylan6, recipe Thin-Ann; 785C 20m dry O2

Measured ox (37A)

Gate deposition

Tylan19, recipe SiGevar; 600C 50m, 300mT press, 100sccm SiH4, 2sccm PH3

Measured Si (630A)

Hard Mask deposition

Tylan9, recipe 9hoxn2od; 800C 55m, 10sccm SiCl2H2, 100sccm NH3

Measured oxide (160A)

Gate Lithography – Fine features

10m 105C bake

Spin 4000rpm (4% calixarene) (400A Calix)

10m 105C pre-bake

Job name: Q3L2 (dose 21k uC/cm2, dose multipliers 4nm=3.3x, 6nm=2.4x

8nm=1.9x,

10nm=1.6x, 12nm=1.4x, 14nm=1.3x, 16nm=1.2x, 18nm=1.1x)

Develop xylene (30s)

Rinse IPA

Hard Mask etch

Lam 5, Recipe 5702

23s, 20mT, Top RF 200W, Bottom RF 40W, 90cc CHF3, 200cc Ar

Measured etch rates: Si 8.8A/s, Oxide 15A/s, HTO 12A/s

Gate Lithography – Coarse features

G-line resist application

mix 1 g-line: 4thinner

primeoven

Spin 2500rpm for 1m (1000A Gline)

2m 90C bake

G-line exposure

Job Q3L1

pass 1: mask, Q3 Layer 2

Apertures at minimum

41 columns X 11 rows, spacing of 1.905 mm

Alignment marks: 8r 41c, 8r 1c
dropouts: c27-30, c12-14
no soft or hard bake
develop (standard)
Hard bake 120C 1 hour

Gate Etch

ST2 ONLY => 100:1 HF 20s
Lam 5, recipe 5693 /w 5702 BT
Breakthrough: 3s; 20mT, Top RF 200W, Bottom RF 40W, 90cc CHF3, 200cc Ar
Main Etch: 8s; 15mT press, RF top 300W, RF bottom 150W, 50cc Cl2, 150cc HBr
Over Etch: 3s; 35mT press, RF top 250W, RF bottom 120W, 200cc HBr, 5cc O2

Resists Strip

Technics-c 1m 100W O2 plasma
Measured SD Thickness: ST1 122A, ST2 112A, ST3 81A ST4 24A

Spacer Deposition (HTO)

Tylan9, recipe 9hoxn2od; 800C 67m, 10sccm SiCl2H2, 100sccm NH3
Measured thickness (200A)
Tylan7, recipe N2Anneal; 950C 10m

Spacer Etch

Lam 5, recipe 5702
17s, 20mT, Top RF 200W, Bottom RF 40W, 90cc CHF3, 200cc Ar

Measured etch rates: Si 5A/s, HTO 13A/s
Measured SD Thickness: ST1 112A, ST2 100A

Sink8 Clean

Photoresist Protection

I-line photoresist spin 1um
standard bake 90C 1m

Slice cut of e-beam fields for metalization

Disco, 1.904mm spacing, z-index 0.075mm, speed 3 (1mm/s)

Pre metal deposition clean

Acetone 1m
Technics-c 30s 50W O2 plasma (should be reduced in future)
25:1 HF 30s

NMOS branch

NMOS Erbium deposition

Load e/e chamber, 4 hour pump
E/e-buffer chamber transfer
Buffer chamber pump 15m
Buffer-Imetal transfer
Set sample temperature 20% (~200C)
Turn on LN cooling
Stabilize chamber 1h
Outgas source (10m per day of no use)
Dummy evaporation ~100A
Rotate arm to 225 (perpendicular position)
Evaporation (1.1A of Er per 1A of Si)
Set anneal temperature 30%-50% (300°C-450°C)
Anneal 1h
Set temperature 0%
Turn off LN cooling
Unload

Erbium etch

800:1 Nitric acid 10s
Rinse, Blow dry
800:1 Nitric acid 5m

PMOS branch

PMOS Platinum deposition

Load Ultek
Bake chamber 1hour
Evaporate Pt (0.75Å Pt per 1Å Si) 1Å/s

PtSi formation

Nitrogen purge oven 1h
250C 1h, 325C 1h, 400C 1h
Open oven turn off purge
Oxidation of silicide 400C 15m

Platinum removal

Aqua Regia (1 nitric:3 HCl: 4 water) 85C 30s

Metalization branch

Metal Lithography

Nanowriter KRS 2500Å process

Metal Evaporation

Veeco thermal evaporation
Evaporate 150Å Ti (adhesion layer)

Evaporate 450Å Al (conduction layer)

Metal Lift-off

Ultrasonic 5m

Acetone clean, DI rinse

Anneal 450C UHV 30m