Advanced Gate Materials and Processes for Sub-70 nm CMOS Technology

by

Pushkar Sharad Ranade

B.E. (University of Pune, India) 1996
M.S. (University of California, Berkeley) 1998

A dissertation submitted in partial satisfaction of the requirements

for the degree of

Doctor of Philosophy

in

Engineering – Materials Science and Engineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:

Professor Tsu-Jae King, Co-Chair
Professor Timothy D. Sands, Co-Chair
Professor Ronald Gronsky
Professor Chenming Hu

Fall 2002

The dissertation of Pushkar Sharad Ranade is approved:

_____

Co-chair Date

_____

Co-chair Date

_____

Date

_____

Date

University of California, Berkeley

Fall 2002

Advanced Gate Materials and Processes for Sub-70 nm CMOS Technology

by

Pushkar Sharad Ranade

Abstract

Advanced Gate Materials and Processes for Sub-70 nm CMOS Technology

by

Pushkar Sharad Ranade

Doctor of Philosophy in Engineering

Materials Science and Engineering

University of California, Berkeley

Professor Tsu-Jae King, Co-chair

Professor Timothy D. Sands, Co-chair

The continued evolution of CMOS technology beyond the 90 nm technology node will most likely be driven by advances in materials engineering and process integration. Fundamental changes in the materials used in the MOSFET gate stack will become necessary as will novel processing techniques and device structures. These in turn will introduce new process integration challenges. This work addresses two such challenges.

It is generally believed that the performance of deep-sub-micron CMOS transistors can be improved through the use of metal gate electrodes instead of the conventional poly-crystalline silicon electrodes. A major challenge in the introduction of metal gate electrodes is the need to obtain distinct gate work functions for NMOS and PMOS devices. While two metals would ordinarily need to be used on a single silicon substrate, a method that allows the metal gate work function to be tuned over the required range is highly desirable. In this work, a general framework for metal gate work function engineering is developed and the application of such techniques to advanced transistor fabrication is discussed.

The fabrication of ultra-shallow junctions in the source and drain regions of a transistor presents another significant process integration challenge. This task is more challenging for PMOSFET fabrication, given the low solid solubility and high diffusivity of boron in silicon. A novel approach for fabricating ultra-shallow and abrupt p+/n junctions is described. Selective germanium deposition on active silicon regions is used to create an elevated source/drain structure. The implantation of boron into germanium and the subsequent co-diffusion of germanium and boron into the silicon substrate is used to confine the B atoms to a shallow region under the substrate surface and form extensions to the gate electrode. Sub-70 nm elevated source/drain bulk-silicon PMOSFET devices have been fabricated to demonstrate the effectiveness of this approach.

_____

Professor Tsu-Jae King
Dissertation Committee Co-chair

_____

Professor Timothy D. Sands
Dissertation Committee Co-chair

Table of Contents

Acknowledgements

Many thanks are due to numerous colleagues and individuals who have directly or indirectly assisted in the preparation of this manuscript.

My advisor, Professor Tsu-Jae King has been instrumental in directing the progress of my doctoral research over the last four years. I would like to thank her for giving me the awesome opportunity to address some of the most daunting challenges facing the semiconductor industry today. She gave me ample freedom to pursue several different projects during the course of my research while always providing valuable insight and making sure that I did not lose sight of my primary research objectives.

I would also like to acknowledge the support of Professor Chenming Hu during my graduate research. He has been closely associated with a significant part of my research and his knowledge and mastery of the field have been truly inspirational. I would like to thank Professor Tim Sands, my advisor in the department of Materials Science and Engineering and co-chair of my dissertation committee. I have had the pleasure of knowing him ever since I joined Berkeley and he has always been supportive of my research endeavors and provided valuable guidance all along. My thanks also go to Professor Ron Gronsky for serving on my dissertation committee and Professors Oscar Dubón and Bill Morris for serving on my qualifying examination committee. Many thanks also go to Professors Vivek Subramanian and Jeff Bokor for their guidance during the course of my research.

I have had the privilege of collaborating with several exceptionally talented graduate students and colleagues over the last few years. I would like to thank Hideki Takeuchi for his guidance on various aspects of my research. He has been a terrific mentor in the Microfabrication laboratory and I have benefited immensely from his expertise. I would also

like to acknowledge Igor Polishchuk, Yee-Chia Yeo, Leland Chang, Kevin Yang, Ronald Lin and Qiang Lu for their support and close friendship which I will always cherish.

I have also had the pleasure of working with numerous graduate students and visitors of the Device Research Group, both past and present. I would like to thank Yang-Kyu Choi, Daewon Ha, Charles Kuo, Kevin Cao, Xuejue Huang, Sunil Bhave, Patrick Xuan, Jakub Kedzierski, Nick Lindert, Stephen Tang, Michael Orshansky, Andrea Franke, Qing Ji, Min She, Mark Cao, Pin Su, Sriram Balasubramanian, Jane Xi, Kenichi Goto, Koji Watanabe, Atsushi Yagishita, Kozo Watanabe and Renichi Yamada for their friendship and support.

I would like to acknowledge the fabulous administrative support of Judy Fong, Linda Manly and the staff of the Electronics Research Laboratory. I could always rely on them for processing the innumerable purchase orders and check requests that I generated over the last few years.

A large part of my doctoral research was performed in the Microfabrication Laboratory at Berkeley and many thanks go to Katalin Voros and the numerous highly talented lab technicians and staff members for keeping the lab equipment up and running. A part of my dissertation research was performed in the Summer of 2002 during my internship at the IBM T.J. Watson Research Center in Yorktown Heights, New York. I would like to thank Sandro Callegari, my mentor at IBM and my managers Raj Jammy and Wilfried Haensch for their valuable support during my stay at IBM. I would also like to thank Vijay Narayanan, Cyril Cabral, Sufi Zafar and Roy Carruthers for many useful discussions during my internship. Many thanks also go to Aditya Agarwal and Michael Ameen at Axcelis Technologies, Beverly, Massachussetts for their help with ultra-low energy ion implantation.

I would also like to acknowledge the agencies that have funded my graduate research. Over the years, this research has been supported in part by the University of California SMART program, the Semiconductor Research Corporation (SRC) and SEMATECH.

Finally, I would like to express my deep gratitude to my brother Shreekar and my parents Varsha and Sharad Ranade, who have always encouraged my academic endeavors inspite of the enormous physical distance that has separated us.

Pushkar Ranade

December 2002,

Berkeley

Chapter 1

Introduction

1.1 Historical perspective

The invention of the field-effect transistor (FET) and the subsequent development of monolithic silicon based integrated circuit fabrication techniques has led to unprecedented levels of growth in the semiconductor industry through the latter half of the 20th century. Through the years, complementary metal-oxide-semiconductor field-effect transistor (CMOSFET) device scaling to ever smaller physical dimensions has led to continuous improvement in device performance and to the widespread proliferation of microelectronic devices in day-to-day life as microelectronics based consumer appliances kept becoming *"smaller, cheaper and faster"*. The scaling of silicon based microelectronic devices and integrated circuits was originally predicted by Gordon Moore [1] in 1965 to last at least a decade. Phenomenal advances in photolithography techniques, tools and pattern transfer processes and equipment have however, ensured the validity of *Moore's Law* well into the 21st century. In an evolution driven primarily by photolithography, minimum feature sizes on devices have been scaled down from several microns in the 1970's to less than 100 nm today. This has enabled consistent improvement in device packing densities on chips, enhancement in circuit speeds and vastly improved performance-to-cost ratios for microelectronics based products. The resulting rise in the demand for microelectronics based products has led to an exponential growth in the global market for the semiconductor industry and a proportional

increase in research and development expenditures to ensure sustained growth. In 2001, the worldwide sales for the semiconductor industry totalled US $139B. However, trends for future device scaling are greatly dependent on significant changes in basic transistor materials and device structures. Scientists and engineers will have to address the rapidly approaching physical scaling limits for many conventional transistor materials. The turn of the century is beyond any doubt an exciting time for device and process integration engineers and materials scientists alike and the need for enhanced cross-disciplinary research has seldom been more pressing.

While industry estimates indicate that it might be possible to extend bulk-Si CMOS technology into the 50 nm gate length regime, it is becoming increasingly evident that the materials traditionally used in the front-end fabrication of the MOSFET, i.e. the transistor gate stack, are reaching fundamental physical limits. Figure 1.1 is a schematic cross-section across a CMOS structure which indicates the areas where novel materials and process

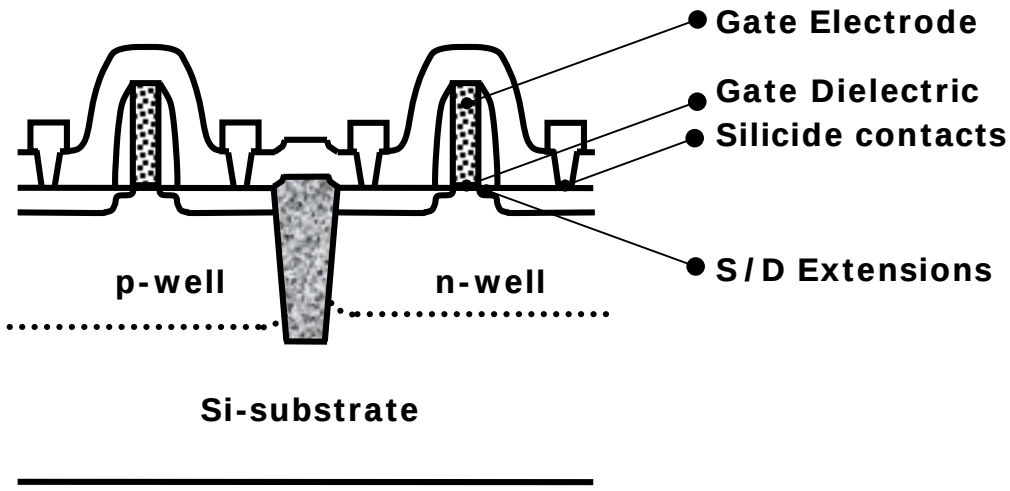


Figure 1.1 Schematic cross-section through a conventional CMOS gate stack indicating the areas where novel materials and integration challenges exist.

integration solutions are deemed necessary for continued device scaling. As an example, continued device scaling will necessitate the use of silicon dioxide gate dielectric films less than five monolayers (~1 nm) thick. At these thicknesses, direct tunneling currents through the dielectric will become much larger than desired (from power dissipation considerations). Moreover, the uniform growth of these ultra-thin films over 300 mm diameter Si substrates will pose new processing challenges. Just as silicon dioxide has been the gate dielectric of choice for several decades, poly-crystalline silicon has traditionally been the gate material of choice. One of the key advantages in favor of poly-silicon is the ability to convert undoped (*semi*-conducting) Si films to heavily doped (*highly*-conducting) *n* or *p* type Si films, making poly-silicon a CMOS compatible gate material. However, the functionality of doped silicon as a conductor does not prevent the formation of a thin depletion layer at the Si/SiO_2 interface when used as a transistor gate electrode. For nominal values of supply voltage, the depletion layer in a fairly heavily doped poly-silicon gate (1×10^{19} cm^{-3}) can be several angstroms thick. The presence of a depletion layer at the dielectric interface reduces the capacitive coupling between the gate and the channel and leads to a reduced density of inversion layer charge (for the same gate overdrive). In addition, a thicker effective dielectric layer increases the device threshold voltage (V_{TH}). With the effective gate dielectric thickness specification rapidly approaching sub-nm dimensions, poly-silicon depletion is bound to become a significant roadblock in the continued scaling of CMOS devices. Another limitation of doped poly-silicon gate electrodes is the penetration of dopant atoms from the gate into the transistor channel through the ultra-thin gate dielectric film. Dopant penetration and the resulting “autodoping” of the channel have been severe for boron doped poly-silicon owing to the lower atomic mass and small size of boron [2,3]. Apart from the gate stack of the transistor, the source and drain extension regions also present

significant materials and process integration challenges as CMOS technology moves beyond the 100 nm technology node. In an effort to improve the short channel performance of the transistor, source and drain extension junction depths have been continuously scaled down over the last few decades. Junction depth scaling has hitherto been satisfactorily achieved by the control of ion implantation parameters and the development of low-energy (sub-10 keV) implantation equipment capable of introducing a high dose of dopant ions into the silicon substrate at reasonable throughput. However, as nominal junction depth requirements approach sub-50 nm and as the need for highly conductive (< 200 Ω/) and very abrupt (< 3 nm/decade) junctions becomes more urgent, ultra-low energy (< 5 keV) ion implantation into silicon might no longer remain the most efficient junction formation technique. The need to fabricate ultra-shallow and abrupt extension junctions to the gate electrode with very high electrically active dopant concentrations warrants the development of alternative dopant introduction schemes and/or alternative device structures, viz. the elevated source/drain structure. In addition to minimizing junction sheet resistance, it is also essential to develop processes that ensure very low resistance contacts (specific contact resistivities $<1\times10^{-7}$ Ω-cm^2) between the front-end of the transistor (gate stack, source and drain) and the back-end metallization schemes. These and related materials and process integration challenges are deemed essential and must be overcome to ensure the continued scaling of convential bulk-Si based CMOS technology.

Alternatives to the conventional bulk-Si CMOS transistor structure are also being explored and have attracted attention as viable candidates to extend Si-based CMOS technology into the sub-15 nm gate length regime. These novel approaches involve the use of silicon-on-insulator (SOI) substrates to minimize leakage current paths through the bulk of the Si wafer and improve the subthreshold characteristics of the device. The development

of the ultra-thin body (UTB) Si transistor [4] and double-gate transistor structures like the FinFET [5] are two examples of SOI based transistor structures. The use of strained Si or SiGe films in the channel regions of a transistor has been proposed to enhance carrier mobilities in the channel and improve transistor drive currents [6,7]. While such novel approaches have traditionally been developed independent of the materials integration challenges mentioned earlier, it must be noted that the development of advanced gate stack materials and processes for bulk-Si CMOS technology will indeed also serve to enhance the performance of these advanced CMOS structures.

It is thus clear that the continued evolution of Si-based CMOS technology beyond the 100nm technology node will be driven by advances in materials engineering and process integration. CMOS scaling will no longer be driven by photolithography alone; rather it will be driven by the ability to *identify*, *characterize* and *integrate* novel materials into CMOS transistor fabrication processes and the timely transfer of these processes to reliable manufacturing setups.

The International Technology Roadmap for Semiconductors (*ITRS*) [8] is a highly coordinated global effort to plan the evolution of the semiconductor industry over the next decade and beyond. In annual revisions, the *ITRS* constantly monitors advances in semiconductor device research, development and manufacturing sectors and suggests areas which are expected to limit technology scaling and warrant additional resources. The 2001 edition of the *ITRS* lists several areas as major challenges in the near term as well as the long term. Among the most pressing near term challenges are alternative gate dielectric and gate electrode materials and alternative dopant introduction technologies for ultra-shallow and abrupt junction formation. The *ITRS* guidelines on gate dielectric and extension junction

scaling are highlighted in Fig. 1.2, which also indicates the roadmap estimates on the introduction of alternative gate stack materials and alternative junction engineering schemes.

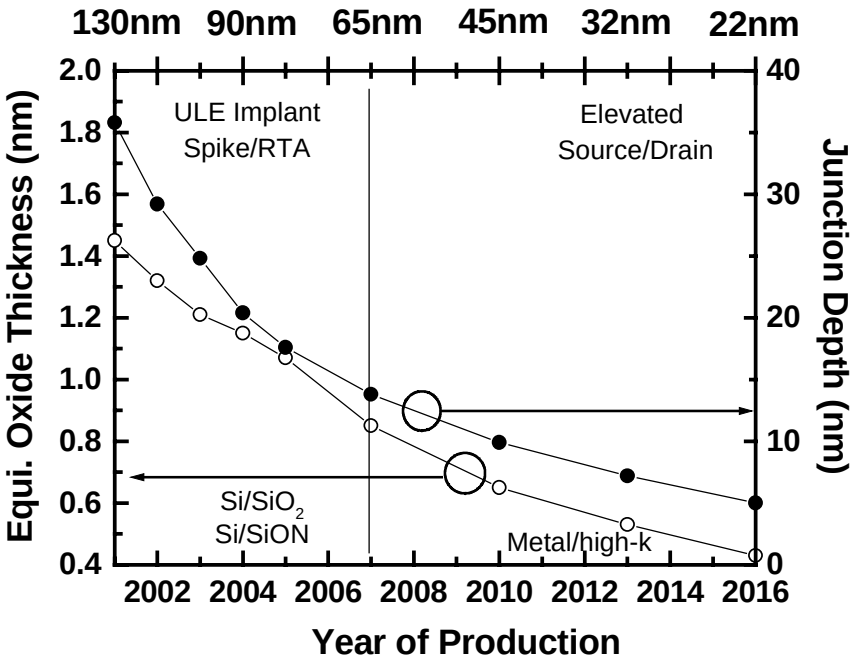


Figure 1.2 *ITRS* guidelines on CMOS gate dielectric and junction depth scaling and introduction of alternative gate stack materials and junction engineering schemes. Top and bottom horizontal axes indicate the lithography nodes and year of production respectively.

It has been the goal of this thesis to address in depth two critical challenges in the development of sub-90 nm CMOS devices.

1. Scaling limitations imposed by the use of poly-silicon gate electrodes
2. The fabrication of ultra-shallow (< 30 nm) and abrupt junctions in silicon

The scope of this work in general has been to analyze the above-mentioned challenges, to identify and characterize novel materials and to propose integration schemes that will enable the incorporation of these materials into manufacturable CMOS fabrication processes.

1.2 Advanced Gate Electrode Materials

CMOS transistor technology has traditionally relied on Si (elemental or as part of a compound) in several morphological forms (bulk single crystalline, thin film polycrystalline, thin film epitaxial) to serve the varied functions that are needed for the satisfactory operation of a field-effect transistor. Silicon and Si based materials have thus been used in several functional forms as semiconductor (channel, body) , insulator (gate dielectric, inter-metal dielectric) and conductor (gate electrode, silicide contacts) in a conventional field-effect transistor. While thermodynamic stability, ease of formation and acceptable dielectric properties have been some of the motivations behind the use of SiO_2 as a gate dielectric; ease of process integration and device performance issues like threshold voltage control have been important considerations in the use of poly-silicon gate electrodes. Research over the last few years indicates an increasing use of metals and metal based gate stack materials for future CMOS devices [9].

This work discusses several techniques developed to fulfil the need for dual (or multiple) work function metal gate CMOS technology. In order to be acceptable for use as gate electrodes in future CMOS devices, candidate metals must possess several general physical properties and must lend themselves easily to conventional Si processing techniques (lithography, deposition (physical/chemical vapor), etching (dry/wet), etc.). Candidate metals must have high (> 1000°C) melting points and must be thermodynamically stable on the MOSFET gate dielectric (SiO_2 or other alternative gate dielectrics). These metals must be very conductive and must have thermal expansion coefficients close to Si (to minimize the possibility of thermally induced stresses after rapid thermal processing). In addition to these requirements, the selection of materials for this application is also constrained by the need to obtain low and symmetric threshold voltages on the N- and PMOSFET devices. The gate

electrode of a MOSFET influences the threshold voltage of the transistor through its work function. In order to obtain low and symmetric threshold voltages on the two types of devices, while maintaining acceptable short channel performance, it is essential that the two gate electrodes have different work functions. The ideal gate work function for bulk-Si NMOSFETs (PMOSFETs) is ± 0.2 V of the E_C (E_V) of Si respectively [10] while that for ultra-thin body or double gate SOI-MOSFETs which use undoped channels is ± 0.2 eV of E_I [11]. It is thus evident that the integration of multiple metals on a single substrate or the development of single metal tunable work function schemes will become necessary in future CMOS gate stacks. From a process integration perspective, the latter is highly desirable. Engineering the work function of a metal thin film at dielectric interfaces is however not trivial. This thesis will briefly discuss two different approaches developed to fabricating dual-metal gate CMOS devices and focus on a novel "single-metal tunable work function" approach that might assist in the integration of metal gates into bulk-Si and SOI-CMOSFET devices. While there are a few potential work function engineering schemes, a novel approach has been developed to modulate the work function by modulating the structure and/or the chemistry in thin molybdenum (Mo) films. This approach involves the implantation and subsequent diffusion of argon or nitrogen ions in the metal films towards the dielectric interface. Preliminary results show a substantial shift in the Mo work function upon nitrogen or argon ion implantation and subsequent annealing.

1.3 Ultra-shallow Junction Formation

The fabrication of shallow and abrupt source-drain (S/D) extension junctions is another significant challenge in the continued scaling of CMOS transistor technology beyond the 100 nm technology node. According to the *ITRS*, bulk-Si CMOS devices with physical gate

lengths $L_g \leq 50$ nm will require ultra-shallow S/D extension junctions (junction depth $x_j \leq$ 30 nm) for acceptable short channel performance. In addition to being very shallow, extension junctions will also need to be heavily doped in order to meet roadmap requirements for the S/D series resistance ($R_s \leq 200\ \Omega/\square$) to improve device drive current and extremely abrupt (slope ≤ 3 nm/decade) for improved short channel performance [12]. While such aggressive specifications appear to be essential for the continued scaling of CMOS technology, it must be noted that the solid solubility and diffusion of dopants in Si impose fundamental limitations on the scalability of conventional processing schemes and materials used in the fabrication of extension S/D junctions.

The task of achieving ultra-shallow S/D junctions is more challenging for PMOS devices owing to the higher diffusivity and lower solid solubility of B in Si (as compared to As and P) [13,14]. The transient enhanced diffusion (TED) [15] of implanted dopant atoms, B in particular [16] is also perceived to be a major obstacle in the fabrication of ultra-shallow p+/n junctions.

The maximum achievable electrically active dopant concentrations are fundamentally limited by the solid solubilities of dopants in Si at the annealing temperature. Boron in particular has a fairly low peak solid solubility in Si (3×10^{20} atoms/cm^3 at 900°C). The fairly high diffusion coefficient of B in Si at 900°C and the transient enhanced diffusion of B during ramp-up impose additional constraints on the fabrication of p+/n junctions in Si. Such fundamental and practical constraints might thus limit the scaling of junction depths in Si using conventional processing techniques and device structures. This work has been an attempt to explore alternative device structures and materials that could enable the fabrication of ultra-shallow ($x_j \leq 30$ nm) p+/n S/D extension junctions with relatively low

annealing temperatures and thermal budgets. This thesis will present two simple approaches to incorporating Ge in the S/D regions of sub-100 nm conventional bulk-Si PMOSFETs and report on enhanced short-channel device performance obtained using these approaches.

1.4 Organization

The remainder of this thesis has been organized as follows. Chapter 2 will discuss in detail the requirements for metal gate CMOS technology – viz. materials and process integration challenges, potential candidate materials and integration schemes developed. Chapter 3 will discuss gate work function engineering schemes developed using Mo and W gate electrodes and cover experimental results that demonstrate Mo work function tunability using N and Ar implantation and physical characterization results that offer insight into the mechanisms behind the observed results. Chapter 4 covers dual work-function metal gate MOSFET device fabrication. Chapter 5 will discuss the application of $Si_{1-x}Ge_x$ in the formation of ultra-shallow S/D extension junctions. $Si_{1-x}Ge_x$/Si heterojunction diode characteristics and the impact of Ge on the formation of p+/n junctions in Si substrates will be presented alongwith a discussion of dopant diffusion phenomena across Ge/Si heterojunctions. Chapter 5 will also present a novel elevated S/D $Si_{1-x}Ge_x$ PMOSFET fabrication process and discuss the performance of sub-70 nm transistors fabricated using this technique. Chapter 6 concludes the thesis by summarizing the results obtained and the contributions made to the field alongwith suggestions for future work in related areas. The process flow used for transistor fabrication has been presented as an appendix.

1.4 References

1. Gordon Moore, “Cramming More Components onto Integrated Circuits”, Electronics, vol. 38, no. 8, April 1965
2. B. Yu, D.-H. Ju, W.-C. Lee, N. Kepler, T.-J King and C. Hu, “Gate engineering for deep-submicron CMOS transistors”, IEEE Transactions on Electron Devices, vol.45, 6, p. 1253, June 1998
3. K. Onishi et al, “Dopant penetration effects on polysilicon gate HfO_2 MOSFETs”, 2001 Symposium on VLSI Technology, Digest of Technical Papers, p.131, June 2001
4. Y. -K. Choi, Y. -C. Jeon, P. Ranade, H. Takeuchi, T. -J. King, J. Bokor, C. Hu, “Ultra-thin body SOI Device with 30nm Gate Length and Selectively Raised Ge S/D”, Conference Proceedings, IEEE 58th Device Research Conference, p.23, June 2000
5. X. Huang, W.-C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor and C. Hu**,** “Sub-50 nm P-channel FinFET”, IEEE Transactions on Electron Devices, vol.48, no.5, p. 880, May 2001
6. K. Rim, S. Koester, M. Hargrove, J. Chu, P. Mooney, J. Ott, T. Kanarsky, P. Ronsheim, M. Ieong, A. Grill, H.–S.P. Wong, “Strained Si NMOSFETs for high performance CMOS technology”, 2001 Symposium on VLSI Technology, Digest of Technical Papers, p.59, June 2001
7. Y.-C. Yeo, Q. Lu, T.-J. King, C. Hu, T. Kawashima, M. Oishi, S. Mashiro and J. Sakai, “Enhanced performance in sub-100 nm CMOSFETs using strained epitaxial silicon-germanium”, International Electron Devices Meeting 2000, Technical Digest, p. 753, Dec. 2000
8. International Technology Roadmap for Semiconductors, Semiconductor Industry Association, 2001 update
9. H.–S.P. Wong, “Beyond the conventional transistor”, IBM Journal of Research and Development, vol.46, no.2-3, IBM, p. 133, 2002
10. I. De, D. Johri, A. Srivastava, C. M. Osburn, “Impact Of Gate Workfunction on Device Performance at the 50 nm Technology Node”**,** Solid-State Electronics, vol.44, no.6, June 2000, p.1077

11. L. Chang, S. Tang, T.-J. King, J. Bokor and C. Hu, “Gate Length Scaling and Threshold Voltage Control of Double-Gate MOSFETs”, International Electron Devices Meeting 2000, Technical Digest, p. 719, Dec. 2000
12. Y. Taur, C. H. Wann and D.J. Frank, “25 nm CMOS Design Considerations”, International Electron Devices Meeting 1998, Technical Digest, p. 789, Dec. 1998.
13. F. A. Trumbore, “Solid Solubilities of Impurity Elements in Germanium and Silicon”, Bell Syst. Tech. Journal, 39, p. 205, 1960
14. R. A. Colclaser, Microelectronics: Processing and Device Design, Wiley Interscience, New York, 1980
15. P. Stolk et al, “Physical Mechanisms of Transient-Enhanced Dopant Diffusion in Ion Implanted Silicon”, Journal of Applied Physics, v. 81, 9, p. 6031, May 1997
16. A. Agarwal et al “Boron Enhanced Diffusion of Boron: The Limiting Factor for Ultra-Shallow Junctions”**,** International Electron Devices Meeting 1997, Technical Digest, p. 467, Dec. 1997

Chapter 2

Metal Gate CMOS Technology

2.1 Introduction

Polycrystalline silicon (poly-Si) has been used as a MOSFET gate material for several decades. One of the primary reasons for this is its high compatibility with CMOS processing. Poly-Si can be very easily deposited by low pressure chemical vapor deposition (LPCVD) techniques and it displays excellent thermal stability on SiO_2, which has been the gate dielectric of choice for MOSFETs. From a device performance perspective, the most important property of the gate electrode is its work function at the dielectric interface. In an MOS system, it is the work function of the gate that controls the threshold voltage of the device, by controlling how much applied voltage is necessary to bring about inversion in the transistor substrate. For a CMOS technology, it is generally desirable to have low and symmetric threshold voltages on the N- and PMOS devices. As device channel lengths have been scaled down over successive technology generations, dopant concentrations in the channel have been increased in order to maintain good short channel performance (i.e. in order to prevent the channel depletion region from becoming too large and the drain depletion charge from significantly penetrating into the channel region). This in turn requires the gate work functions to be close to E_C and E_V for bulk-Si N- and PMOS devices respectively (so that reasonably low threshold voltages can be maintained). The device threshold voltage can be also controlled by adjusting the dopant concentrations in the

transistor channel. However, this is not a very ideal way to control the threshold voltage since only a lightly doped transistor substrate can allow the use of a single (mid-gap) gate material for the two types of devices. While this would allow for symmetric threshold voltages, it would not allow the threshold voltages to be kept very low and more importantly, would deteriorate device short channel performance. Thus, dual-doped poly-Si has been a preferred CMOS gate material. The primary process integration benefit comes from the ease with which an undoped poly-Si film can be converted to a low work function ($n+$) or a high work function ($p+$) film using ion implantation of appropriate dopants (donors or acceptors) followed by a thermal annealing.

Perhaps the most pressing reason to look at alternative gate materials for deep-submicron CMOS devices arises from the fact that a heavily doped poly-Si gate will become slightly depleted at the gate dielectric interface when the channel is in inversion [1]. A schematic band diagram across the MOS stack is shown in Fig. 2.1 to illustrate this phenomenon.

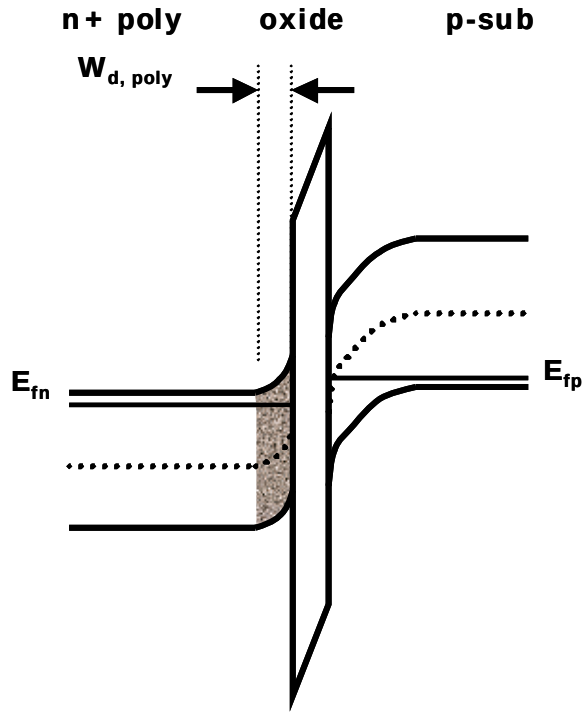


Figure 2.1 The poly-silicon gate depletion effect typically adds several angstroms to the dielectric EOT ($W_{d,\ poly}$) when the channel is in inversion.

A depletion layer in the gate can be harmful in several ways. It adds a capacitance in parallel to the gate dielectric capacitance and thus adds to the effective dielectric thickness between the gate and the channel. This implies a reduction in the capacitive coupling between the gate and the channel in inversion as shown in the MOSFET *C-V* curve of Fig. 2.2.

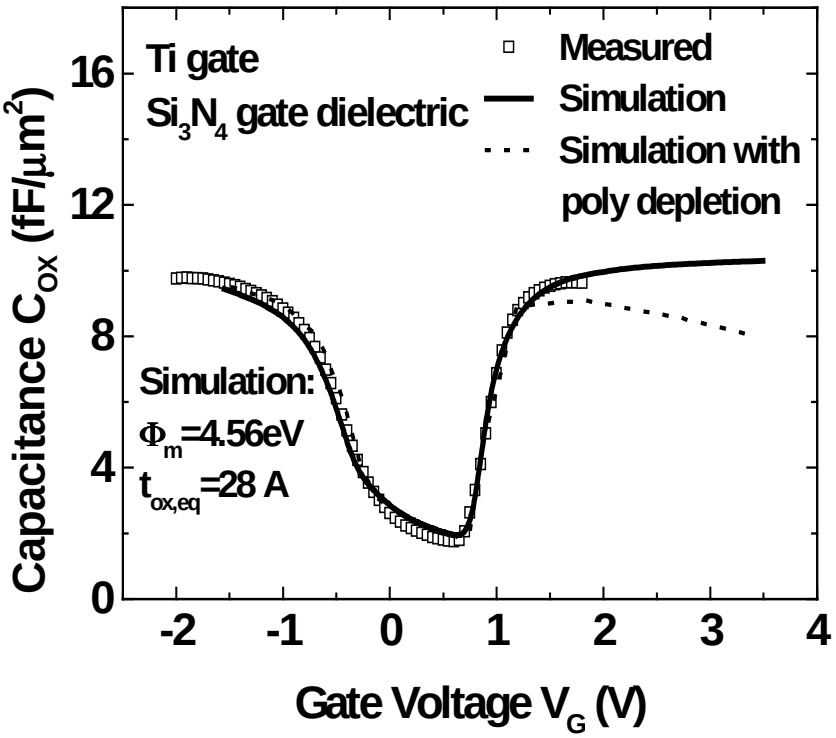


Fig. 2.2 NMOSFET *C-V* characteristics for a metal gate FET (Ti). Also shown are the results of a quantum mechanical *C-V* simulator for the same gate work function with and without poly-gate depletion effects. It can be seen that the absence of gate-depletion leads to a significant increase in the inversion capacitance. (From Ref.[2])

For comparable gate overdrives, the reduction in the inversion capacitance in the case of a poly-Si gated device lowers the drive current significantly [3]. The gate depletion effect is of particular concern for deep-submicron CMOS technology. As the dielectric thickness is scaled to 1 nm and below in sub-50 nm [4] devices, even the addition of a few tenths of a

nm in the form of a depletion layer can significantly increase the effective dielectric thickness of the device. This would also be a serious setback to all efforts to scale down the gate dielectric thickness using high-permittivity gate dielectrics. Because of this, alternative gate electrode materials which can either reduce or eliminate the gate depletion effect have been investigated by many researchers. The use of metal gate materials is most beneficial for devices with very thin EOT (< 1 nm) where high performance (high drive current) is most important. While refractory metals and their metallic derivatives are attractive candidates, there are numerous process integration challenges that must be overcome before a viable metal-gate CMOS technology can be developed.

In addition to the gate depletion effect, thermodynamic incompatibility with high-k gate dielectrics is also an issue for poly-Si gate technology. There have been successful demonstrations of MOS capacitors and transistors with ultra-thin EOT using high-k gate dielectrics viz. oxides of Hf, Zr, Al and La [5-8]. Many of these materials are not thermally stable in contact with poly-Si above certain temperatures, and therefore cannot be used in a conventional dual poly-Si gate CMOS process, in which the highest activation annealing temperature can exceed 1000°C. Therefore, alternative gate materials that are thermally stable with various advanced gate dielectric materials will be needed. It must be mentioned however, that the high-k dielectric is typically separated from the Si substrate by a thin interfacial layer (SiO_2 or Si_3N_4 or an *oxy-nitride*) and hence the bottom interface can withstand elevated temperatures.

When ultra-thin gate dielectrics are used in conjunction with poly-Si gate electrodes, dopant penetration from the gate into the channel can be a significant problem. It has been especially severe for *p*+ gate electrodes given the small mass and high diffusivity of B in Si. Boron penetration has also been observed to occur through high-k dielectrics [9] and is a

potentially significant concern for future gate stacks since it leads to instabilities in device operation.

An additional concern with poly-Si gate electrodes is the active dopant concentration limited conductivity. For CMOS applications, the poly-Si gate electrode is deposited undoped and subsequently converted to *n*- or *p*- type using appropriate ion implantation and annealing. The electrically active dopant concentration is thus inherently limited by the physical dose of dopant ions implanted into the gate, the thermal budget allowed to activate the dopants (i.e. move them to substitutional sites in the Si lattice) and the solid solubility of the dopant atoms in Si. The use of high-k gate dielectrics and the need to achieve abrupt and shallow source and drain extension junctions might not allow the use of high temperatures for dopant activation in future technology generations (with a conventional *gate-first* approach). This might lead to unreasonably high resistivities on the gate lines. A further difficulty arises with the continued scaling of the physical dimensions of the gate electrode. According to the *ITRS* [10], the aspect ratio for the gate electrode is expected to remain constant (Height/Length = 2/1) as the gate length is scaled down. Concomitantly, the gate sheet resistance is to be maintained at 5 Ω/ . This implies the need for a steady increase in gate active dopant concentrations as the gate length is scaled down. If the use of high-k gate dielectrics precludes the use of high thermal budgets for dopant activation, it is unclear how the specified values of active dopant concentration and gate sheet resistance would be achieved in a poly-Si gate technology. The solid solubility of the commonly used dopants in Si poses physical limitations on the gate sheet resistance. In order to achieve required gate sheet resistance of 5 Ω/ , the active dopant concentration in the gate would have to be $\sim 1.8\times10^{20}$ cm^{-3} for the 65 nm technology node. The solid solubilities of B, As and P at 1000°C are $\sim 1\times10^{20}$ cm^{-3}, 2×10^{20} cm^{-3} and 3×10^{20} cm^{-3} respectively [11]. It should be noted

that the required active dopant levels would imply super-saturated solutions of B in Si and near 100% activation of P and As.

Metal gate materials thus appear to be very attractive candidates for future CMOS devices. However, there are significant process integration challenges that must be overcome before a metal gate technology can be considered to be viable for integrated circuit manufacturing. These are discussed in the following sections.

2.2 Selection of Materials

Candidate gate materials must satisfy several criteria in order to be viable for use in a Si-based CMOS fabrication process. As mentioned earlier, for sub-65 nm technology nodes, the *ITRS* specifies a gate sheet resistance of $\leq 5\ \Omega/\ $. Thus these materials must be highly conductive and must have very high melting points in order to withstand the high thermal budgets commonly used in CMOS processing. It is also important that these materials lend themselves easily to conventional thin-film deposition (physical or chemical vapor deposition) and reactive ion etching (RIE) techniques. These requirements will ensure that advanced gate CMOS devices can still be fabricated using conventional tools. The new gate materials also need to have thermal expansion coefficients that closely match those of the single crystalline Si substrate in order to ensure that no significant thermal stresses are introduced in the film during rapid temperatures changes (as used for dopant activation). The general requirements described above already limit the candidate materials to some of the high melting point refractory metals, *e.g.* W, Ti, Ta, Mo, Nb, Re, Ru and their binary or ternary metallic derivatives, *e.g.* WN, TiN, TaN, MoN, MoO_2, TaSiN.

The most significant constraint in the choice of gate material, however, relates to the need to precisely engineer the transistor threshold voltages. As mentioned earlier, in order to obtain low and symmetric N- and PMOSFET threshold voltages (V_T) while suppressing short-channel effects, it is essential to have the gate work function between ±0.2eV of E_C and E_V for bulk-Si N- and PMOSFETs, respectively [12]. On the other hand, advanced transistor structures such as the ultra-thin body silicon-on-insulator (SOI) MOSFET [13] and the double-gate MOSFET [14] that use undoped Si channels require the gate work functions to be between ± 0.2eV of E_I, the intrinsic Si Fermi level [15]. There is thus a need to identify metallic materials that have work functions close to the above values and to develop process integration schemes that will provide multiple gate work functions for NMOS and PMOS devices integrated on a single Si substrate. A further challenge is imposed by the dependence of the metal work function on the properties of the underlying dielectric film. In general, metal work functions at dielectric interfaces differ from their values in vacuum [16]. This observation and its theoretical origins are discussed in a subsequent section and indicates that the search for metal gate materials must be conducted in tandem with the search for alternative gate dielectrics. It should also be noted that metals with complementary work functions also display inherent differences in physical properties like reactivity. Low work function metals are typically easily oxidized while high work function metals are inherently inert. This imposes additional constraints on the choice of metal gate material and process integration schemes.

In general, there are two major approaches to integrating novel gate-stack materials into a CMOS process. One is the "*gate-first*" approach, where the gate stack is formed before source and drain activation, as in a conventional CMOS process. The other is the "*gate-last*" approach, where the gate stack is formed after source and drain activation. An example of

the latter is a replacement gate (*damascene*) process that uses chemical-mechanical polishing (CMP) of the metal gate [17]. This process is difficult and relatively complex, and its adoption in mass production will require significant optimization and changes to processing tools and flows. The *gate-first* approach is generally preferred for its simplicity and compatibility with existing CMOS process flows. Two *gate-first* processes using dual metal gates are discussed in a subsequent section.

2.3 Candidate Materials

The choice of materials for this application is governed by two sets of requirements alluded to earlier in this chapter. One set encompasses general physical and chemical properties that are required by the constraints imposed by the device fabrication methods. The other set includes specific properties needed for optimum performance of the device.

2.3.1 Physical Properties

Candidate gate metals need to be thermodynamically stable on the gate dielectric and must be able to withstand thermal budgets commonly used in CMOS fabrication. The highest temperatures used in device fabrication are for the activation of dopant atoms in the source, drain and gate regions of the transistor. Typical activation thermal budgets include rapid thermal annealing at temperatures just above 1000°C (either a spike anneal with no soak time or for a few seconds soak at the highest temperature). This implies minimum melting temperatures of well above 1000°C for assured thermal stability. While SiO_2 is the currently used dielectric, it would be very likely that a change from SiO_2 to higher permittivity dielectrics will occur simultaneously with that from poly-Si to alternative gate electrode

materials. Hence, the new gate electrode material must be chosen in tandem with potential high-k materials being explored. The choice of high-k gate dielectrics is governed by similar thermal stability concerns with the Si substrate and specific physical and electronic properties such as the permittivity, band gap, electron and hole effective masses and band offsets with respect to the Si substrate. A modest permittivity ($15 \leq k \leq 20$) is expected to be an optimum value given the inverse relationship between permittivity and band gap (a smaller band gap is undesirable as it would lower the barrier height for electron (or hole) tunneling between the gate and the channel. Over the last several years, many research groups have actively pursued the search for new gate dielectrics starting from purely theoretical thermodynamic considerations [18,19] to experimental demonstration of high-k MOS capacitors and transistors [20,21]. While a lot of work remains to be done in this field, it appears that metal based dielectrics (binary or ternary oxides and silicates of refractory or rare earth metals) are most likely candidates to replace SiO_2 or SiO_2 based (Si_3N_4 or SiON) gate dielectrics. It is essential that the thermal stability of alternative gate materials on at least a subset of candidate dielectrics be experimentally verified. Typically, low work function metals are fairly reactive (prone to oxidation) (eg. Al, Ti, Ta) while high work function metals tend to be inert (eg. Ni, Pt, Pd, etc.) and naturally immune to oxidation.

The deposition technique used to deposit the gate electrode affects a number of important properties of the gate electrode. The morphology of the gate electrode is important for several reasons. An amorphous gate electrode is likely to have a work function significantly different from that of a crystalline one. In addition, different crystalline facets of a single crystal typically display a wide range of work functions. Thus, it is critical to identify deposition conditions that will lead to appropriate film morphologies and orientations. Furthermore, the deposition technique will also affect the quality of the interface between

the gate electrode and the dielectric. In general, a deposition technique that permits sharp interfaces to be obtained is desirable. Most metal thin films are deposited using physical vapor deposition techniques viz. sputtering (using DC, magnetron or RF sources) or evaporation (using electron beam or thermal evaporation). Unfortunately, most of these techniques involve the use of energetic particles (ions, electrons) and are likely to impart physical damage to the gate dielectric (rough interfaces, metal ions penetrating into the dielectric, etc.) and result in degraded gate dielectric reliability [22-25]. Chemical vapor deposition techniques would lead to minimal damage to the underlying dielectric and are desirable for this application. Chemical vapor deposition will also provide a number of variables, viz. temperature, pressure or gas flow to control the film microstructure. However, apart from a few exceptions (W and Ta) precursors for refractory metal CVD have not been very well identified or characterized. This is a challenging area that can be expected to become even more important as CMOS technology continues to be scaled down.

Yet another important consideration in the choice of a gate material is its patternablility. Gate films will need to be etched with high selectivity to the underlying dielectric. In addition, the gate etch should ideally result in vertical sidewalls without any protrusions or notches at the dielectric interface. This implies an optimization of reactive ion etch chemistries and process conditions that ensure high etch selectivity and ideal sidewall profiles. While wet etch chemistries might also be used to remove metal films with high selectivity to the gate dielectric, the use of wet etchants for sub-50 nm line patterning is unlikely given the isotropic nature of wet etching processes. The choice of gate material will have to account for the inherent reactivity or inertness of the metal under consideration.

State of the art CMOS fabrication processes involve the use of rapid thermal annealing (RTA) to activate dopant atoms in the source, drain and gate with minimal

diffusion. Typical RTA temperatures are between 900-1000°C and typical annealing times range from a few seconds to a minute. In some cases, spike annealing (no soak at peak temperature) is also used. Temperature ramp-up is typically between 200-250°C/s and cool-down of the wafer is also very rapid. This rapid thermal cycling can become a source of thermally induced stress in the MOS stack and might lead to cracking or peeling of gate films depending on the magnitude of the stress induced. In light of this, it is necessary to choose a gate material that has a coefficient of thermal expansion that is close to that of bulk Si.

2.3.2 Sheet Resistance

In addition to the abovementioned physical properties, potential gate electrode materials must also satisfy certain criteria specific to the type of device under consideration. As mentioned earlier, the sheet resistance of the gate line needs to be maintained to 5 Ω/ or less for sub-100 nm gate length nodes. While this is a general *ITRS* requirement, it is of particular concern for CMOS device design for high frequency (e.g. RF) communications products. One of the major impediments in implementing an RF CMOS technology is the relatively large resistances of the gate lines and the resulting loss in high frequency gain. Poly-silicon depletion and gate sheet resistance are thus of particular concern for RF applications.

2.3.3 Work Function

The most direct impact of the gate electrode on the operation of a MOSFET is through its control of the device threshold voltage (V_{TH}). The voltage required for the onset of inversion in the MOSFET channel is determined by the work function in the gate electrode. The threshold voltage of a MOSFET is typically given by the following expression:

$$V_{TH} = V_{FB} + \frac{\sqrt{2\varepsilon_{Si}.q.N_a.2\phi_B}}{C_{OX}}$$

where ε is the permittivity of Si, q is the electronic charge, N_a is the dopant concentration in the channel, ϕ_B is the band bending in the Si substrate and C_{OX} is the oxide capacitance. V_{FB} is the flat band voltage across the MOS stack and denotes the voltage applied on the gate electrode that eliminates any band bending across the stack. V_{FB} is given by the following expression:

$$V_{FB} = \Phi_{MS} + 2\phi_B - \frac{Q_f}{C_{OX}}$$

where Φ_{MS} indicates the work function difference between the metal gate and the Si substrate and Q_f denotes the magnitude of fixed charge in the oxide film.

The threshold voltage of the MOSFET is thus directly controlled by the gate electrode through the work function difference between the gate and the channel. The band bending in the channel is given by the following expression:

$$\phi_B = \frac{kT}{q}\ln\left(\frac{N_a}{n_i}\right)$$

where n_i is the intrinsic Si carrier concentration. For typical channel doping levels used (1×10^{16}-1×10^{18}cm^{-3}), ϕ_B is approximately 0.3-0.4 eV. To minimize V_{TH} the flat band voltage should thus be ~ –0.8 V for an NMOS device and ~ +0.8 V for a PMOS device. This in turn implies gate work function values of ~ 4.1 V for NMOS and ~ 5.1 V for PMOS devices. More precise calculations performed using 2D device simulations [12] indicate that for an optimum combination of short channel performance and device drive current, gate work functions for sub-50 nm bulk-Si CMOS tranistors should be between ± 0.2 V of E_C (E_V) for NMOS (PMOS). Sub-100 nm gate length bulk-Si MOSFETs typically need fairly

high dopant concentrations in the channel regions to prevent the drain depletion region from penetrating excessively in the channel leading to poor short-channel performance. However, high levels of channel doping typically degrade carrier mobility in the transistor channel. Ultra-thin body FETs and vertical transistor structures that use a double gate structure (e.g. FinFET) typically use undoped Si channels since the thin Si body automatically enhances the short channel performance of the device by eliminating sub-surface leakage paths between the source and the drain. The use of undoped channels also prevents degradation in channel mobility owing to the reduced impact of impurity scattering and a lower transverse electric field [26]. Another advantage of using undoped Si channels is the elimination of threshold voltage fluctuations which arise from statistical fluctuations in the dopant levels when the channel lengths become very small (< 10 nm). For such advanced device structures using fully depleted Si channels, the gate work functions on the N- and PMOS devices need to be closely centered around the intrinsic Si Fermi level ± 0.2 V of E_I, i.e. 4.4 V and 4.9 V for N- and PMOS devices respectively [15].

2.4 Metal Gate/High-k Dielectric Integration

An important consideration in the selection of metal gate electrodes is the work function dependence in contact with high-k dielectrics. In general, the work function of a metal at a dielectric interface is different from its value in vacuum. This variation needs to be taken into account when designing transistor gate stacks with alternative gate dielectrics.

The dependence of the metal work function on the permittivity of the gate dielectric has been explained by Yeo et al [16] using experimental data in conjunction with interface dipole theory [27-28]. In order to accurately predict metal work functions at dielectric

interfaces, one needs to take into account charge transfer across such interfaces. In general, there are always intrisic interface states at metal-dielectric interfaces [29] which can be predominantly donor like (closer to E_V) or acceptor like (closer to E_C). When an interface is formed between a metal and a dielectric, these interface states can get charged up resulting in a dipole that will drive the bands to align in a way that minimizes the charge transfer. The energy level at which the dominant character of the interface states changes between donor- and acceptor-like is called the charge neutrality level, E_{CNL}. For example, in the case where the metal Fermi level ($E_{F,M}$) is above the dielectric charge neutrality level ($E_{CNL,D}$), the dipole layer created at the interface will be charged negatively on the dielectric side. This dipole will tend to drive the metal Fermi level towards $E_{CNL,D}$ and hence the effective metal work function ($\Phi_{M,eff}$) will be different from the vacuum metal work function ($\Phi_{M,vac}$). This workfunction change is proportional to the difference between $E_{CNL,D}$ and $E_{F,M}$, or equivalently, the difference between $\Phi_{M,vac}$ and $\Phi_{CNL,D}$. Taking into account the effects of charge screening in the dielectric, the effective metal work function $\Phi_{M,eff}$ is given by

$$\Phi_{M,eff} = \Phi_{CNL,D} + S(\Phi_{M,vac} - \Phi_{CNL,D})$$

where S is a slope parameter that accounts for dielectric screening and depends on the electronic component of the dielectric constant, ε_∞, [30] according to

$$S = [1 + 0.1(\varepsilon_\infty - 1)^2]^{-1}$$

A larger dielectric screening lowers the slope parameter S and increases the degree of pinning of the metal work function to $E_{CNL,D}$. Values for ε_∞ for various dielectrics of interest (SiO_2, Si_3N_4, HfO_2 and ZrO_2) have been used to calculate the slope parameters for these dielectrics. Comparison of theoretical and measured interfacial work functions for several metals of interest (Al, Ti, W, Mo and Pt) led to reasonable agreement indicating the validity

of the theory. An important conlusion of this work was highlighted in [16], where it was shown that in order to obtain $\Phi_{M,eff}$ of 4.05 V (5.17 V) for NMOS (PMOS) gate electrodes, metals with even smaller (larger) $\Phi_{M,vac}$ have to be used as the permittivity of the gate dielectric is increased. The consequence of this theory on the selection of gate metals is that PMOS compatible metals will most likely be very inert (high work function), while NMOS compatible metals will be very reactive (low work function). These complementary properties are likely to have a significant impact on future CMOS gate stack process integration since very inert metals are by definition difficult to etch (a general lack of volatile compounds) while reactive metals are likely to react with the underlying gate dielectric (typically an oxidation reaction where oxygen from the dielectric can preferentially bond with the gate metal).

2.5 Process Integration

This section discusses process integration approaches to incorporate metal gate electrodes into transistor gate stacks. In general, one may use a combination of two or more metals on a single Si substrate to achieve the work function requirements discussed earlier. While the processes described herein have been successfully implemented, it must be noted that the integration of multiple metals on a single wafer poses significant process integration challenges and it would be highly desirable to develop a single metal tunable work function gate CMOS process. Such a process would allow for minimal process complexity and would be relatively easy to integrate on a single Si substrate. The following chapter is devoted to a discussion of metal gate work function engineering and the development of single metal tunable work function gate CMOS processes.

2.5.1 Dual Metal Gate CMOS Technology

Probably the most intuitive approach to developing a metal gate CMOS process involves the use of two metals, one serving as the NMOS gate and the other as the PMOS gate. Process integration is however not very straightforward since two film deposition steps and atleast two etch steps would be involved.

The general process involves the deposition of one metal over the entire substrate after active area definition and gate dielectric deposition. Following the first deposition, a well lithography is performed to expose either n- or p-well regions and the exposed metal is removed using a wet etch chemistry. For example, if the first metal is appropriate for NMOS gates, then it is retained over the p-well and removed from the n-wells and vice-versa for a PMOS gate metal. The choice of which metal to deposit first depends on the relative ease of removal of the two metals and the selectivity of the etch chemistries on the gate dielectric. Following the removal of the first metal, the photoresist is stripped and after rinsing the wafer, a second metal is deposited over the substrate. In doing so, a bilayer metal stack is formed on one half of the substrate, while a single layer gate is formed over the other half and each well has a different metal in direct contact with the gate dielectric. In order to prevent interdiffusion between the two metals on the side where a bilayer is formed, a thin diffusion barrier metal (< 5 nm) can be deposited over the first metal before patterning. Figure 2.3 (a) shows a schematic cross-section through the gate stacks after the two deposition steps. Following the second deposition, photolithography is used to define the gate lines on the N- and PMOS sides.

The process described above was demonstrated using Mo as the PMOS gate electrode and Ti as the NMOS gate electrode [2] and it was shown that normal output characteristics can be obtained using such a dual metal gate process. It is surmised that an

optimized process can lead to the fabrication of high performance N- and PMOSFETs with appropriate threshold voltages and acceptable short-channel performance. It is however important to examine the merits of this process and identify potential integration concerns.

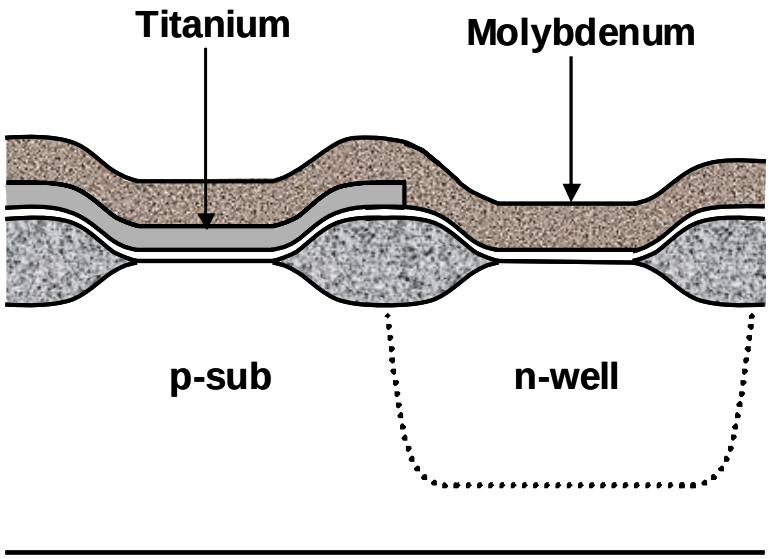


Fig. 2.3 Schematic cross-section through n- and p-channel transistors on a single Si substrate where the n-channel gate is Ti and the p-channel gate is Mo. Subsequent processing would involve photolithography to define the gate lines followed by metal etching on the N- and PMOS sides.

One concern arises from the fact that the first gate metal is chemically removed over the gate dielectric on one of the wells. In doing so, it is likely that the chemicals used for the metal etch can also attack the thin gate dielectric film. This was also observed in the case of the Mo and Ti gate process where the Ti etchant also attacked the Si_3N_4 gate dielectric on the n-well. This discrepancy is generally intolerable as it would lead to substantial mismatch in the performance of the two types of devices. If a different combination of gate metals and dielectric is used, it might be possible to avoid this drawback, however, any exposure of the dielectric to metal etchants is likely to affect the long term reliability of the gate dielectric film and should be avoided as much as possible. Hence, it is unlikely that the process mentioned herein can be ported to manufacturing as described. One solution is to intentionally remove the gate dielectric entirely over the n-well (or p-well) after removing the first metal layer. This way the n-well (or p-well) active region is exposed and after appropriate cleaning procedures, a new gate stack can be deposited over that well. While this solution can provide for a more reliable gate dielectric film, it introduces substantial process complexity (removal of dielectric film over first metal film on one well).

Gate lithography of the asymmetrical gate stacks presents another concern since the optimal focus depths for the two wells can be substantially different. Another concern lies with developing a high selectivity reactive ion etch process for the asymmetrical gate stacks on the n- and p- wells. Since one well will have a bilayer metal stack and the other well will have only a single metal layer, a high etch selectivity is needed between the second gate metal and the gate dielectric.

2.5.2 Metal Interdiffusion Gate CMOS Technology

A technique that alleviates some of the abovementioned concerns was developed [30] and referred to as metal-interdiffusion-gate (MIG) CMOS process since it relies on interdiffusion between two (or more) metal films. This approach can be implemented as follows.

Following gate dielectric deposition, a first metal layer is deposited and immediately capped with a second metal layer. The first metal film is intentionally kept thin (≤ 10nm) while the second metal film can be much thicker. In the next step, resist lithography is used to expose and remove the top metal film over one of the two wells. For example, if the first metal has a work function corresponding to that for NMOS devices, then the second metal is removed over those areas (p-wells) and vice-versa. Following selective removal of the second metal, gate lithography can be performed to define gate lines on both the wells. After gate definition, an interdiffusion anneal is performed to interdiffuse the two metals in the bilayer. Depending on the choice of the metals in the pair, this anneal can lead two one of a several outcomes.

If the metals do not thermodynamically prefer to alloy over the temperature range used and simply prefer to *intermix*, the resulting bilayer gate will have a work function that corresponds to that of the mixture. The exact value is however difficult to predict and will depend on the extent of the intermixing (which in turn depends on the relative thicknesses of the two metal films, the interdiffusion coefficients, etc.).

Another outcome can lead to the diffusion of the top metal through the bottom layer and its segregation at the dielectric interface. This will lead to the threshold voltage of the resulting transistor being determined entirely by the segregated metal layer. In effect, the bottom layer can act as a sieve to allow the top metal to pass through. This approach was demonstrated with the use of Ni and Ti bilayer gate stacks [30] and was shown to lead to low

and symmetrical threshold voltage n- and p-channel transistors. Ni and Ti were chosen for this demonstration since they have complementary work functions (~5 V and ~4 V respectively). Ni was used as the top metal. Ni is a fast diffuser in Ti [32,33] and is also known to segregate at the SiO_2 interface. Even at the relatively low temperature used for interdiffusion (400°C), x-ray photoelectron spectroscopy (XPS) results indicated that interdiffusion was complete and elemental Ni was observed at the SiO_2 interface. This indicates that Ni is stable on SiO_2 up to 400°C and does not preferentially form metal silicides or oxides.

Another possible outcome is the preferential alloying of the two metals to form a binary alloy with a work function determined by the exact composition of the alloy. This approach was recently demonstrated using the Ru-Ta system [34]. Co-sputtering of several alloy compositions in the Pt-Ti-Ta system have also shown desirable work function values and the MIG approach can potentially be applied to that system as well [35]. It is thus important to note that the general MIG concept can also be extended to ternary systems.

The primary advantage of a MIG approach over the dual-metal gate approach described earlier lies in the protection of the gate dielectric at all times since the bottom metal layer is always present to protect the dielectric. Hence, reliability problems associated with the earlier approach can be prevented.

2.6 Summary

This chapter summarized the primary requirements for a viable metal gate technology including challenges in the selection of materials, important physical properties, process compatibility and interaction with other materials in the transistor gate stack. It is shown that two metals will ordinarily be required for a low and symmetric V_{TH} CMOS technology. Two

schemes developed to integrate two metals on a single Si substrate were developed and discussed in terms of their relative ease of integration. It is concluded that although the integration of multiple metals on a single Si substrate is possible, process integration can be quite difficult. This highlights the need for a single-metal tunable work function gate CMOS technology that will been discussed in the next chapter.

2.7 References

1. C.-L. Huang, N. Arora, A. Nasr, D. Bell, “Effect of polysilicon depletion on MOSFET I-V characteristics”, Electronics Letters, vol.29, 13, p.1208-9, June 1993
2. Y.-C. Yeo, Q. Lu, P. Ranade, H. Takeuchi, K. J. Yang, I. Polishchuk, T. -J. King, C. Hu, S.C. Song, H. F. Luan, D. -L. Kwong, “Dual-Metal Gate CMOS Technology with Ultra-Thin Silicon Nitride Gate Dielectric”, IEEE Electron Device Letters, v. 22, no. 5, p. 227, May 2001
3. Y. Taur, C. H. Wann and D.J. Frank, “25 nm CMOS Design Considerations”, International Electron Devices Meeting 1998, Technical Digest, p. 789, Dec. 1998
4. R. Chau, J. Kavalieros, B. Roberds, R. Schenker, D. Lionberger, D. Barlage, B. Doyle, R. Arghavani, A. Murthy, G. Dewey, “30 nm physical gate length CMOS transistors with 1.0 ps n-MOS and 1.7 ps p-MOS gate delays”, International Electron Devices Meeting 1998, Technical Digest, p. 45, Dec. 2000
5. E. P. Gusev et al, “Ultrathin high-K gate stacks for advanced CMOS devices”, International Electron Devices Meeting 1998, Technical Digest, p. 20, Dec. 2001
6. M. Copel, E. Cartier, E. P. Gusev, S. Guha, N. Bojarczuk, M. Poppeller, “Robustness of ultrathin aluminum oxide dielectrics on Si (001)”, Applied Physics Letters, vol.78, no.18, p. 2670, April 2001
7. S. B. Samavedam, et al, “Metal gate MOSFETs with HfO_2 gate dielectric”, 2002 Symposium on VLSI Technology, Digest of Technical Papers, p.24, June 2002
8. R. K. Sharma, A. Kumar, J. M. Anthony, “Advances in high-k dielectric gate materials for future ULSI devices”, Journal of Materials, vol.53, no.6, p. 53, June 2001
9. K. Onishi, L. Kang, R. Choi, E. Dharmarajan, S. Gopalan, Y. Jeon; C. S. Kang, B. H. Lee; R. Nieh, J. C. Lee, “Dopant penetration effects on polysilicon gate HfO_2 MOSFETs”, 2001 Symposium on VLSI Technology, Digest of Technical Papers, p.131, June 2001
10. International Technology Roadmap for Semiconductors, Semiconductor Industry Association, 2000
11. S. M. Sze, “Physics of Semiconductor Devices”, 2nd Edition, Wiley, 1981

12. I. De, D. Johri, A. Srivastava, C. M. Osburn, “Impact Of Gate Workfunction on Device Performance at the 50 nm Technology Node”, Solid-State Electronics, vol.44, no.6, p. 1077, June 2000
13. Y.-K. Choi, Y. -C. Jeon, P. Ranade, H. Takeuchi, T. -J. King, J. Bokor, C. Hu, “Ultra-thin body SOI Device with 30nm Gate Length and Selectively Raised Ge S/D”, Conference Proceedings, IEEE 58th Device Research Conference, p.23, June 2000
14. X. Huang, W.-C. Lee, C. Kuo, D. Hisamoto, L. Chang, J. Kedzierski, E. Anderson, H. Takeuchi, Y.-K. Choi, K. Asano, V. Subramanian, T.-J. King, J. Bokor and C. Hu, “Sub-50 nm P-channel FinFET”, IEEE Transactions on Electron Devices, vol.48, no.5, p. 880, May 2001
15. L. Chang, S. Tang, T.-J. King, J. Bokor and C. Hu, “Gate Length Scaling and Threshold Voltage Control of Double-Gate MOSFETs”, International Electron Devices Meeting 2000, Technical Digest, p. 719, Dec. 2000
16. Y.-C. Yeo, P. Ranade, T.-J. King and C. Hu, “Effects of high-κ gate dielectric materials on metal and silicon gate workfunctions”, IEEE Electron Device Letters, vol.23, no.6, p. 342, June 2002
17. A. Yagishita, et al, “High performance metal gate MOSFETs fabricated by CMP for 0.1 μm regime”, International Electron Devices Meeting 1998, Technical Digest, p. 785, Dec. 1998
18. K. J. Hubbard, D. G. Schlom, “Thermodynamic stability of binary oxides in contact with silicon”, Journal of Materials Research, vol.11, no.11, p. 2757, Nov. 1996
19. A. Kawamoto, J. Jameson, K. Cho and R. W. Dutton, “Challenges for atomic scale modeling in alternative gate stack engineering” IEEE Transactions on Electron Devices, vol.47, no.10, p.1787, Oct. 2000
20. W.-J. Qi et al, “MOSCAP and MOSFET characteristics using ZrO_2 gate dielectric deposited directly on Si”, International Electron Devices Meeting 1998, Technical Digest, p. 145, Dec. 1999
21. W.-J. Qi et al, “High-K gate dielectrics”, Proceedings of the SPIE, vol.3881, p.24, Sep. 1999
22. T. Amazawa and H. Oikawa, “Surface state generation of Mo gate metal oxide semiconductor devices caused by Mo penetration into gate oxide”, J. Electrochem. Soc. 145, 1297, 1998

23. P. Lundgren, “Impact of the gate material on the interface state density of metal-oxide-silicon devices with an ultrathin oxide layer”, J. Appl. Phys., v. 85, no. 4, 2229, 1999
24. T. Ushiki, et al, “Gate oxide reliability concerns in gate-metal sputtering deposition process: an effect of low-energy large-mass bombardment”, Microelectronics Reliability, v. 39, no. 3, p. 327, 1999
25. D.-G. Park, et al, “Effects of TiN deposition on the characteristics of $W/TiN/SiO_2/Si$ metal oxide semiconductor capacitors”, J. Electrochem. Soc. 148, no. 9, F189, Sep. 2001
26. D.A. Antoniadis, “MOSFET Scalability Limits and New Frontier Devices”, 2002 Symposium on VLSI Technology, Digest of Technical Papers, p.2, June 2002
27. J. Tersoff, “Theory of semiconductor heterojunctions: The role of quantum dipoles”, Physical Review B, vol. 30, p. 4874, Oct. 1984
28. F. Capasso and G. Margaritondo, Eds., “The theory of heterojunction band lineups”, in Heterojunction Band Discontinuities, Amsterdam, The Netherlands, Elsevier, 1987
29. S. G. Louie and M. L. Cohen, “Electronic structure of a metal-semiconductor interface”, Physical Review B, vol. 13, p. 2461, Mar. 1976
30. W. Mönch, “Electronic properties of ideal and interface-modified metal-semiconductor interfaces”, Journal of Vacuum Science and Technology B, vol. 14, p. 2985, July/August 1996
31. I. Polishchuk, P. Ranade, T. -J. King and C. Hu, “Dual Work Function Metal Gate CMOS Transistors by Ni-Ti Interdiffusion”, IEEE Electron Device Letters, v. 23, no. 4, p. 200, April 2002
32. Askill, Gibbs, “Tracer Diffusion in β-Titanium”, Phys. Stat. Solidi, v. 11, p. 559, 1965
33. G. B. Gibbs, D. Graham, D. H. Tomlin, Phil. Mag, 8, p. 1269, 1963
34. V. Misra, H. Zhong, H. Lazar, “Electrical properties of Ru-based alloy gate electrodes for dual metal gate Si-CMOS” IEEE Electron Device Letters, v.23, no.6, p.354, June 2002 (to be replaced with the IEDM reference)
35. B.-Y. Tsui, et al, “Wide Range Work Function Modulation of Binary Alloys for MOSFETs Application”, IEEE Electron Device Letters (to be published)

Chapter 3

Metal Gate Work Function Engineering

3.1 Introduction

The work function of a metal depends on several physical and chemical properties. The conventional definition of the work function of a metal quantifies it in terms of the energy required to remove an electron from the metal surface [1]. Implicit in this definition is the idea that the energy required to remove an electron from a metal is primarily a property of the surface and specifically a property of the interface between the metal and its surrounding medium. From a purely thermodynamic perspective, the work function of a metal can also be defined in terms of the chemical potential within the metal (the Fermi level is in fact one manifestation of the Gibbs free energy of a metal). From a thermodynamic perspective then, structural and/or chemical changes at a metal surface are likely to change the free energy at that surface and hence affect the work function. The notion that structural or chemical changes in a thin metal film are likely to affect its work function is one of the primary hypotheses of this work. This chapter will discuss various approaches developed to modulate thin film work functions and present experimental results demonstrating the feasibility of some of these approaches.

3.2 Approaches to Work Function Engineering

Three general approaches to work function engineering are proposed as shown in Fig. 3.1 and involve changes to the structure, the chemistry or to both the structure and the

chemistry of thin metal films. Techniques are described in this chapter for the implementation of some of these approaches.

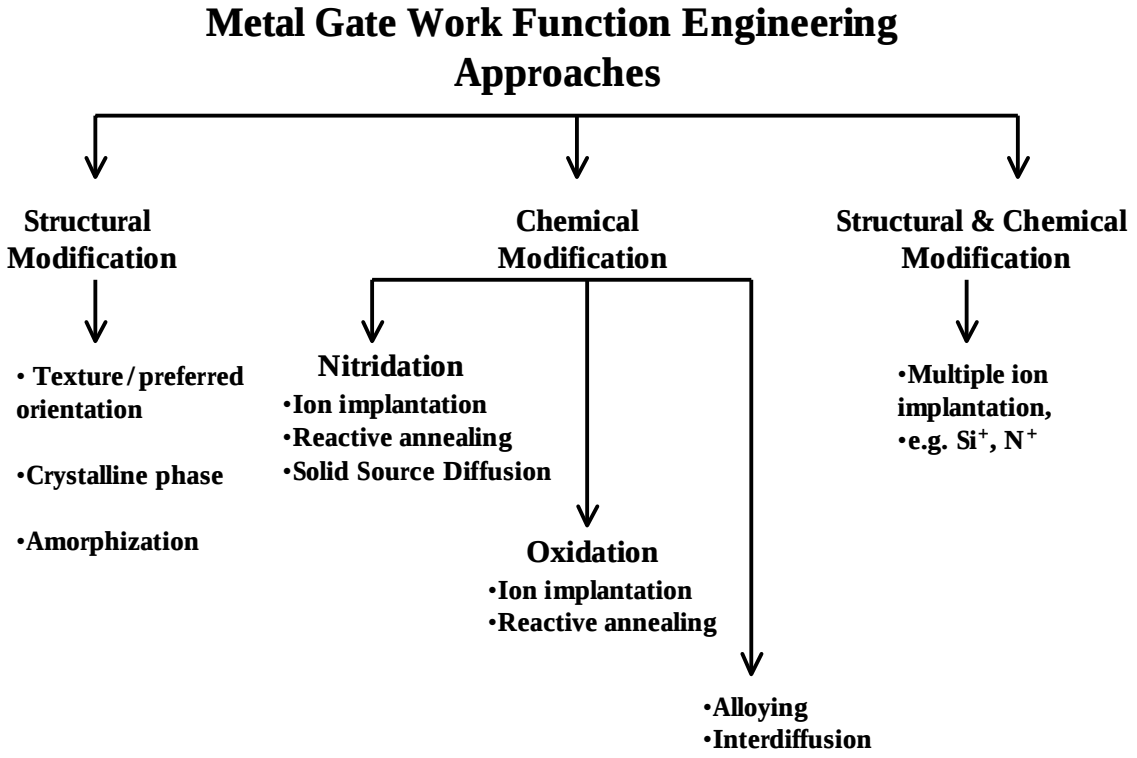


Fig. 3.1. Schematic indicating the general approaches to work function engineering of thin metal films.

3.2.1 Structural Modification

As shown in Fig. 3.1, changes to the structural morphology of thin metal films can be further classified into modification to film texture or preferred orientation, modification to crystalline phase and modification (disruption) in the long range order within the crystal lattice. These approaches are demonstrated using thin molybdenum and tungsten films.

In general, metals display anisotropy in their work function [2]. The origin of this anisotropy is believed to be the difference in atomic packing density along the various crystallographic planes. Densely packed crystallographic surfaces display high work functions

since these surfaces are smooth and relatively inert with fewer broken atomic bonds (most bonds are within the plane). Open crystallographic surfaces display lower work functions owing to a greater number of broken bonds (larger fraction of bonds outside the plane). Smoluchowski [3] has developed a formal semi-quantitative theory to explain this anisotropy based on a similar crystallographic analysis. Amorphous metal films are expected to have lower work functions than their crystalline counterparts.

Single crystalline Mo has a work function in vacuum that is highly anisotropic and varies from ~ 4.4 V for the relatively open and less dense (112) crystallographic planes to ~ 4.95 V for the closest packed and smoothest (110) planes in the body-centered cubic structure [1]. Figure 3.2 shows the variation of the Mo work function with crystalline orientation.

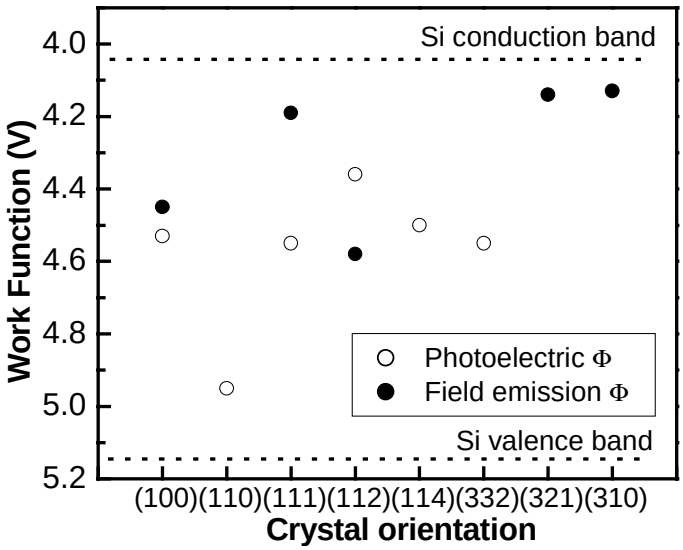


Fig. 3.2 Variation of Mo work function with crystal orientation [2]. The (110) , (321) and (310) work functions are best suited for bulk Si PMOSFET and NMOSFET gate electrode applications respectively, while some of the other planes have work functions suitable for UTB or double-gate FET gate electrode applications.

While most metals display a modest variation in the work function (between 0.2-0.3 V), Nb and Mo display a relatively wide variation (~ 0.7 V). This study focuses primarily on Mo, owing to a lack of information on the thermal stability of Nb with SiO_2 and a general incompatibility with CMOS processing. In addition to the magnitude of the variation, it is also the range of work function values obtained that make Mo attractive for CMOS applications. As discussed in Chapter 2, advanced FD-SOI based CMOS devices would require gate work functions within ± 0.2 V of E_I (4.4-4.8 V) for NMOS and PMOS devices [4] while sub-70 nm bulk-Si CMOS devices would require gate work functions between ± 0.2 V of E_C (E_V) for NMOS (PMOS) transistors respectively [5]. Thus, Mo appears to be a promising gate material if thin film microstructure can be engineered through control of deposition techniques and variables or the use of post deposition structural modification techniques.

Mo gate capacitors were fabricated on thermally oxidized p-type Si wafers. In order to account for the influence of oxide fixed charge, multiple SiO_2 thicknesses were obtained on a single substrate by selective etchback using dilute HF solution. Thin Mo films (15 nm) were sputter deposited on the thermally oxidized substrates. Metal deposition was performed in a DC magnetron sputtering tool with a base pressure of 1×10^{-7} T and Ar plasma and a chamber pressure of 6 mT. Substrate temperature during deposition was maintained at 200°C and plasma power used was 100 W. Dropout lithography was used to open up portions of the wafer which were then implantated with inert ions (Ar^+, 3 keV, $1\times10^{16}cm^{-2}$). This implant effectively amorphizes the microstructure as shown in the x-ray diffraction (XRD) results of Fig. 3.3. The flat-band voltages (V_{FB}) of the capacitors were determined from *C–V* measurements (Fig. 3.4) and used to extract the gate work functions using the

mathematical relations described in Chapter 2. Figure 3.5 shows the variation of V_{FB} with oxide thickness (T_{OX}). It can be seen that Ar^+ implantation lowers the Mo work function by 0.8 V (to 4.0 V), indicating that amorphous Mo has a lower work function. The wafer was then annealed at 500°C for 20 min and the measurements were repeated. The x-ray diffraction (XRD) analysis shown in Fig. 3.3 also indicates that the Mo film is crystallized after the annealing as is evident by the appearance of the 110 diffraction peak. As shown in Fig. 3.5, the Mo work function after implantation and annealing matches that of unimplanted and annealed Mo (4.95 V) and reflects this change in the film morphology.

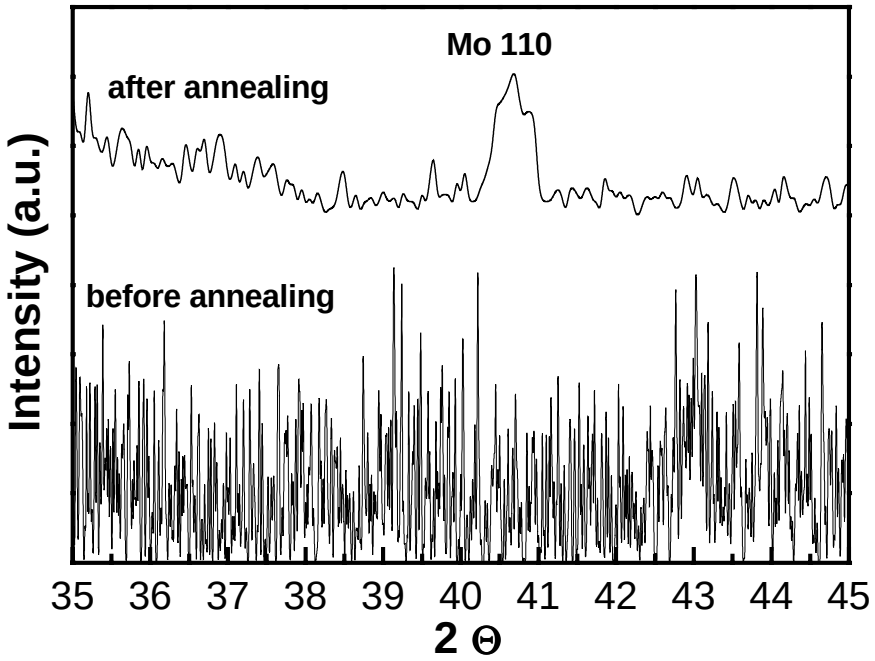


Fig. 3.3 XRD analysis on Ar implanted Mo films before and after annealing. Appearance of the 110 Mo peak is indicative of crystallization of the Mo film upon annealing

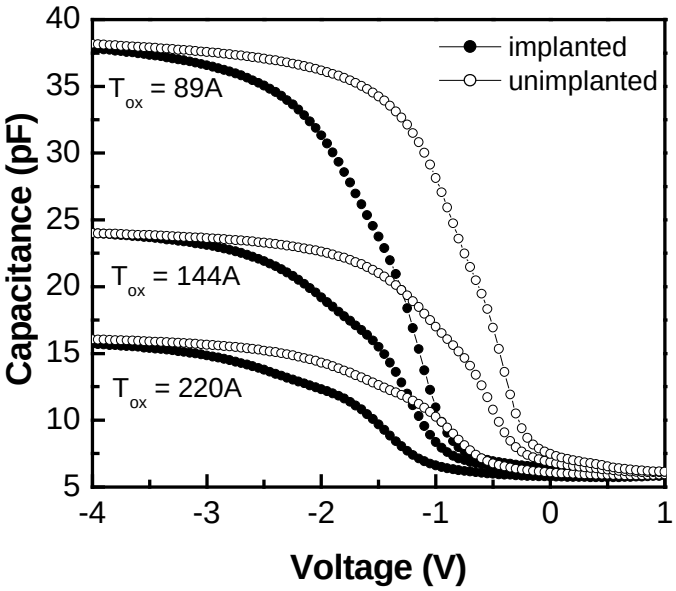


Fig. 3.4 *C-V* curves for $Mo/SiO_2/Si$ capacitors with and without Ar implantation. The negative shift in the curves indicates a lowering of the flat band voltage (V_{FB}) after Ar implantation

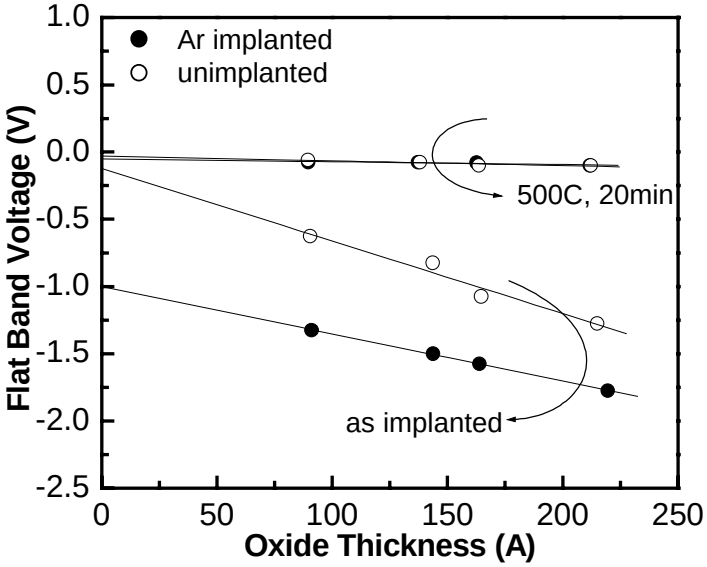


Fig. 3.5 V_{FB} vs. T_{OX} relationship with and without Ar implantation. The y-intercept indicates Φ_{MS}. The work functions of the implanted and unimplanted films before annealing are 4 eV and 4.8 V respectively, while they are identical (4.95 V) after annealing

To evaluate the impact of deposition parameters on the morphology of Mo films, the sputtering pressure and temperature of the chuck were varied and the deposited films were

analyzed using x-ray diffraction. Figure 3.6 shows the effect of deposition temperature on the morphology of Mo films. As the deposition temperature was increased from 25°C to 500°C, the intensity of the 110 peak of Mo increased significantly. This indicates that the 110 orientation of Mo is the energetically favored one. Body-centered-cubic (bcc) metal films generally deposit in the (110) orientation.

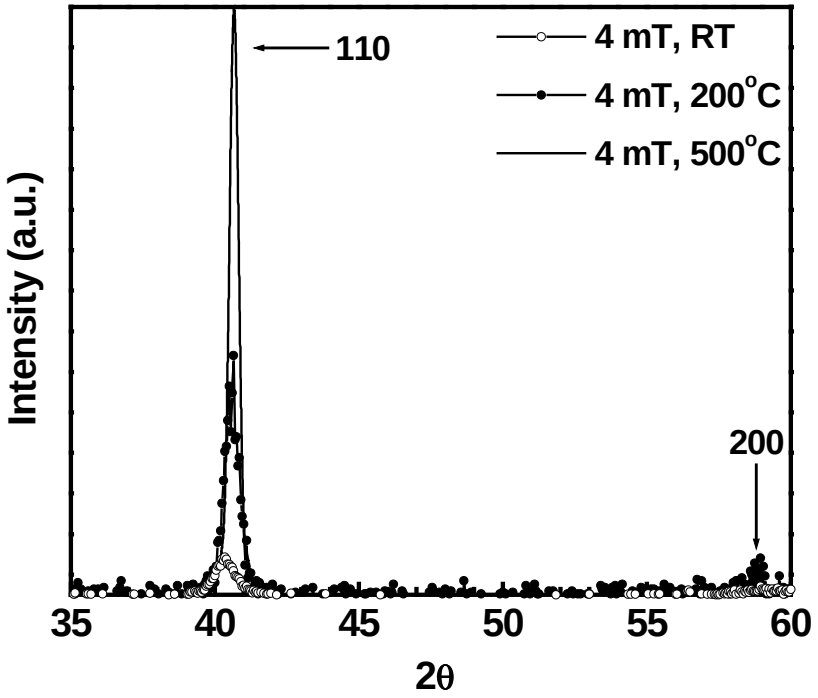


Fig. 3.6 XRD analysis of Mo thin films deposited on thermally oxidized Si substrates. The films were 50 nm thick and deposited at three different temperatures. As seen in the figure, increasing the sputter deposition temperature, increases the intensity of the 110 peak

Figure 3.7 shows the impact of deposition pressure (Ar plasma) on the morphology of the films. While increasing the pressure from 1 mT to 4 mT suppresses the 110 peak substantially, increasing the pressure to 20 mT suppresses it almost entirely while simultaneously exhibiting a 222 peak. This indicates a change in the preferred orientation of the film with increase in the deposition pressure.

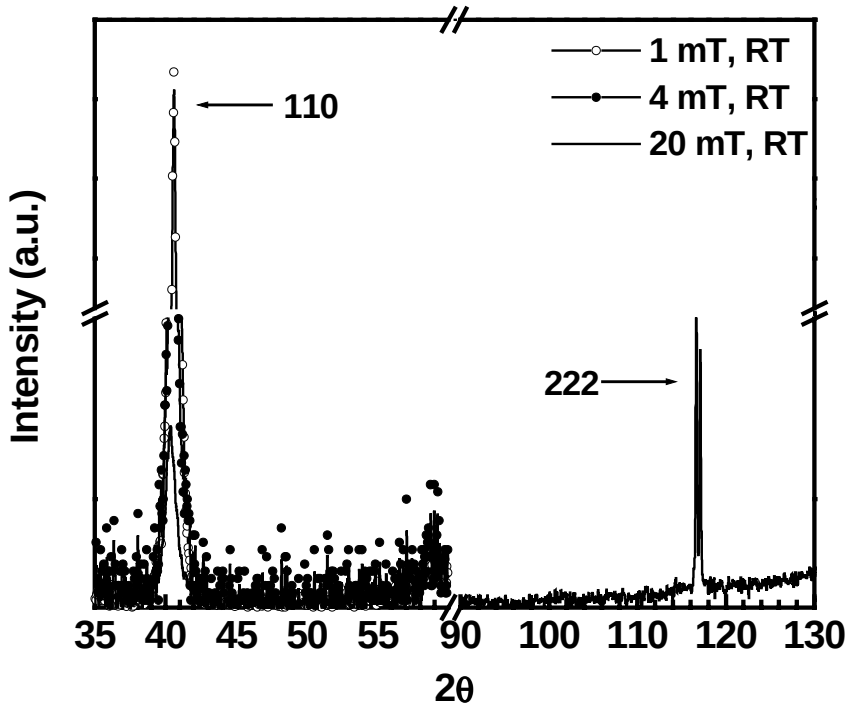


Fig. 3.7 XRD analysis of Mo thin films deposited on thermally oxidized Si substrates. The films were 50 nm thick and deposited at three different Ar plasma pressures. As seen in the figure, increasing the sputter deposition pressure suppresses the 110 peak while enhancing the 222 peak intensity

Similar experiments were performed with tungsten films instead of molybdenum to evaluate the effect of deposition parameters on film microstructure. As shown in Fig. 3.8, temperature has a smaller impact on the preferred orientation of W. Conversely, varying the deposition pressure leads to a preferential nucleation of face centered cubic (β) W instead of the low temperature body centered cubic (α) W as shown in Fig. 3.9. It is interesting to note that while the impact of pressure on Mo morphology is a change in the preferred orientation, the predominant effect on W is a change in the crystalline phase.

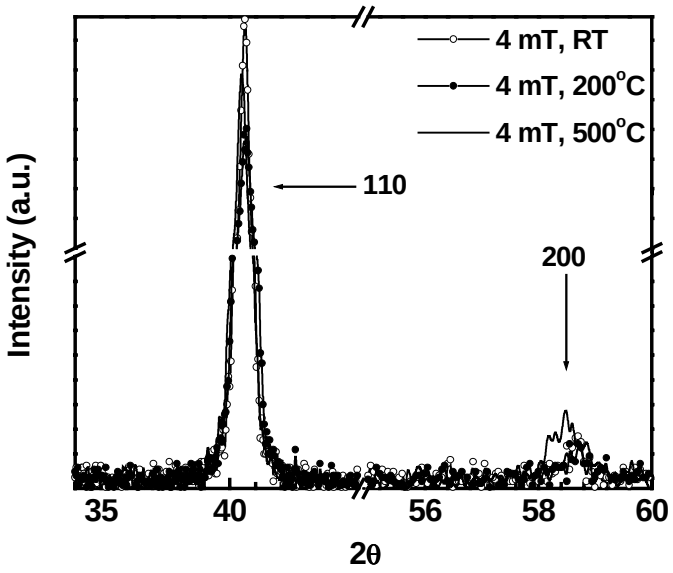


Fig. 3.8 XRD analysis of W thin films (thickness = 50 nm) deposited on thermally oxidized Si substrates. The films were deposited at three different temperatures. As seen in the figure, increasing the sputter deposition temperature does not significantly change the morphology of the W film (unlike the case of Mo).

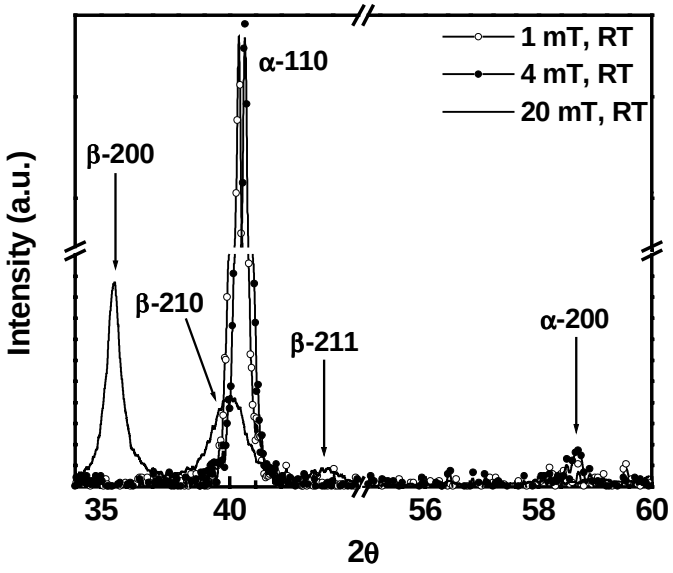


Fig. 3.9 XRD analysis of W thin films (thickness = 50 nm) deposited on thermally oxidized Si substrates. The films were deposited at three different Ar plasma pressures. As seen in the figure, increasing the sputter deposition pressure changes the preferred nucleation phase from α-W to β-W.

Figure 3.10 shows the *C-V* characteristics for the two W films. It can be seen that the flat band voltage for β-W is ~ 0.2 V lower than that for α-W. This also implies a proportional lowering in the work function with phase. The work function of α-W is measured to be ~ 4.6 V while that of β-W is ~ 4.4 V. These results indicate that with suitable modification, W films are potential candidates for fully-depleted SOI (FD-SOI) NMOS transistors.

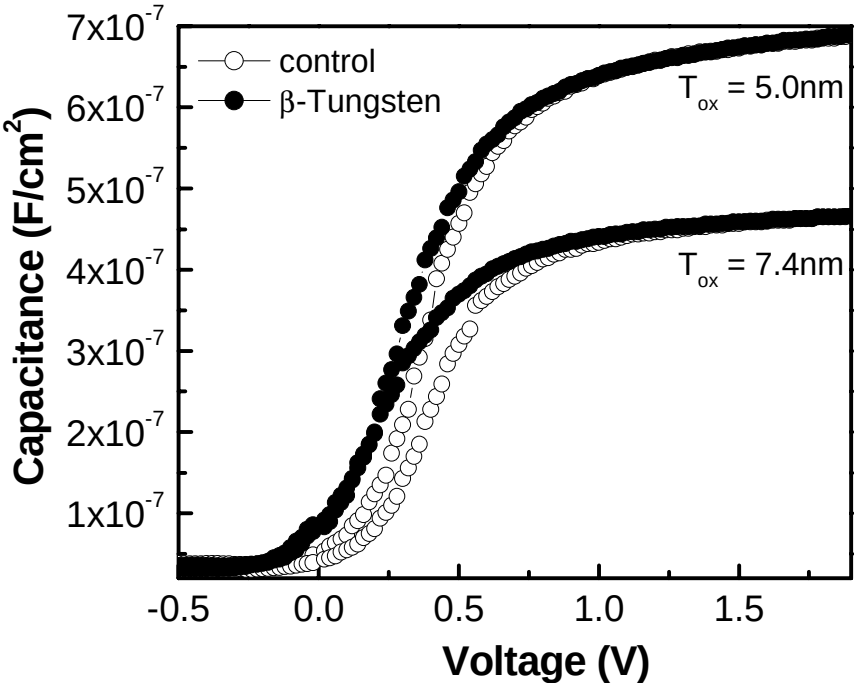


Fig. 3.10 *C-V* curves for $W/SiO_2/Si$ capacitors where the W films (50 nm) were deposited at different Ar plasma pressures (1 mT for the control devices and 20 mT for the β-W devices). The negative shift in the curves upon increasing the deposition pressure indicates a lowering of the flat band voltage (V_{FB}) and hence a lower work function for the β-W.

3.2.2 Chemical Modification

There are several ways to selectively change the chemistry of thin metal films deposited on Si substrates as shown in Fig. 3.1. It should be noted however, that the final chemistry of the film should be such that the film retains near metallic conductivity. As such, conducting

metal nitrides, metal silicides and conducting metal oxides are potential candidates. This section describes a few techniques developed to implement this approach.

A very powerful yet straightforward way to introduce foreign elements into a film is ion implantation. Ion implantation is an attractive approach for a number of reasons. If the work function can indeed be modulated over the desired range, ion implantation can provide for a highly integrable and CMOS compatible process since photoresist masking can be used to selectively implant the desired ions. In addition, ion implantation is a technique that has been used in CMOS processing for decades and has achieved a high level of sophistication (very precise control of implantation parameters like beam currents, voltages and tilts is possible). In addition, fairly sophisticated simulators are available to model implanted depth profiles in a variety of substrates with fairly high precision. The use of inert ion implantation to modify the structure of thin Mo films has already been described in the previous section. An alternative approach which modifies the chemistry of the deposited Mo films involves the implantation of N^+ ions instead of inert Ar^+ ions. The experiment described earlier was repeated with implantation of N^+ ions (1 keV – $5\times10^{15}cm^{-2}$; 2 keV – $5\times10^{15}cm^{-2}$, $1\times10^{16}cm^{-2}$). Very thin Mo films and ultra-low energy implantation were used in order to minimize ion penetration into the gate dielectric (by lowering the projected range and straggle (ΔR_p)) while allowing for high atomic N content within the Mo films. *SRIM* simulations were used to estimate the N concentration-depth profiles in the $Mo/SiO_2/Si$ stacks (Fig. 3.11). A dose of $5\times10^{15}cm^{-2}$ is estimated to introduce ~5 % N in the 15 nm Mo film. The evolution of the Mo work function (extracted from flat-band voltage measurements as described earlier) with annealing temperature is shown in Fig. 3.12. The work function of unimplanted Mo films increased slightly upon annealing, but is fairly stable at ~5 V. This increase can be attributed to an improvement in the crystalline quality of the film (increase in the columnar grain size).

It should be noted however, that the final work function is in agreement with the published values for the (110) Mo work function. As shown in the XRD analysis in this chapter, the as deposited Mo film has a preferred (110) orientation.

The work function of the N implanted films continues to decrease with increasing annealing temperature, saturating at ~4.4 V at an annealing temperature between 800-900°C. The origins of this change can be traced to a simultaneous evolution of the Mo film morphology and chemistry at the dielectric interface.

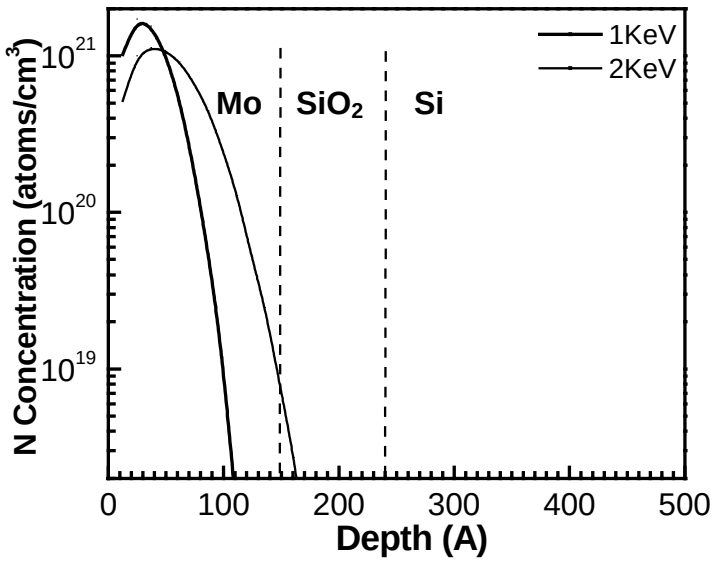


Fig.3.11 Simulated N concentration-depth profiles in the $Mo/SiO_2/Si$ stack (Dose = $1\times10^{15}cm^{-2}$). The use of low implant energies enables high peak concentrations of N to be obtained in the film while preventing significant penetration of the N ions into the gate dielectric.

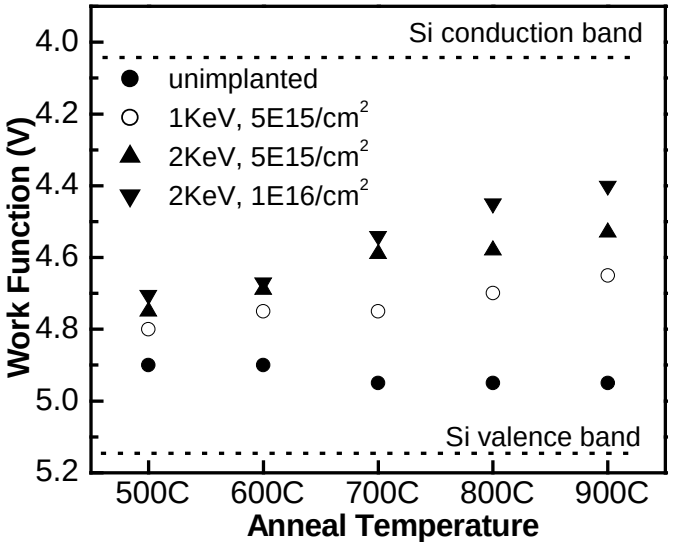


Fig. 3.12 Variation of Mo work function with thermal annealing. All anneals were 15 min long except for the 900°C anneal (15 s)

Figure 3.13 (a) shows the results of XRD analysis performed on N^+ implanted and control samples. The N^+ implanted samples show several diffraction peaks corresponding to Mo_2N. The implanted samples show several diffraction peaks corresponding to Mo_2N. Figure 3.13 (b) compares the intensities for the 110 and 112 peaks and indicates the predominantly (110) texture of the Mo film. Nitrogen implanted into $Mo/SiO_2/Si$ stacks has been observed to segregate at the Mo/SiO_2 interface upon annealing [6]. The nitride formation is thus most likely restricted to the interfacial regions. Figures 3.14 (a-d) show the details of diffraction peaks highlighted in Fig. 3.13 (a).

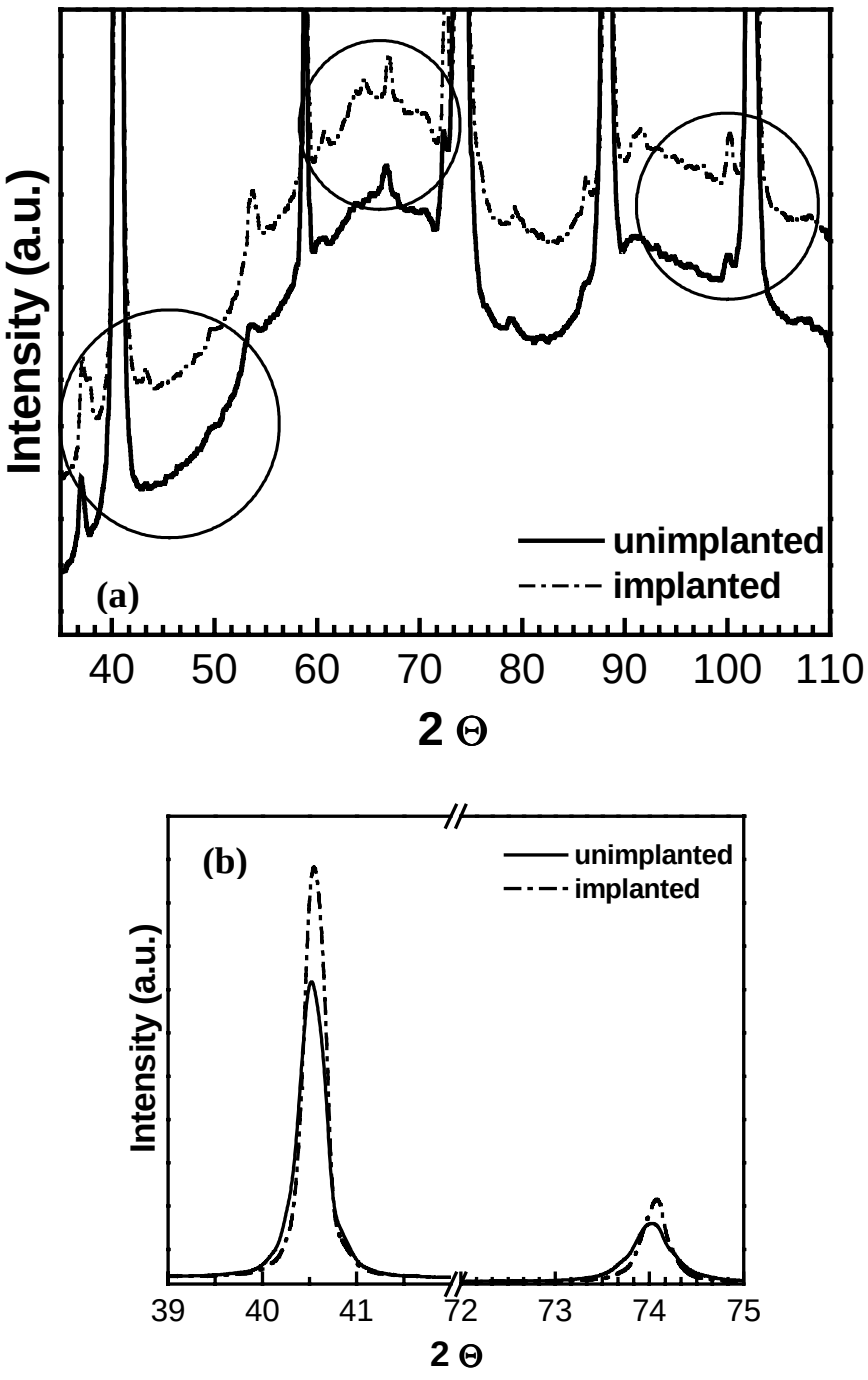


Fig. 3.13 XRD analysis of N implanted and unimplanted Mo films. Both samples received a 900°C anneal (15 s). Implant dose = $1\times10^{16}cm^{-2}$. A syncrotron x-ray source was used for the analysis. (a.) The entire range of peaks, (b.) Comparison of the 110 and 112 peak intensities indicating the predominantly (110) texture of the film

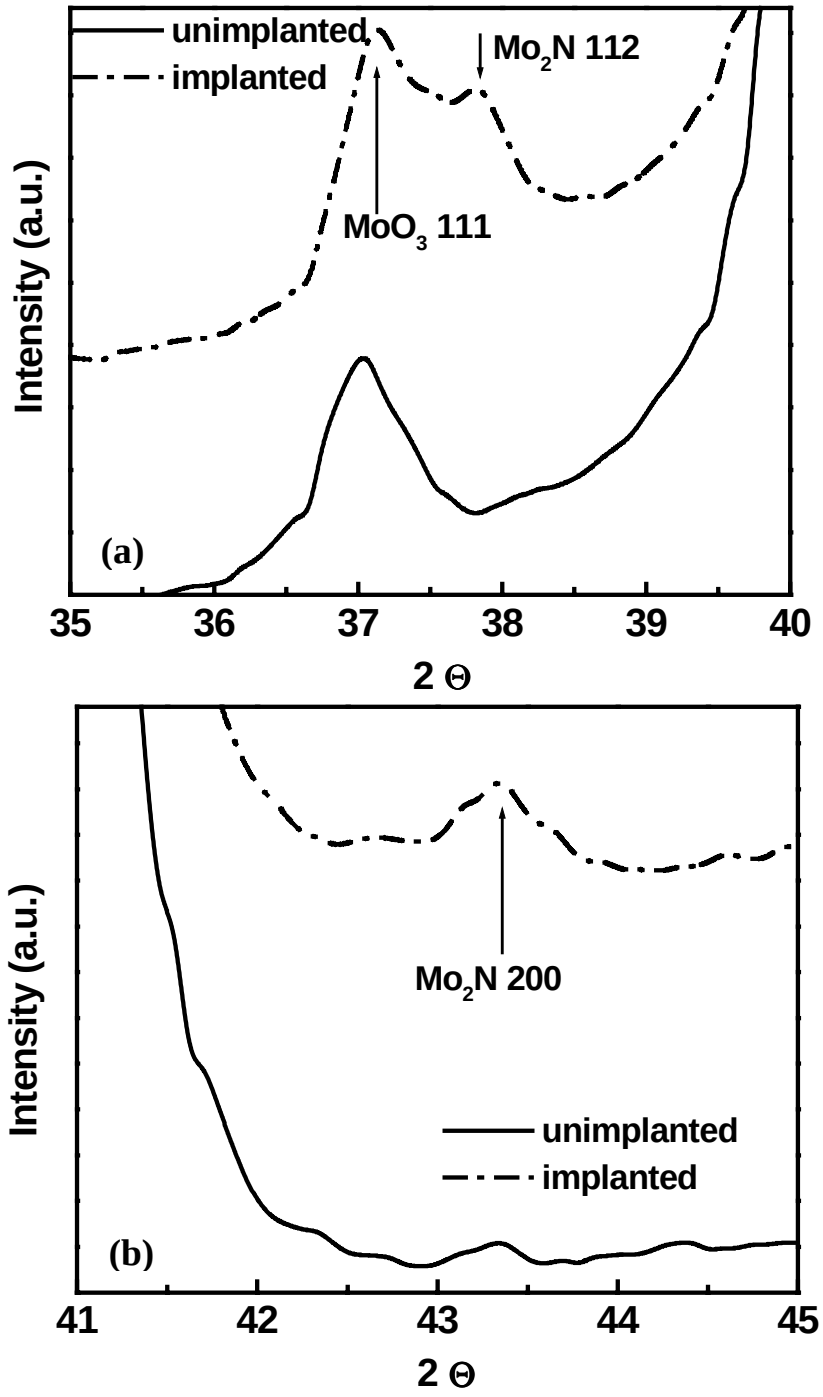


Fig. 3.14 (a-b). Details of XRD peaks highlighted in Fig. 3.13 (a.). Peaks indicative of Mo_2N formation are observed in the N implanted sample while no such peaks are observed in the control sample.

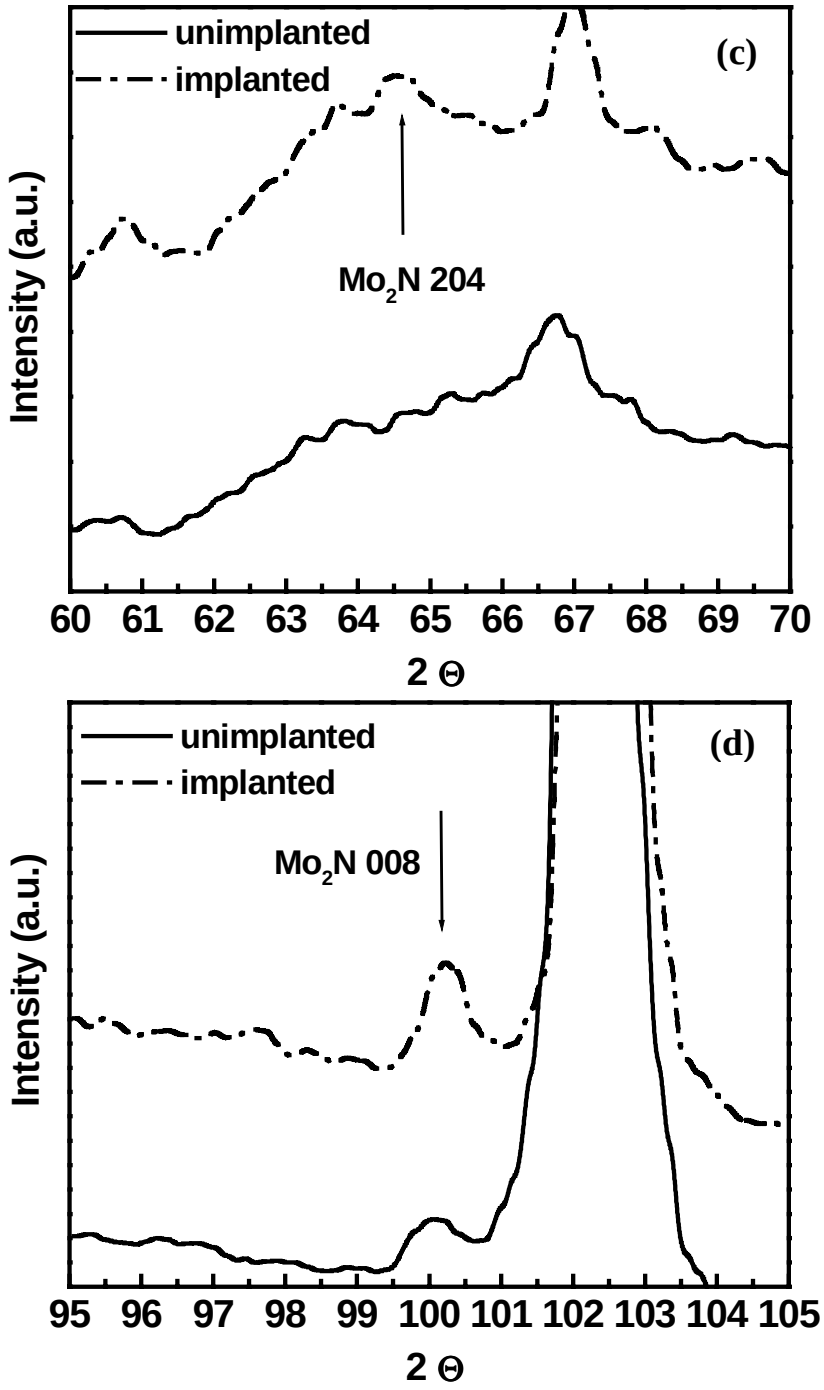


Fig. 3.14 (c-d). Details of XRD peaks highlighted in Fig. 3.13 (a.). Peaks indicative of Mo_2N formation are observed in the N implanted sample while no such peaks are observed in the control sample.

Figures 3.15 (a-b) show the impact of implant dose and annealing temperature on the diffraction peak intensity. It can be seen that increasing the dose or temperature leads to a slight increase in the Mo_2N diffracted peak intensity.

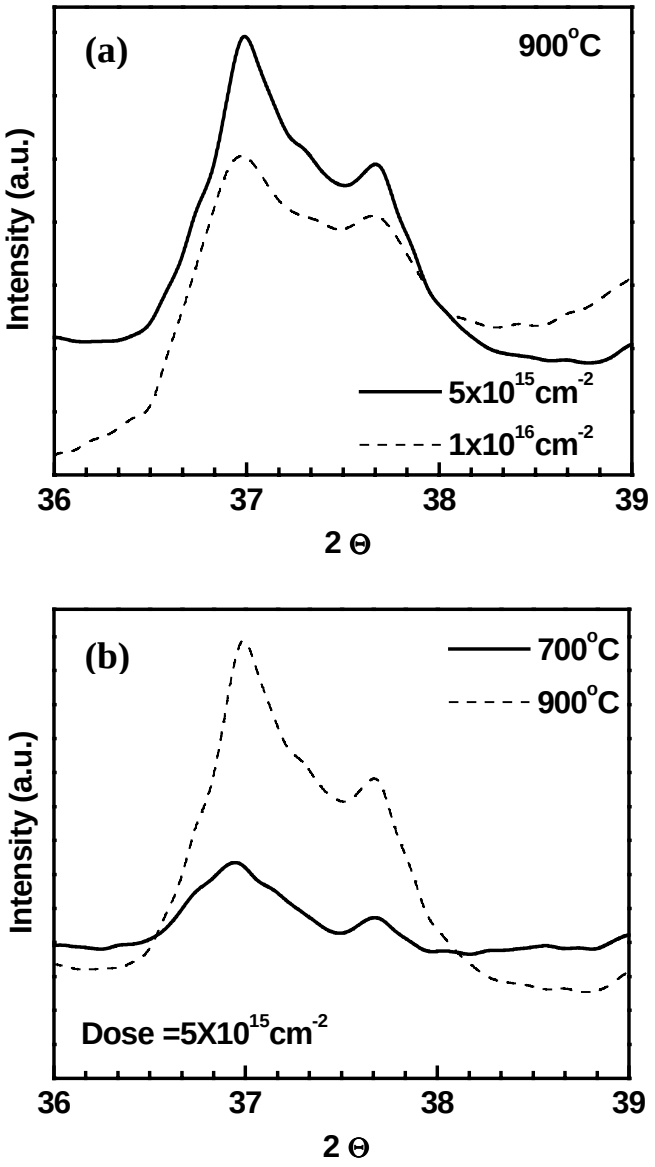


Fig. 3.15 (a-b). The varation of the Mo_2N 112 peak intensity with N^+ implant dose and annealing temperature

To further analyze the chemical changes in the N^+ implanted Mo films, x-ray photoelectron spectroscopy (XPS) analysis was performed. The results of XPS analysis on a sample after

900°C annealing are summarized in Fig. 3.16 which plots the electron count vs binding energy for Mo and N signals at several points through the thickness of the Mo film. There is a significant overlap between the Mo 3p peak and the N 1s peak since they are separated by ~4 eV. Hence, the overall peak was deconvolved to isolate the Mo and the N contributions. It can be seen that a fairly strong N signal accompanies the Mo signal throughout the thickness of the Mo film. As observed by Wei et al [7], the Mo valence state in the nitride cannot be calculated directly from the chemical formula Mo_2N; rather a combination of valence states was observed in the nitrided Mo films.

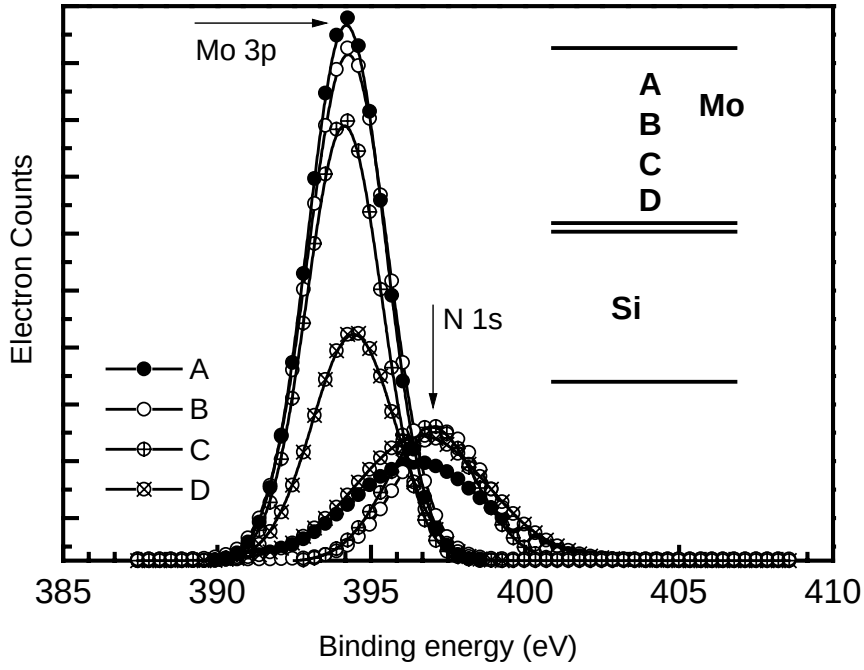


Fig. 3.16. Mo 3p and N 1s binding energy at different positions across the Mo film thickness. (Mo film thickness = 15 nm, Implant dose = $1\times10^{16}cm^{-2}$, 900°C, 15 s RTA)

An alternative way to introduce nitrogen into metal films is to use a sacrificial layer as a solid diffusion source [8]. In this approach, thin metal films (50 nm, Mo or W) were deposited onto thermally oxidized Si substrates. The elemental metal films were then capped with 100

nm over-stoichiometric TiN (TiN_{1+x}). The over-stoichiometric films were prepared by sputtering a Ti target using a N/Ar partial pressure ratio 3 times higher than that used for stoichiometric TiN deposition. The motivation behind this approach is that excess N in these films can (and most likely will) diffuse out into the underlying metal (Mo or W) films upon thermal annealing. The TiN films can thus be used as solid diffusion sources to introduce N into the metal films. Figure 3.17 (a-b) shows the N depth profiles across the $TiN/Mo/SiO_2$ and $TiN/W/SiO_2$ stacks obtained using Secondary Ion Mass Spectroscopy (SIMS). The depth profiles for the as deposited films and those after two different annealing conditions are shown.

Several distinctions between the diffusivity of N into W and Mo are evident from an analysis of Fig. 3.17. It should be mentioned however, that these SIMS profiles should not be construed as quantitative evidence of the diffusion phenomena since it was difficult to calibrate the sputter etch rates and the peak intensities accurately in these stacks owing to a lack of standard reference samples. Qualitatively, the extent of nitrogen diffusion is seen to be much faster in Mo than in W. At 800°C, N does not diffuse through the thickness of the W film, while it diffuses completely into the Mo film and a slight pileup at the SiO_2 interface is evident. Interestingly, a comparison between the W and Mo profiles into the TiN layer indicate that W diffusion into TiN is faster than that of Mo.

The solubility and diffusivity of N in Mo at high temperatures (> 1000°C) have been evaluated earlier [9-10]. While the relevant data for lower temperatures is not readily available, the extrapolated N diffusion coefficient at 900°C is very high ($\sim 2\times10^{-8} cm^2/s$). Low temperature diffusivities in poly-crystalline thin films are expected to be much higher than the high temperature extrapolated values since grain boundary diffusion (lower activation energy) would tend to dominate over bulk diffusion in the low temperature regime.

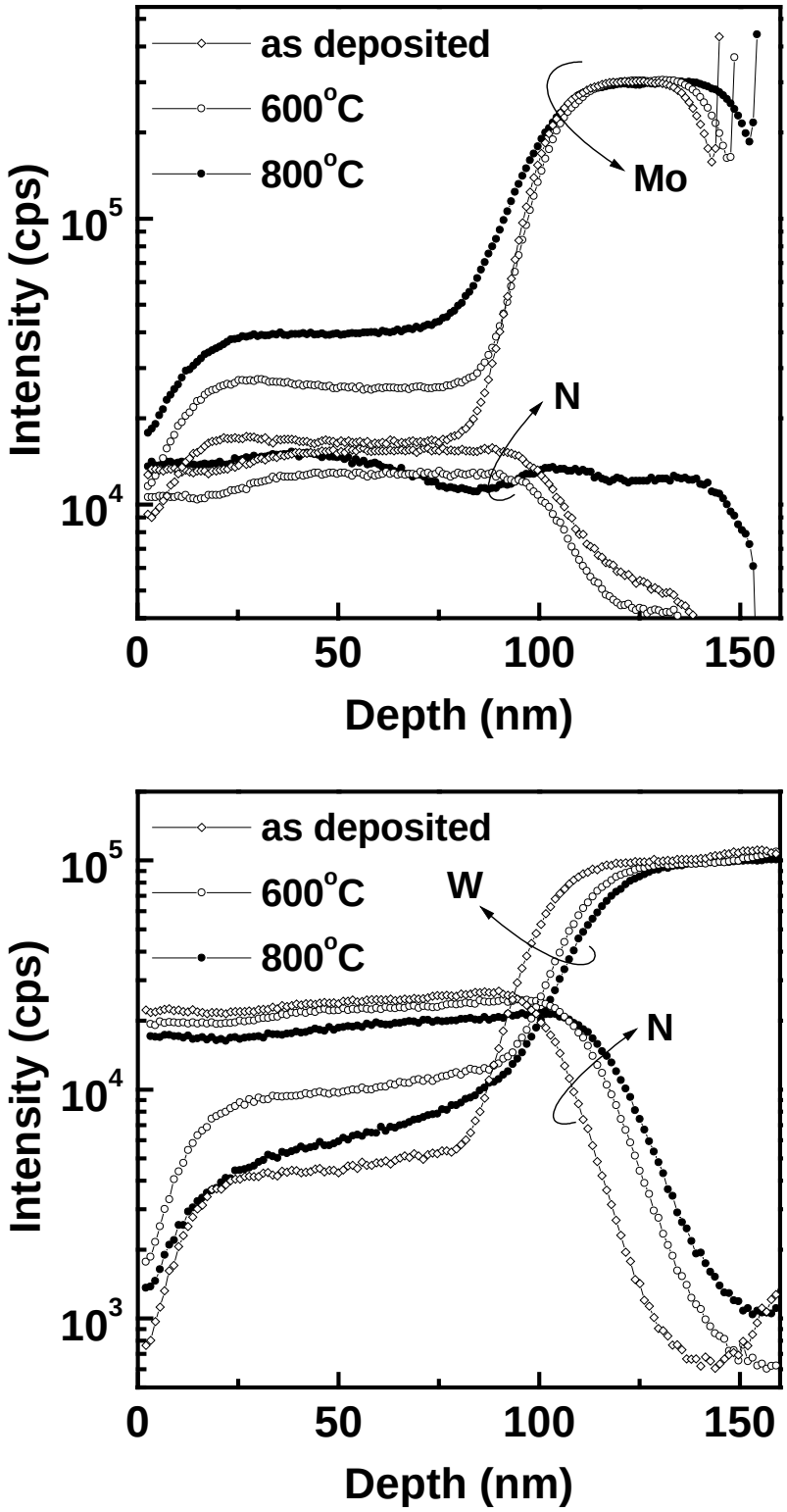


Fig. 3.17 (a-b). SIMS concentration-depth profiles for the $TiN_{1+x}/Mo/SiO_2$ (a) and the $TiN_{1+x}/W/SiO_2$ (b) stacks before and after annealing (600°C and 800°C, 5 min. each)

The low solid solubility of N in Mo (< 1% at 1000°C) supports the observed formation of Mo_2N after N implantation. The implanted N atomic content for a dose of 1×10^{16} cm^{-2} into a 15 nm Mo film is ~10%. It can thus be hypothesized that a combination of enhanced diffusivity and low solubility can lead to the segregation of N and formation of Mo_2N in N implanted Mo films or in Mo films exposed to N from a solid diffusion source like TiN.

3.3 Summary

Several approaches to metal work function engineering have been explored and the relationship between thin film structure and chemistry and work function has been experimentally demonstrated. Ultra-low energy Ar and N ion implantation was used to modify the structure and/or chemistry of Mo gate films, resulting in a lowering of the work function. In the case of N, the effects are primarily chemical (Mo_2N formation in the interfacial regions) and hence continue to evolve with increasing annealing temperature (up to 900°C). In the case of Ar however, the effects are primarily structural (amorphization) and hence easily reversible with annealing (500°C). Solid source diffusion of N into Mo and W films was also analyzed and presented as a potential technique to alter metal thin film chemistry. These techniques provide straightforward approaches for achieving multiple metal gate work functions for CMOS applications and may also be adapted to other metals. The next chapter describes the application of nitrogen implantation into Mo gate films to advanced MOSFET fabrication.

3.4 References

1. C. Kittel, “Introduction to Solid State Physics”, 7th Edition, Wiley, New York, 1996
2. H. B. Michaelson, “The Workfunction of the Elements and its Periodicity”, J. Appl. Physics, v. 48, no.11, November 1977, p. 4729, Nov. 1977
3. R. Smoluchowski, Physical Rev., 60, p. 661, 1941
4. L. Chang, S. Tang, T.-J. King, J. Bokor and C. Hu, “Gate Length Scaling and Threshold Voltage Control of Double-Gate MOSFETs”, International Electron Devices Meeting 2000, Technical Digest, p. 719, Dec. 2000
5. I. De, D. Johri, A. Srivastava, C. M. Osburn, “Impact Of Gate Workfunction on Device Performance at the 50 nm Technology Node”**,** Solid-State Electronics, vol.44, no.6, p. 1077, June 2000
6. P. Ranade, H. Takeuchi, T.-J. King and C. Hu, “Work Function Engineering of Molybdenum Gate Electrodes By Nitrogen Implantation”, Electrochemical & Solid State Letters, vol. 5, no. 11, p. G85, Nov. 2001
7. Z. B. Wei, P. Grange and B. Delmon, “XRD and XPS Studies of Fresh and Sulfided Mo_2N”, Applied Surface Science, vol. 135, p. 107, 1998
8. R. J. P. Lander et al, “A Tunable Metal Gate Work Function Using Solid State Diffusion of Nitrogen”, Proceedings of the ESSDERC, p. 103, Sep. 2002
9. G. Qi, W. M. Small and T. Debroy, “Thermochemistry and diffusion of nitrogen in solid molybdenum”, Met. Trans. B, v.22, p. 219, April 1991
10. J. H. Evans and B.L. Eyre, “The heat of solution and diffusivity of nitrogen in molybdenum”, Acta Metallurgica, vol. 17, no. 8, p. 1109, Aug. 1969

Chapter 4

Molybdenum Gate FET Fabrication

4.1 Introduction

This chapter describes the application of one of the primary metal gate work function engineering approaches described earlier to advanced MOSFET fabrication. As described in Chapter 3, MOS capacitor experiements have shown the effectiveness of nitrogen implantation as a technique to alter the Mo gate work function over a fairly wide range. It was also evident that the range of work functions obtained would make Mo more suitable for advanced MOSFET structures using fully-depleted Si channels. In this chapter, the application of this tunable work function approach to two promising device structures, namely the ultra-thin body MOSFET (UTBFET) and the FinFET are described. In addition, this chapter also discusses the breakdown characteristics of Mo gate capacitors with thermally grown SiO_2 gate dielectrics. The effect of N implantation on the work function of the gate electrode is highlighted using *I-V* measurements and the resulting differences in the breakdown characteristics of the capacitors are compared. The change in the gate work function is evident from a change in the slope of the Fowler-Nordheim (*F-N*) relationship. The lower time to breakdown (t_{BD}) of the devices that received the highest N implant energy and dose is attributed to the reduced barrier height at the gate-dielectric interface which increases the electron tunneling probability across the interface.

4.2 Molybdenum Gate Ultra-Thin Body (UTB) FETs

The ultra-thin body FET structure is considered to be a promising alternative to the conventional bulk-Si FET for gate lengths less than 15 nm [1-5]. This structure was first demonstrated by Choi et al [6] using a highly scaled Si body thickness (< 5 nm) and was shown to provide excellent control of short-channel effects with reasonable current output. A schematic of the proposed UTBFET is shown in Fig. 4.1.

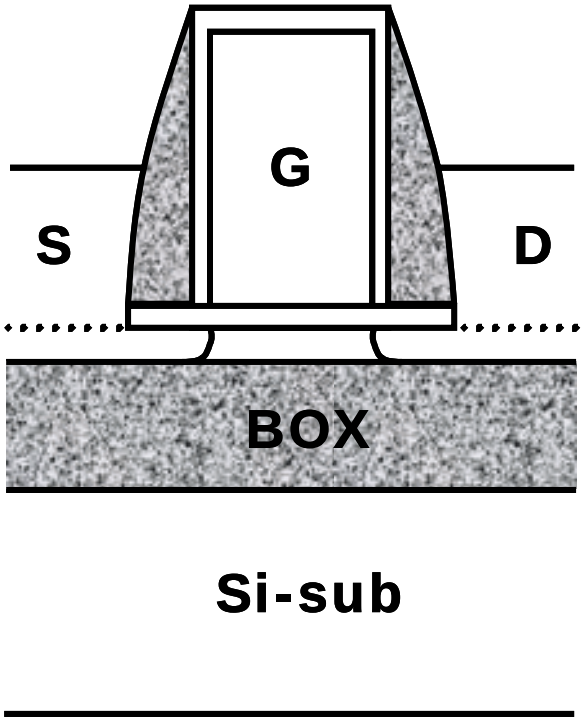


Fig. 4.1 Schematic of an ultra-thin body (UTB) SOI MOSFET. The presence of the buried oxide prevents sub-surface leakage between the source and drain regions.

It can be seen that the structure is very similar to a conventional bulk-Si MOSFET, the primary distinction being the presence of a thick buried oxide (BOX) (typically 400 nm) under a very thin single-crystalline Si layer (<10 nm). Such SOI substrates are generally manufactured using either a high dose O^+ implant and high temperature annealing (the

SIMOX process) or a layer transfer process that involves the implantation of H^+ into shallow depths of "carrier" Si substrates followed by physical bonding of these carrier wafers to thermally oxidized Si substrates and subsequent separation along the H^+ implanted regions to leave behind thin single crystalline Si layers on thermal oxide (the SmartCut® process). Typical as-received Si body thicknesses are generally much greater than 10 nm. Further thinning can be performed using controlled thermal oxidation to consume Si and the resulting oxide can then be removed using chemical etch solutions (HF). Subsequent processing for UTBFET formation is similar to conventional FET formation until the source-drain (S/D) formation step. As shown in Fig. 4.1, the S/D regions in the UTBFET are elevated above the Si substrate plane. This is necessary since the presence of the buried oxide prevents the formation of deep S/D regions. In order to reduce the series resistance to carriers along the channel, the S/D regions have to be raised above the substrate plane as shown. The S/D regions can be formed in a number of ways, the most common being a Si selective epitaxy to deposit thin film Si over the avtive S/D regions and not over the gate or the exposed buried oxide. Another technique that can also provide for abrupt junctions and improved dopant activation efficiencies involves selective Ge deposition. This technique was originally developed for bulk-Si applications as described in detail in the following chapter and has also been successfully used in UTBFET fabrication and FinFET fabrication described in the subsequent section.

Some of the distinct advantages of the UTBFET structure are summarized below:

1. The buried oxide eliminates sub-surface leakage paths between the source and the drain regions, making this structure more immune to short-channel effects (improved subthreshold swing and drain-induced-barrier-lowering (DIBL)).

2. Since the BOX provides immunity from short-channel effects, high levels of channel doping are not required as in bulk FET design. The ability to leave the channel undoped reduces transverse electric fields and improves carrier mobility in the inversion layer. In addition, for highly scaled MOSFETs, statistical fluctuation of dopant atom densities can lead to significant variations in device performance across a wafer. By using an undoped Si body, such random dopant fluctuations can be avoided.

The aggressively scaled body thickness also introduces additional constraints that must be considered in device design:

1. If the body is left undoped, it is clear that the device threshold voltage (V_{TH}) can only be controlled using the work function of the gate. As discussed at length in earlier chapters, the use of $n+$ or $p+$ poly-silicon gates will lead to significantly higher V_{TH} devices since the work functions are much lower (higher) than the desired values for NMOS (PMOS) devices respectively. This in turn highlights the need for developing a tunable work function metal gate technology for FDSOI CMOS.
2. The presence of the BOX necessitates the use of elevated S/D regions which can be fairly challenging.

In spite of the above constraints, detailed simulations and experiments have indicated that the UTBFET structure is a promising alternative to scale Si-CMOS beyond the 15 nm gate length regimes [2]. Figure 4.2 (a.) shows a schematic cross-section of a Mo-gated UTBFET with raised Ge S/D regions [7]. The Si body thickness was ~10 nm. The gate stack consisted of thermal SiO_2 as the dielectric (2 nm), a thin layer of Mo (40 nm) capped with in-situ n+ doped poly-Si. The n+ poly-Si layer was further capped with a layer of SiO_2 (LTO) which served as a hard mask to prevent Ge deposition on the gate oxide during the S/D formation steps. After Mo deposition, some devices were implanted with N to change the gate work

function. The purpose of the poly-Si capping layer was to provide mechanical support to the thin Mo layer and also to prevent dopant ions from entering the metal gate during the subsequent self-aligned S/D formation steps. After spacer formation, Ge was deposited in the S/D regions to form the elevated structures shown. Following Ge deposition (60 nm), a capping layer of SiO_2 was deposited and the wafer was implanted with B ions to convert the undoped Ge to p+ Ge. A 700°C, 60 s anneal was used to activate the dopant atoms. Figure 4.2 (b.) shows the I_D-V_G characteristics of PMOSFETs fabricated using this process. It is evident that unimplanted Mo gate PMOSFETs display the desired low V_{TH} for PMOSFETs (- 0.2 V). The effect of N implantation is a lowering of the Mo work function. In this case, a lower work function gate for a PMOS device implies a higher $|V_{TH}|$. As shown in Figure 4.3, the I_D-V_G curves for the N implanted devices are shifted to the left implying higher V_{TH}s. In addition, the V_{TH} can be adjusted using the implant dose as shown in the figure. As discussed in [7], a V_{TH} shift of 65 mV for every $1\times10^{15}cm^{-2}$ of implanted N was observed. NMOS devices with N implanted Mo gates are expected to display low V_{TH}s since the work gate work function would be appropriate for such devices.

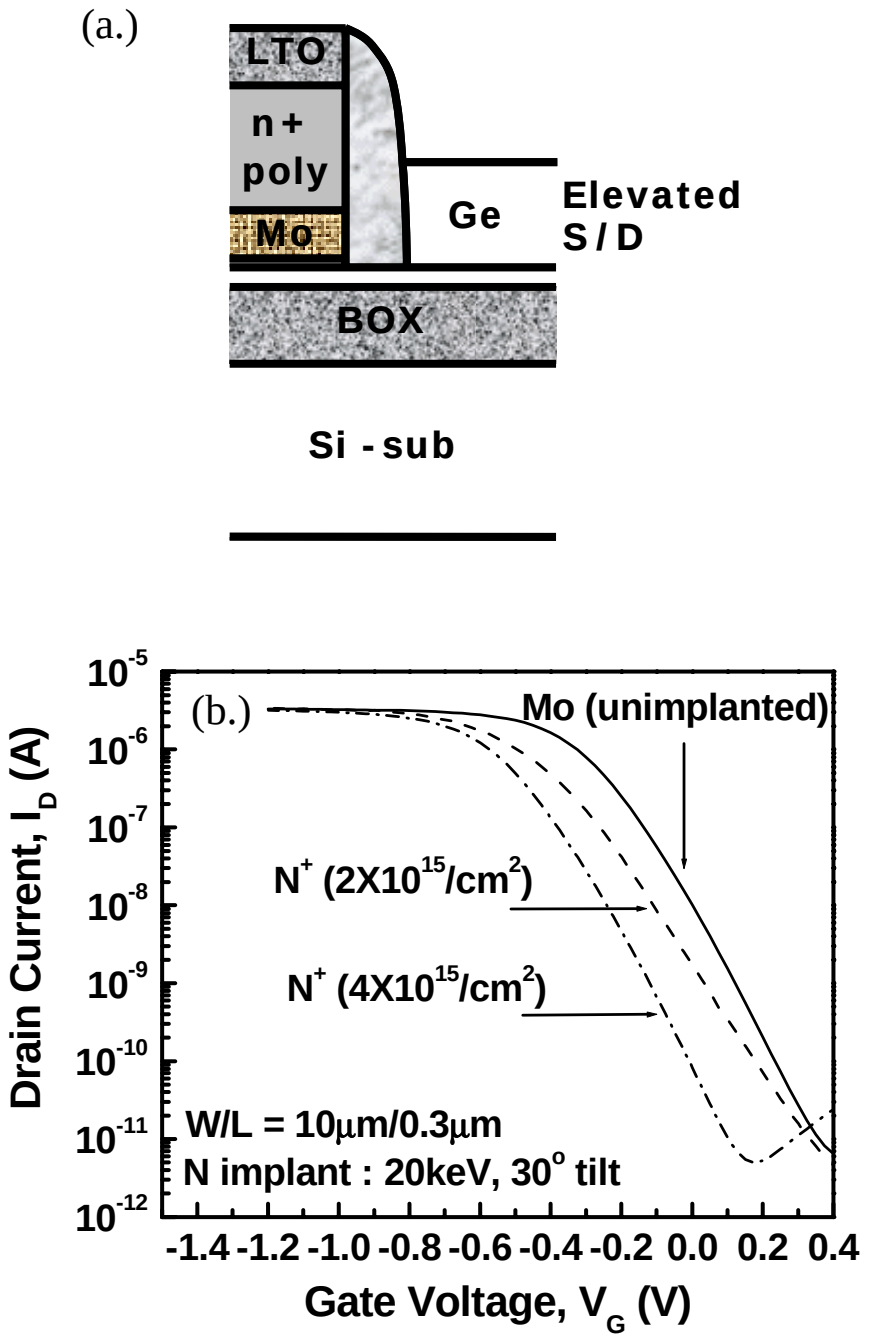


Fig. 4.2 (a.) Schematic of a Mo gate UTB p-FET, (b.) I_D-V_G characteristics of the Mo gate p-FETs showing the dependence of the V_{TH} on the N implant dose.

4.3 Molybdenum Gate FinFETs

Another promising structure that might allow scaling of Si-based CMOS technology into the sub-15 nm technology generations is the FinFET [8-10]. The FinFET is a vertical self-aligned double gate transistor wherein a gate electrode straddles a thin vertical Si body on either side. The structure is so named owing to the shape of the thin Si body (a few nm wide and several hundred nm thick) which resembles the back fin of a fish. The Si fin can be defined on a thin SOI layer. One manifestation of the FinFET is shown in Fig. 4.3.

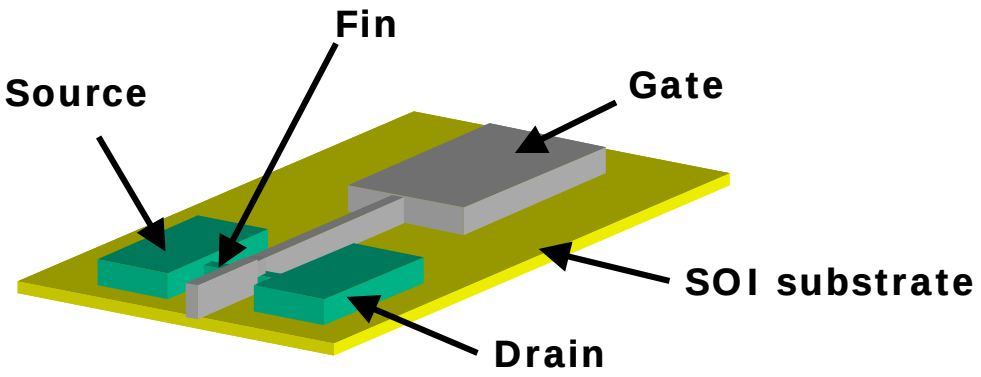


Fig. 4.3 Schematic illustration of a FinFET structure showing a gate electrode straddling a thin, vertical Si fin on an SOI substrate to form two back-to-back active channel regions on either vertical surface of the fin.

As such, the FinFET possesses all the advantages of the ultra-thin body MOSFET described earlier. In addition, it has the additional benefit of two self-aligned gates that serve to increase the coupling with the inversion layer charge. Such a structure is thus more immune to short-channel effects and the FinFET can likely be scaled even beyond the single-gate structures like the UTBFET. In addition, a lower transverse electric field in the FinFET channel can further improve carrier field-effect mobilities and lead to higher performance.

Another advantage of thin-body FETs is a fairly significant reduction in the gate leakage current owing to reduced electric fields. When compared with a bulk-Si device, the reduction in gate leakage current for UTBFETs and FinFETs is expected to be around 3X and 4X respectively [11]. Over the last few years, aggressively scaled FinFET devices (L_G < 15 nm) have been demonstrated by several semiconductor chip manufacturers, indicating that industry adoption in the future is very likely [12-14]. Since the fin is undoped in an ideal FinFET (similar to the Si-body in a UTBFET), gate work function engineering forms a very important aspect in device design. The tunable work function Mo gate technology described earlier is an ideal approach for FinFET fabrication since the work function range observed (4.4 – 4.9 V) is precisely the one needed for CMOS FinFET devices.

In the first demonstration of a tunable work function metal-gate FinFET technology, molybdenum gate p-channel quasi-planar FinFETs were fabricated using aggressively scaled fin widths (10 nm) [15,16]. Figure 4.4 is a schematic view of the FinFET gate stack indicating the large angle tilted implantation that was used to introduce N into the vertical side-walls. As shown in the I_D-V_G curves of Fig. 4.5, the V_{TH} of an unimplanted Mo gate p-channel FinFET is – 0.2 V as desired, while N implantation controllably increases the $|V_{TH}|$ since the gate work function is now much lower than desired. It should be noted that the low gate work functions would however be ideal for n-channel devices and NMOS FinFETs fabricated using this approach should display low V_{TH}s. In addition, the use of N implantation can also enable the fabrication of a multiple-V_{TH} technology on a single chip. Such a technology is desirable for applications where high performance (high on-current, low V_{TH}) and low power (low off-state leakage, high V_{TH}) transistors are needed on a single chip.

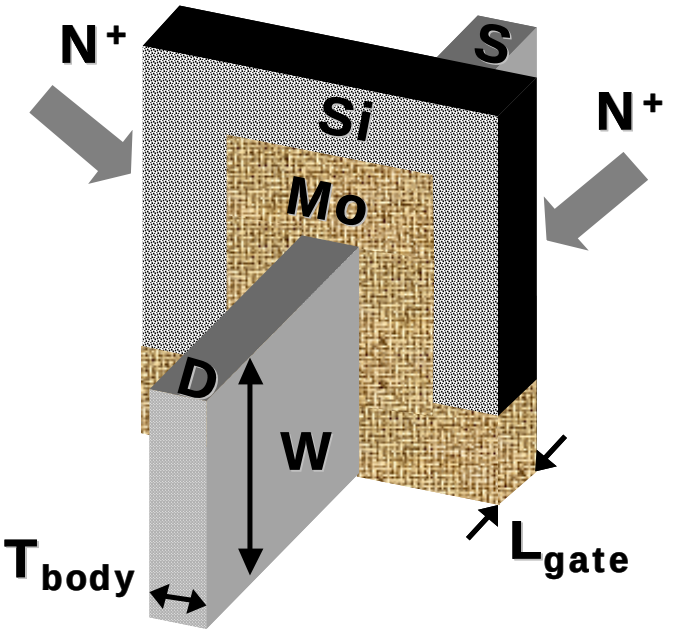


Fig. 4.4 Schematic view of a Mo gate FinFET where N implantation into the gate was used to tailor the device threshold voltage

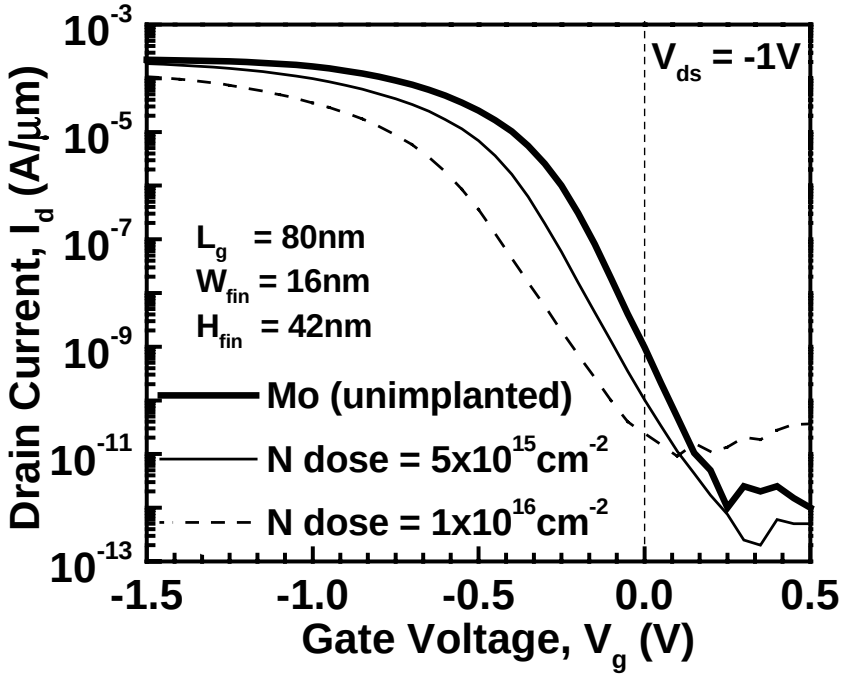


Fig. 4.5 I_D-V_G curves of a Mo gate p-FinFET. The gate dielectric was 2 nm SiO_2. The V_{TH} of the unimplanted Mo gate device is – 0.2 V, while increasing the N implant dose progressively increases the V_{TH}

4.4 Dielectric Reliability after N Implantation

A single metal tunable work function gate technology such as the one described in this work can be very attractive for future CMOS devices. However, the effect of heavy dose N implantation into the gate electrode on the reliability of the underlying gate dielectric warrants further investigation. While the effect of gate work function on the breakdown characteristics of poly-Si/SiO_2/Si capacitors has been investigated earlier [17,18], a similar analysis for metal gate capacitors is lacking. In this section, the effect of N implantation into Mo gate capacitors on the breakdown characteristics of these devices is studied. From Fowler-Nordheim (*F-N*) tunneling measurements, it is shown that the implantation of N into Mo changes the barrier height at the SiO_2 interface, thus confirming the lowering of the work function observed earlier from *C-V* measurements [5]. Leakage measurements were performed at elevated temperature to examine the influence of temperature on the tunneling mechanism and constant voltage stress measurements were performed to ascertain the effect of N implantation on the breakdown characteristics of the devices.

4.4.1 Experiment

Molybdenum gated capacitors were fabricated on thermally oxidized p-type Si wafers. Dry oxidation of Si was used to grow ~ 11 nm SiO_2, after which thin Mo films (150 nm) were deposited onto the thermally oxidized substrates. The Mo deposition was performed in a DC magnetron sputtering system with an Ar plasma (base pressure = 1×10^{-7} T, process pressure = 6 mT). Substrate temperature during deposition was maintained at 200°C and plasma power used was 100 W. The thickness of the gate oxide was quite large compared to the ultra-thin gate dielectrics used in state-of-the-art technology. However, since the primary

objective of this work was to evaluate the impact of N implantation on the gate work function and breakdown characteristics of Mo gated capacitors, conclusions drawn using a relatively thick gate dielectric should nevertheless be useful. It should be noted that from a device reliability standpoint, the use of chemical vapor deposition (CVD) for metal gate electrodes is most desirable, since sputter deposition can lead to degradation of ultra-thin dielectric films (via metal penetration into the dielectric during deposition).

After metal deposition, dropout lithography was used to open up portions of the wafer which were then implanted with N ions (N^+, 60 keV or 80 keV) in doses ranging from $2\times10^{15}cm^{-2}$ to $1\times10^{16}cm^{-2}$. Following N^+ implantation, the metal films were patterned using a Cl_2/O_2 based reactive-ion-etch (RIE) to define gate electrodes. After a rapid thermal annealing treatment (900°C, 10 s), the devices were characterized using *I-V* measurement in the Fowler-Nordheim (*F-N*) tunneling regime. The data was plotted according to the *F-N* equation using the oxide thickness and flat-band voltages obtained from *C-V* measurement. For the sample that received the highest implant energy and dose, the *I-V* measurements were repeated for different temperatures ranging from 25-150°C to see if the tunneling mechanism was dependent on temperature. Constant-voltage stress measurement with a negative bias on the gate electrode was used to estimate charge-to-breakdown (Q_{BD}) and time-to-breakdown (t_{BD}). The results are discussed in detail in the following section.

4.4.2 Results and Discussion

Figures 4.6 (a-d) show the *J-V* characteristics of the capacitors obtained using a negative bias on the Mo electrode. The gate area was 1×10^{-4} cm^2 and t_{OX} (obtained from *C-V* measurement) was 11 nm.

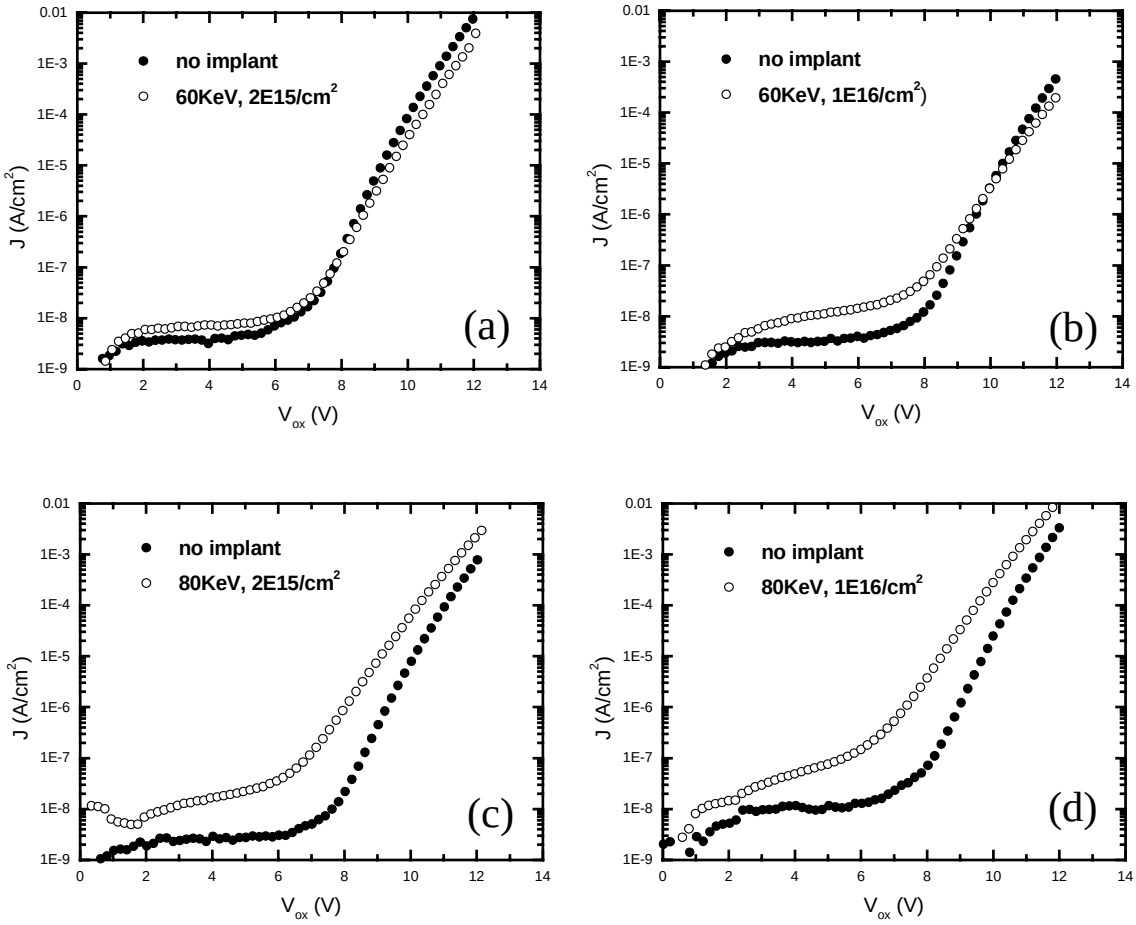


Figure 4.6 (a-d) Tunneling current vs oxide voltage for Mo gate capacitors with and without N implantation. The gate voltage was maintained negative w.r.t. the substrate. Results for four different implant conditions are shown. Capacitor area = 1×10^{-4}cm^2, t_{ox}= 11 nm

The flat band voltages for the different devices (obtained from *C-V* measurements) were used to plot the data of Fig. 4.6 according to the *F-N* equation shown below:

$$J = K.E_{OX}{}^{2}.e^{(-B/E_{OX})}$$

where J is the current density ($\mathrm{A/cm^2}$), E_{OX} the electric field ($\mathrm{V/cm}$) and the parameters K and B are dependent on the barrier height at the metal-dielectric interface (ϕ_B). The exact expressions for B and K are as follows:

$$B = \frac{8\pi\sqrt{2m_{OX}}.\phi_B^{(3/2)}}{3hq} \text{ and}$$

$$K = \frac{q^3 m_n}{8\pi h m_{OX}\phi_B},$$

where m_{OX} denotes the electron effective mass in the dielectric, h is the Planck constant, m_n is the electron mean free path and q is the electronic charge.

From the *F-N* equation it is evident that the slope of the *F-N* plot (J/E^2 vs $1/E$) is proportional to $\phi_B^{3/2}$ at the metal-dielectric interface. A comparison of the slopes for the N implanted devices and the unimplanted devices (Fig. 4.7 (a-d)) shows that there is a lowering of the barrier height after N implantation in agreement with work function measurements reported earlier [5].

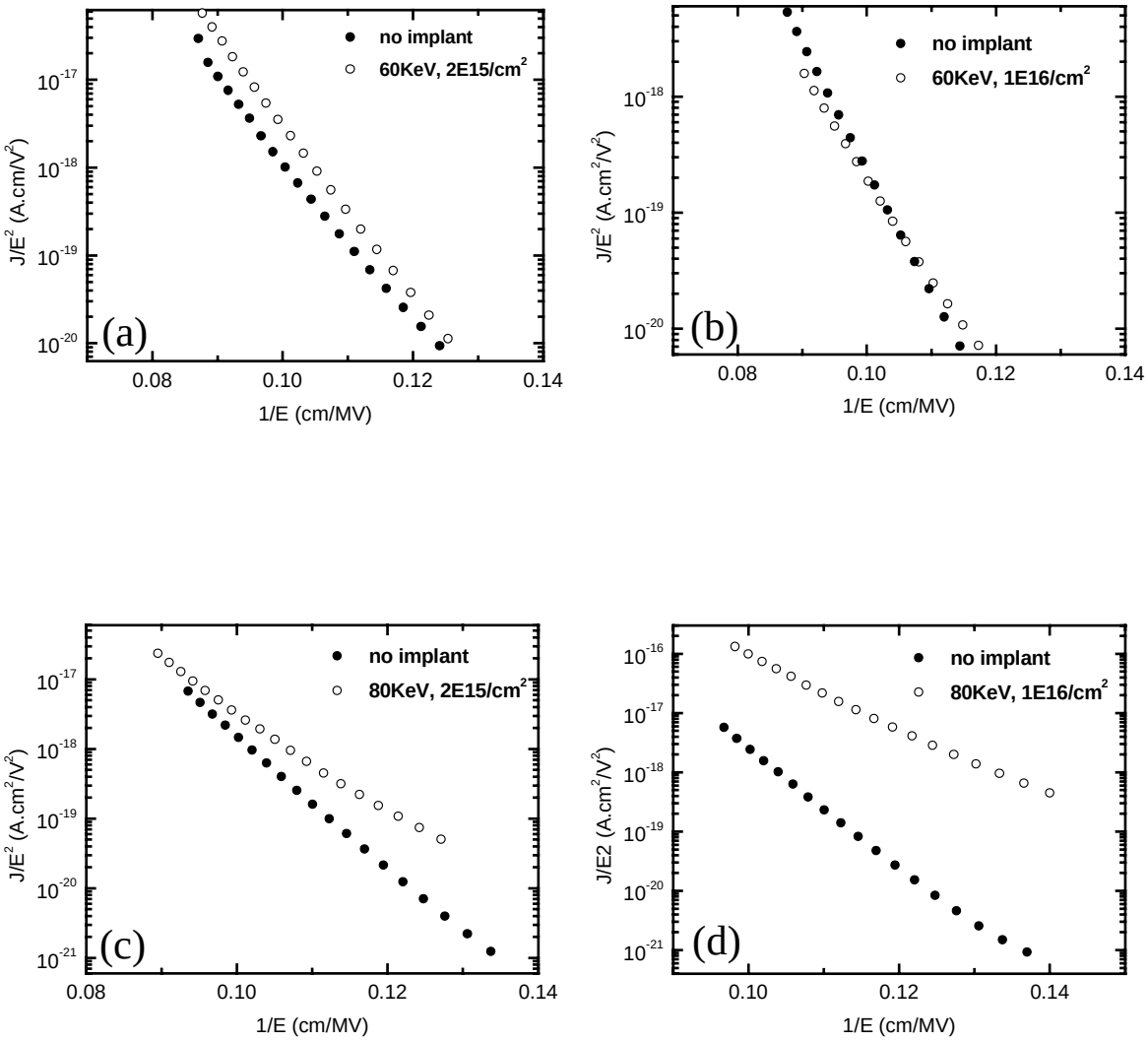


Fig. 4.7 (a-d) Fowler Nordheim tunneling relationships plotted using the data of Fig. 4.6. Comparing the slopes of the curves for the implanted devices with the control devices, it can be concluded that the barrier height at the metal-dielectric interface (proportional to the slope) is lowered by N implantation.

Figure 4.8 shows the temperature dependence of the *J-V* relationship for a control device and a device that received the highest N implant dose and energy in this study (80 keV, $1\times10^{16}/\mathrm{cm}^2$). These measurements were made to evaluate the impact of temperature on the

tunneling mechanism. It can be observed that the tunneling mechanism is independent of temperature and predominantly *FN* tunneling (atleast up to 150°C). However, the larger spread of the current for the implanted device indicates a larger susceptibility to breakdown and a possible degradation of gate oxide quality after N implanation. It should also be noted that the *J-V* curves for the unimplanted (control) sample are also not very tightly distributed (a spread in *J* of 0.002 mA/cm^2 at a V_G = -9 V).

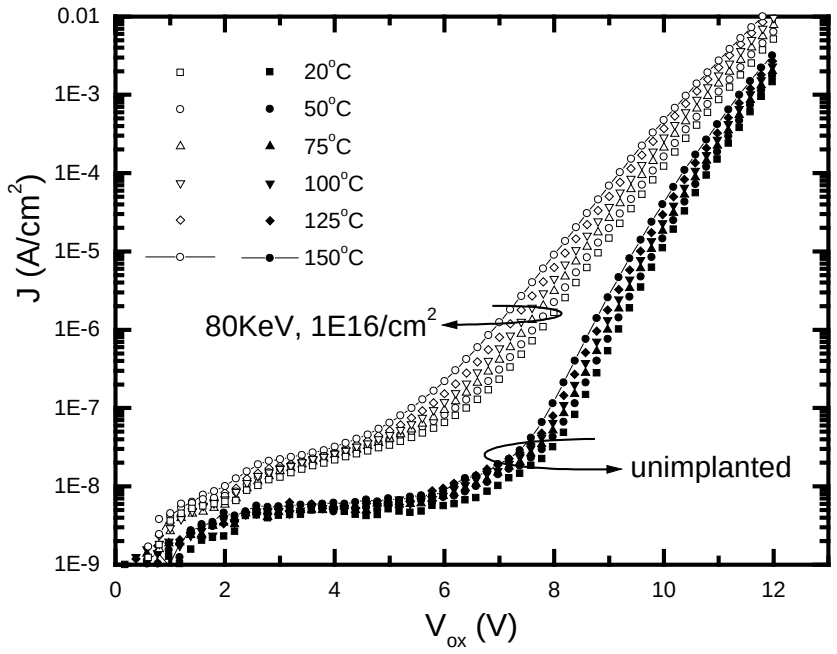


Fig. 4.8 The effect of temperature on the *J-V* relationship for the device of Figs. 4.6 and 4.7 (d). While the tunneling mechanism remains unchanged over the range of temperatures used, the increased spread in the current density indicates a deterioration in the oxide quality on account of the implantation

This is most likely due to the sputter induced damage of the gate dielectric during metal deposition. The elevated substrate temperature during metal deposition (200°C) can promote the degradation of the dielectric by facilitating the penetration of deposited Mo atoms to a certain depth within the dielectric. It should also be noted that the N implanted

devices show significantly higher leakage current at low (< 6 V) voltages (before the onset of *FN* tunneling). This enhanced leakage is attributed to the N implant. Since the deposited Mo films have a columnar morphology, it is likely that an appreciable fraction of the implanted ions channel through along the grain boundaries and into the gate dielectric. Similar observations of enhanced leakage after N implantation into Mo gate electrodes were made by Amada et al [20], who observed that a large angle (30°) tilted implant significantly lowers the channeling and the leakage current.

Figures 4.9 (a-d) plot the time-to-breakdown (t_{BD}) of the capacitors versus the voltage across the gate dielectric (V_{OX}). V_{OX} is the actual voltage dropped across the gate dielectric (applied gate voltage minus the work function difference between the gate and the silicon substrate). The gate voltage during these measurements was held constant (negative *w.r.t.* the substrate) and the tunneling current (essentially constant during measurement) was monitored until breakdown occurred (detected by an abrupt increase in the current).

From earlier measurements [5], it was observed that the gate work functions after implantation were lower than the unimplanted control value (4.95 V) by 0.025 V and 0.4 V for the samples that received the 60 keV and 80 keV N implants respectively (dose = $1\times10^{16}/\mathrm{cm}^2$). This would also imply a reduction in the barrier height between the Mo and the SiO_2 by a comparable amount. With a negative bias on the gate electrode, electron injection from the gate will be dependent on the barrier height at the metal-dielectric interface.

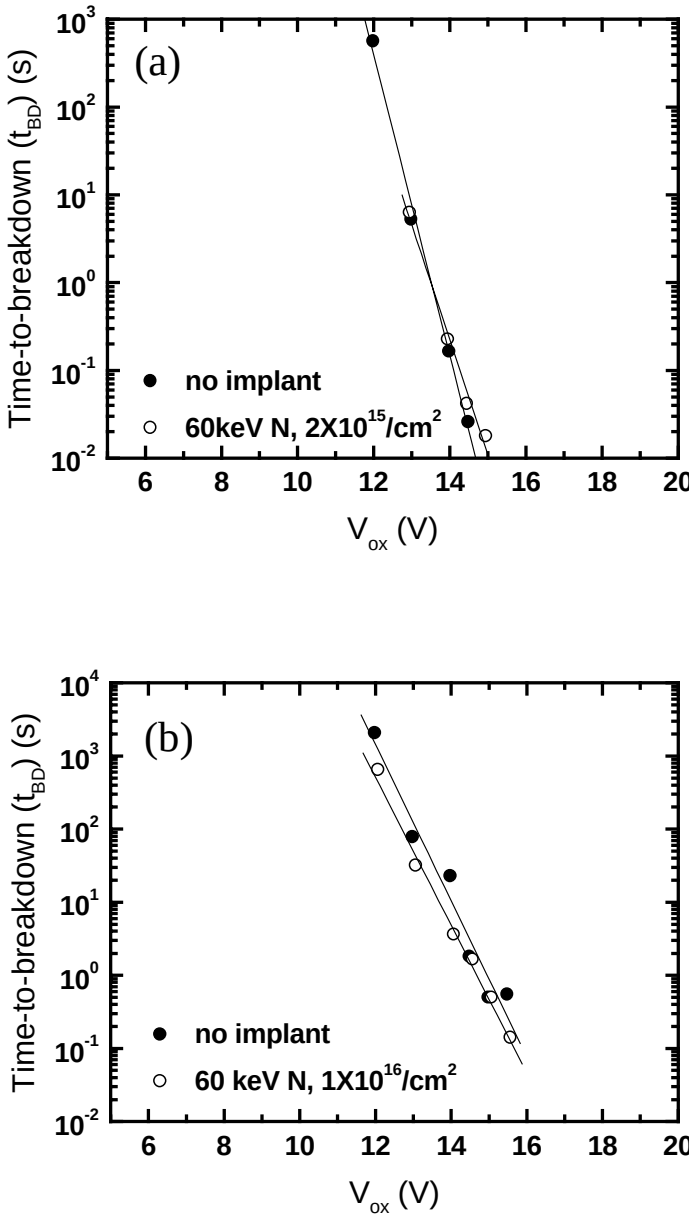
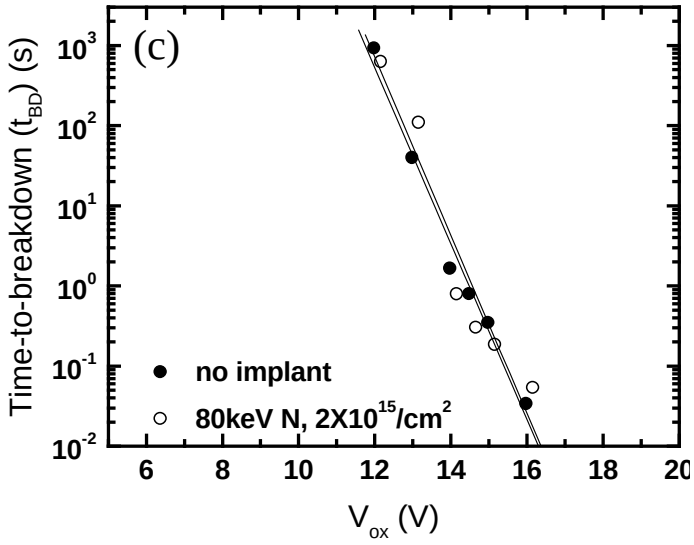


Fig. 4.9 (a-b) Time-to-breakdown (t_{BD}) as function of voltage drop in the gate dielectric (V_{OX}) for the different devices, implant energy = 60 keV



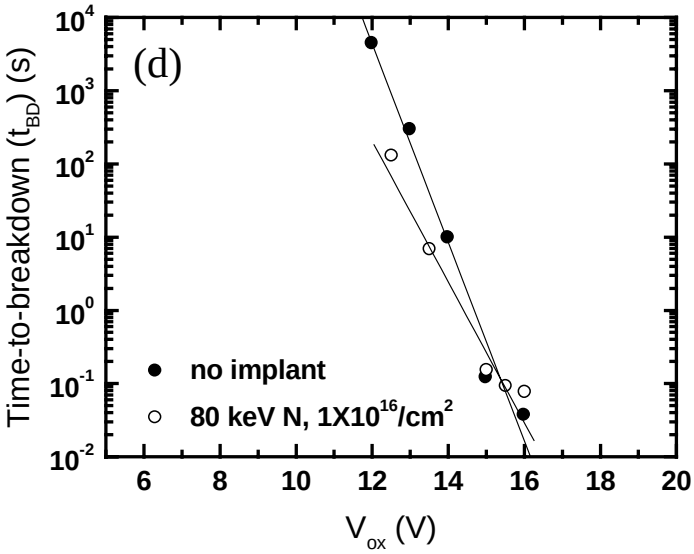


Fig. 4.9 (c-d). Time-to-breakdown (t_{BD}) as function of voltage drop in the gate dielectric (V_{OX}) for the different devices, implant energy = 80 keV

While the change in the work function is relatively small for the 60 keV implanted device, it is appreciable for the 80 keV implanted device. A reduction in the barrier height (and hence the work function) is likely to increase the electron tunneling probability across the metal-dielectric interface. As seen in Figs. 4.9 (a-c), a low implant energy/dose does not lead to a significant change in the t_{BD} vs. V_{OX} relationship. A combination of high implant energy and high dose (80 keV) (Fig. 4.9 (d)) however, leads to a significant reduction in the t_{BD} for a fixed voltage across the gate oxide. This difference is more pronounced at lower applied voltage, presumably because of a larger difference in the tunneling probability when the electric field across the oxide is small. However, at higher electric fields, the exponential term in the *FN* equation dominates, so that both the devices show comparable breakdown times. Similar observations were made with n+ and p+ poly-silicon devices by Yang et al [18].

Figures 4.10 (a-b) show the time-to-breakdown (t_{BD}) variation with the tunneling current density (J) for the implanted and control devices. It can be seen that for a fixed tunneling current density, t_{BD} with and without implantation is comparable at large tunneling current densities (higher applied voltage), while the N implanted devices show smaller t_{BD} at lower current densities. Holland et al [17] have observed that for n+ and p+ poly-silicon gate capacitors, the t_{BD} for a fixed tunneling current density is independent of gate work function. However, in their measurement, the gate oxide thickness was 4 nm. For thicker gate oxides, they also observed similar divergence in t_{BD} at low oxide voltages. The increased t_{BD} for the N implanted device is thus most likely due to the lower barrier height and the resulting increase in tunneling probability through the gate dielectric. Nevertheless, it is important to note that for an optimal combination of gate work function and dielectric reliability, the N implant parameters should be chosen to minimize N penetration into the gate dielectric.

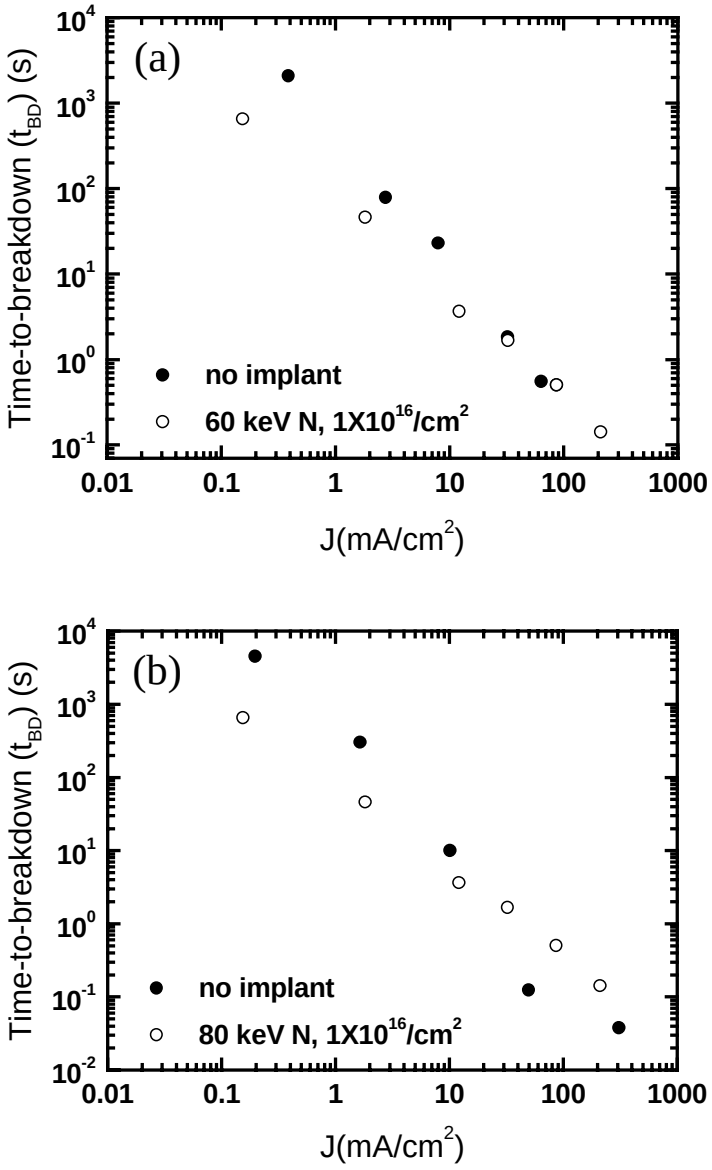


Figure 4.10 (a-b) Time-to-breakdown (t_{BD}) as function of tunneling current density (J) in for the devices that received the highest implant dose (a) 60 keV, (b) 80 keV.

4.5 Summary

This chapter described the application of the tunable work function Mo gate technology described earlier to transistor fabrication. The application of this technology to p-channel UTBFETs and FinFETs has been described. In both cases, the results of N implantation are consistent with the work function lowering described in Chapter 3. The use of unimplanted Mo gates was shown to produce low V_{TH} p-channel devices using either transistor structure. In addition to preliminary transistor performance, the effect of N implantation on the work function of Mo gate electrodes and the reliability of gate oxide films has also been evaluated. The effect of the nitrogen implantation on the barrier height at the metal-dielectric interface is observed as a change in the slope of the Fowler-Nordheim plot. The tunneling mechanism is seen to remain FN dominated even at temperatures as high as 150°C. The lower time to breakdown (t_{BD}) of the devices that received the highest N implant energy and dose is attributed to the reduced barrier height at the gate-dielectric interface which in turn increases the electron tunneling probability across the interface.

4.6 References

1. C. Hu, “SOI and nanoscale MOSFETs”, IEEE 59th Device Research Conference, p.3, June 2001
2. Y. -K. Choi, K. Asano, N. Lindert, V. Subramanian, T.-J. King, J. Bokor and C. Hu, “Ultrathin-body SOI MOSFET for deep-sub-tenth micron era”, IEEE Electron Device Letters, vol.21, no.5, p. 254, May 2000
3. Y. -K. Choi, Y. -C. Jeon, P. Ranade, H. Takeuchi, T. -J. King, J. Bokor, C. Hu, “Ultra-thin body SOI Device with 30nm Gate Length and Selectively Raised Ge S/D”, Conference Proceedings, IEEE 58th Device Research Conference, p.23, June 2000
4. J. Kedzierski, P. Xuan, V. Subramanian, J. Bokor, T.-J. King, C. Hu and E. Anderson, “A 20 nm gate-length ultra-thin body p-MOSFET with silicide source/drain”, Superlattices and Microstructures, vol.28, no.5-6, p.445, 2000
5. R. Chau et al, “A 50 nm depleted-substrate CMOS transistor (DST)”, International Electron Devices Meeting 2001, Technical Digest, p. 621, Dec. 2001
6. Y.-K. Choi, D. Ha, T.-J. King and C. Hu, “Nanoscale ultrathin body PMOSFETs with raised selective germanium source/drain”, IEEE Electron Device Letters, vol.22, no.9, p.447, Sep. 2001
7. D. Ha, P. Ranade, Y.-K. Choi, J.-S. Lee, T.-J. King and C. Hu, “Molybdenum Gate Work Function Engineering for Ultra-Thin Body Silicon-on-Insulator (UTB-SOI) MOSFETs”, Japanese Journal of Applied Physics, April 2003 (to be published)
8. D. Hisamoto et al, “A folded-channel MOSFET for deep-sub-tenth micron era”, International Electron Devices Meeting 1998, Technical Digest, p. 1032, Dec. 1998
9. X. Huang et al, “Sub-50 nm P-channel FinFET”, IEEE Transactions on Electron Devices, vol.48, no.5, p.880, May 2001
10. Y.-K. Choi et al, “Sub-20 nm CMOS FinFET technologies”, International Electron Devices Meeting 2001, Technical Digest, p. 1032, Dec. 2001
11. L. Chang, K. J. Yang, Y.-C. Yeo, Y.-K. Choi, T.-J. King and C. Hu, “Reduction of direct-tunneling gate leakage current in double-gate and ultra-thin body MOSFETs”, International Electron Devices Meeting 2001, Technical Digest, p. 1032, Dec. 2001

12. R. Chau et al, “Advanced Depleted-Substrate Transistors: Single Gate, Double-Gate and Tri-Gate”, 2002 International Conference on Solid State Devices and Materials, p. 68, Sep. 2002
13. J. Kedzierski et al, “Metal-Gate FinFET and Fully-Depleted SOI Devices Using Total Gate Silicidation”, International Electron Devices Meeting 2002, Technical Digest, p. 247, Dec. 2002
14. B. Yu et al, “FinFET Scaling to 10nm Gate Length”, International Electron Devices Meeting 2002, Technical Digest, p. 251, Dec. 2002
15. Y.-K. Choi, L. Chang, P. Ranade, J.-S. Lee, D. Ha, S. Balasubramanian, A. Agarwal, M. Ameen, T.-J. King and J. Bokor, “FinFET Process Refinements for Improved Mobility and Gate Work Function Engineering”, International Electron Devices Meeting 2002, Technical Digest, p. 259, Dec. 2002
16. P. Ranade, Y.-K. Choi, D. Ha, A. Agarwal, M. Ameen and T.-J. King, “Tunable Work Function Molybdenum Gate Technology for FDSOI-CMOS ”, International Electron Devices Meeting 2002, Technical Digest, p. 363, Dec. 2002
17. S. Holland, I. Chen and C. Hu, “Ultra-Thin Silicon Dioxide Breakdown Characteristics of MOS Devices with n+ and p+ Polysilicon Gates”, IEEE Electron Device Letters, vol. 8, no. 12, p. 572, Dec. 1987
18. T. C. Yang, P. Sachdev and K. C. Saraswat, “Dependence of Fermi Level Positions at Gate and Substrate on the Reliability of Ultrathin MOS Gate Oxides”, IEEE Trans. Electron Dev. vol. 46, no. 7, p. 1457, July 1999
19. P. Ranade, H. Takeuchi, T.-J. King and C. Hu, “Work Function Engineering of Molybdenum Gate Electrodes By Nitrogen Implantation”, Electrochemical & Sol-State Letters, 5, 11, p. G85, Nov. 2001
20. T. Amada, N. Maeda and K. Shibahara, “Degradation in a Molybdenum Gate MOS Structure Caused By N^+ Ion Implantation For The Work Function Control”, Materials Research Society Symposium Proceedings, Spring 2002, B7.5, (in press)

Chapter 5

SiGe Ultra-Shallow Junction Technology

5.1 Introduction

The fabrication of shallow and abrupt source-drain (S/D) extension junctions is one of the most significant challenges for scaling of CMOS transistor technology beyond the 90 nm technology node. According to the International Technology Roadmap for Semiconductors (*ITRS*) [1], bulk-Si CMOS devices with physical gate lengths $L_g \leq 50$ nm will require ultra-shallow S/D extension junctions (junction depth $x_j \leq 30$ nm) for acceptable short channel performance. In addition to being very shallow, extension junctions will also need to be heavily doped in order to meet roadmap requirements for the S/D series resistance ($R_s \leq 200$ Ω/□) to improve device drive current and extremely abrupt (slope ≤ 3 nm/decade) for improved short channel performance. While such aggressive specifications appear to be essential for the continued scaling of CMOS technology, it must be noted that the solid solubility and diffusion of dopants in Si impose fundamental limitations on the scalability of conventional processing schemes and materials used in the fabrication of extension S/D junctions.

The task of achieving ultra-shallow S/D junctions is more challenging for PMOS devices owing to the higher diffusivity and lower solid solubility of B in Si (as compared to As and P) [9, 13]. The transient enhanced diffusion (TED) [14] of implanted dopant atoms,

B in particular [15] is also perceived to be a major obstacle in the fabrication of ultra-shallow p+/n junctions.

Ion implantation has been used in the fabrication of p+ and n+ S/D junctions for several decades. Over the last few years, ion implant energies and doses have been progressively scaled down to reduce junction depths and minimize series resistance. There have been several reports on the use of ultra-low energy (≤ 1 keV) ion implantation to fabricate very shallow junctions (x_j ≤ 30 nm) [6,6]. While the physical introduction of a high concentration of dopant atoms is easily achievable with ion implantation, it is difficult to achieve a high concentration of *electrically active* dopants. At very low implant energies, dopant activation can itself become challenging [7,8].

Recently, there have been several reports on the applications of spike and laser annealing techniques [10] to achieve abrupt junctions with very high (non-equilibrium) active dopant concentrations. These techniques rely on extremely high (near melting) temperatures to produce abrupt junctions with very high active dopant concentrations in Si. However, laser annealing introduces additional concerns of thermal instability in other areas of the device viz. the gate stack (since very high peak temperatures might be employed with this process). Transient enhanced diffusion is also expected to affect diffusion profiles during laser annealing [11]. The possible replacement of SiO_2 with a high permittivity dielectric as the gate insulator might impose additional constraints on maximum annealing temperatures since many of the promising high-k dielectrics have low crystallization temperatures and will in some cases separate into constituent phases (in conjunction with the Si substrate and gate electrode) that are thermodynamically preferred at high temperatures. While very high (higher than the solubility limit) dopant concentrations can be achieved using these sophisticated annealing techniques, the excess, super-saturated concentration of dissolved

dopants in the Si lattice easily precipitate out in the form of electrically inactive dopant clusters during subsequent low temperature annealing steps [12].

The maximum achievable electrically active dopant concentrations are fundamentally limited by the solid solubilities of dopants in Si at the annealing temperature. Boron in particular has a fairly low peak solid solubility in Si (3×10^{20} atoms/cm^3 at 900°C). The fairly high diffusion coefficient of B in Si at 900°C and the transient enhanced diffusion of B during ramp-up impose additional constraints on the fabrication of p+/n junctions in Si. Such fundamental and practical constraints might thus limit the scaling of junction depths in Si using conventional processing techniques and device structures. The goal of this work is to explore alternative device structures and materials that could potentially enable the fabrication of ultra-shallow ($x_j \leq 30$ nm) p+/n S/D extension junctions with relatively low annealing temperatures and thermal budgets.

Several authors have reported on novel device structures and processing schemes involving the use of $Si_{1-x}Ge_x$, aimed at fabricating ultra-shallow junctions for future CMOS devices. In an earlier publication, it has been shown that $Si_{1-x}Ge_x$ alloy can provide significantly higher active B concentrations compared to pure Si [13]. It was recently shown that ultra-shallow $Si_{1-x}Ge_x$/Si heterojunctions formed by Ge^+ ion implantation into the S/D regions can be used to achieve low-resistivity, ultra-shallow S/D extensions for improved bulk-Si PMOSFET drive current and short-channel performance [14]. The unique diffusion properties of dopants at Ge/Si interfaces has also been studied recently [15]. An additional motivation for the use of Ge in the S/D regions of PMOSFETs is that a very low thermal budget is sufficient to activate implanted B (or BF_2) in Ge films [16]. Elevated S/D $Si_{1-x}Ge_x$ MOSFETs have been fabricated using a controlled Si substrate etch and rapid thermal CVD to selectively deposit $Si_{1-x}Ge_x$ films in the S/D regions [17]. PMOSFETs with elevated Si_{1-}

$_xGe_x$ S/D regions (formed by ultra-high vacuum CVD) have also been reported by Huang et al [18] to display improved low-temperature characteristics. This chapter discusses two simple approaches to incorporate Ge in the S/D regions of conventional bulk-Si MOSFETs and the applicability of these approaches to sub-100 nm PMOSFET fabrication.

The remainder of this chapter is organized as follows: Section 5.2 presents $Si_{1-x}Ge_x$/Si heterojunction diode characteristics and also highlights the impact of Ge on the formation of p+/n junctions in Si substrates. Section 5.3 presents electrical characteristics of PMOSFETs fabricated using selective Ge deposition in the S/D regions. Section 5.4 provides a discussion of additional motivations for incorporating SiGe in the S/D regions of a MOSFET including low temperature process compatibility and improved contact resistance. Section 5.5 provides a summary of the various results.

5.2 $Si_{1-x}Ge_x$/Si p+/n Heterojunctions : Electrical and Materials Characterization

This section presents the electrical characteristics and materials analysis results for p+$Si_{1-x}Ge_x$/n-Si junctions fabricated using two techniques.

5.2.1 Ge Ion Implantation

Boron doped p+/n $Si_{1-x}Ge_x$/Si junction diodes were prepared using the following fabrication sequence: LOCOS isolation was used to define active areas on single crystalline (100) Si substrates. This was followed by thermal oxidation to grow SiO_2 (~ 6 nm) on the active regions. Ge ions were then implanted into these wafers (energy: 6 keV, dose: $1\times10^{16}cm^{-2}$). This implant was expected to produce a thin (~ 25 nm) $Si_{1-x}Ge_x$ layer at the Si surface with x

≤ 0.2, upon annealing (based on SRIM simulations). Ge implantation was followed by B implantation (energy : 5 keV, dose: $6\times10^{15}cm^{-2}$). Secondary Ion Mass Spectroscopy (SIMS) analysis was used to characterize the co-diffusion of Ge and B in the Si. Figure 5.1 shows the SIMS concentration-depth profiles after high temperature rapid thermal annealing (900°C, 30s + 1050°C, 5s). It can be seen that the Ge implantation and subsequent annealing has produced a thin $Si_{1-x}Ge_x$ layer at the surface with a peak Ge concentration of ~14 at.%. A steep profile and a high peak concentration of B (5×10^{20} cm^{-3}) are obtained at the surface.

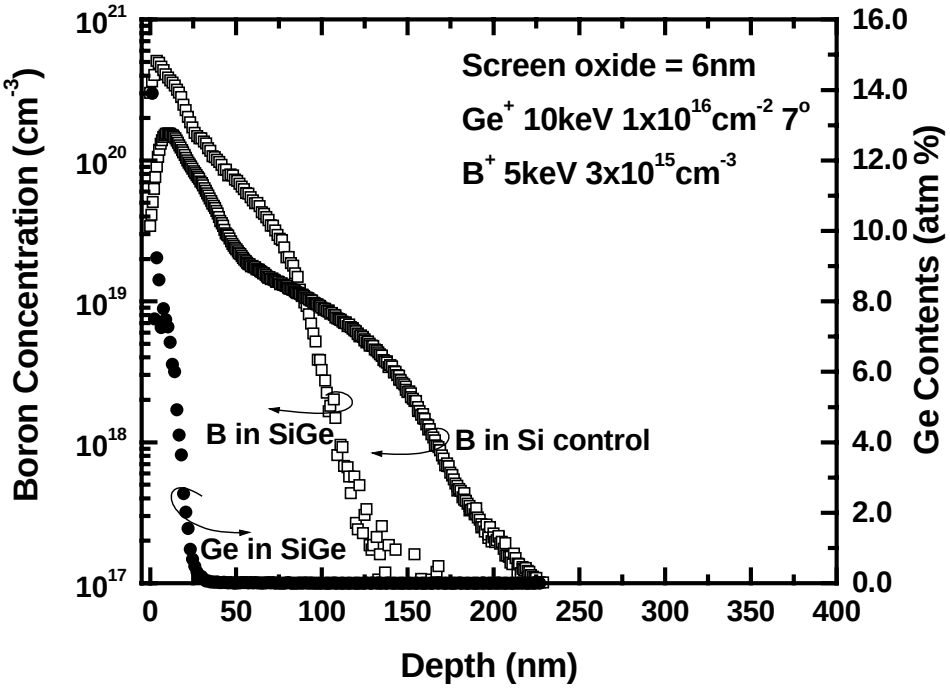


Figure 5.1 SIMS concentration-depth profiles after Ge and B implantation into Si and annealing at 900°C. A thin $Si_{1-x}Ge_x$ layer at the surface with a peak Ge concentration of ~14 at.% is obtained. A steep profile and a high peak concentration of B ($5\times10^{20}cm^{-3}$) are obtained at the surface. The B profile in a control wafer is also shown for comparison. The B profile is markedly shallower with the Ge present

Figure 5.1 also shows the B profile in a control wafer (no Ge implant but an identical B implant and thermal processing) superimposed on the B profile from Ge implanted wafer. It can be concluded that the presence of Ge minimizes the spread of B into the Si and leads to its confinement in the shallow $Si_{1-x}Ge_x$ regions. Figure 5.2 compares the p+/n diode leakage currents for the heterojunction diodes and the control devices. The heterojunction diodes exhibit larger leakage when compared to the control diodes. The heterojunction diodes exhibit larger leakage when compared to the control diodes. Several researchers have observed similar increase in leakage across heterojunctions formed by Ge implantation into Si. The excess leakage is attributed to residual physical damage produced by the high dose Ge implant. Fukami et al [18] have observed that C implantation into the SiGe layer can lower the junction leakage by minimizing the defect density.

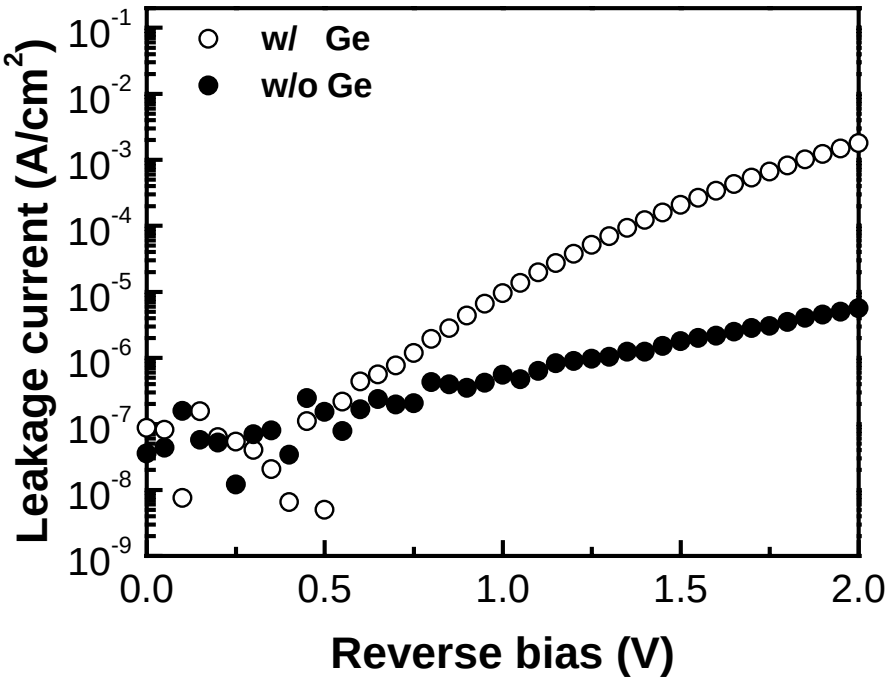


Figure 5.2 Comparison of p+/n diode reverse bias leakage currents for the heterojunction devices and control homojunction devices. The higher leakage current for the heterojunction devices is attributed in part to defects resulting from non-optimal recrystallization annealing after Ge implantation.

Shoji et al [20] have also observed that using a low substrate temperature during Ge ion implantation can lower the end of range defects and improve junction performance. Öztürk et al [21] have also observed that an optimization in the Ge implant conditions can result in high quality p+/n heterojunctions. Such implant induced damage can also be effectively removed with the addition of a low temperature (550°C) solid phase crystallization (SPC) anneal before the high temperature dopant diffusion and activation anneal [22]. The sheet resistance of the annealed p+ $Si_{1-x}Ge_x$ layers and the p+ Si layers was 376 Ω/€ and 2826 Ω/€ respectively. The $Si_{1-x}Ge_x$ layer shows a significant reduction in the sheet resistance which can be attributed to the higher activation efficiency of B in $Si_{1-x}Ge_x$ [13].

5.2.2 Selective Ge Deposition and Interdiffusion

Selective deposition of Ge can also be used to fabricate doped $Si_{1-x}Ge_x$/Si heterojunctions. Germanium films can be selectively deposited onto single crystalline Si substrates (as opposed to SiO_2 or Si_3N_4 regions) in a conventional low pressure chemical vapor deposition system. Thin poly-crystalline films of Ge (60 nm) were selectively deposited onto active regions of a Si substrate to form Ge/Si heterojunctions. Boron was then implanted into the Ge film and co-diffused along with the Ge atoms into the Si substrates to form heterojunction diodes.

The interdiffusion of Ge and Si in the absence of a third diffusing species is generally restricted to the interfacial regions [23,24]. Si diffusion into Ge is expected to provide the dominant diffusion flux in this system owing to the negligible diffusivity of Ge in Si and the relatively larger diffusivity of Si in Ge [24,25]. Figure 5.3 shows SIMS concentration-depth profiles that indicate such restricted interfacial diffusion behavior at undoped Ge/Si

interfaces observed after thermal annealing at 900°C. The co-diffusion of boron and Ge into the Si under similar annealing conditions displays markedly different diffusion mechanisms as shown in Fig. 5.4. In this case, interdiffusion between Si and Ge is significantly enhanced (compared to that of Fig. 5.3) and is seen to extend over the entire deposited Ge film thickness (60 nm). Also noteworthy is the correlation between the diffusion of Ge and B atoms. The B profile is observed to be contained within the Ge profile, giving rise to the in-situ formation of a heavily doped $Si_{1-x}Ge_x$/Si heterojunction.

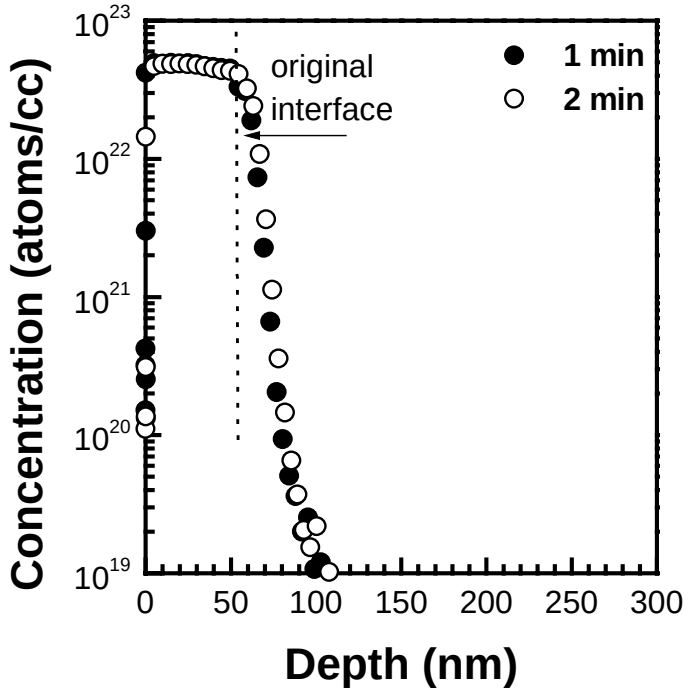


Figure 5.3 SIMS concentration-depth profiles indicating interdiffusion across undoped Ge/Si interfaces after 900°C annealing

In the case of the B doped $Si_{1-x}Ge_x$/Si heterojunctions shown in Fig. 5.4, the vertical diffusion beneath the original Ge/Si interface is ~ 40 nm after annealing for 2 min at 900°C. The B profile is also fairly abrupt with a slope of ~ 9 nm/decade. It can thus be seen that

selective Ge deposition followed by the co-diffusion of Ge and B atoms can lead to the formation of shallow and heavily doped $Si_{1-x}Ge_x$/Si heterojunctions.

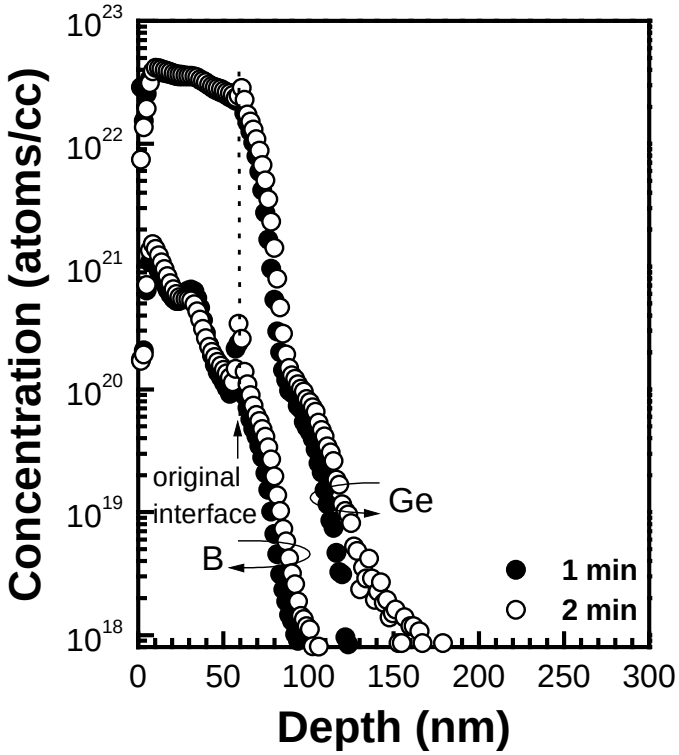


Figure 5.4 SIMS Ge concentration-depth profiles indicating the effect of B on interdiffusion across Ge/Si interfaces after 900°C annealing.

Figure 5.5 compares the sheet resistance and the leakage characteristics of the heterojunctions formed by interdiffusion and corresponding control homojunction devices after a series of annealing steps. The higher leakage seen in the heterojunction diodes is attributed to the presence of crystalline defects around the $Si_{1-x}Ge_x$/Si interface. The higher leakage seen in the heterojunction diodes is attributed to the presence of crystalline defects around the $Si_{1-x}Ge_x$/Si interface. Optimization of the implantation and annealing conditions is expected to lower the heterojunction leakage. The sheet resistance of the heterojunction and control junctions are 350, 300 and 748 Ω/∈ respectively.

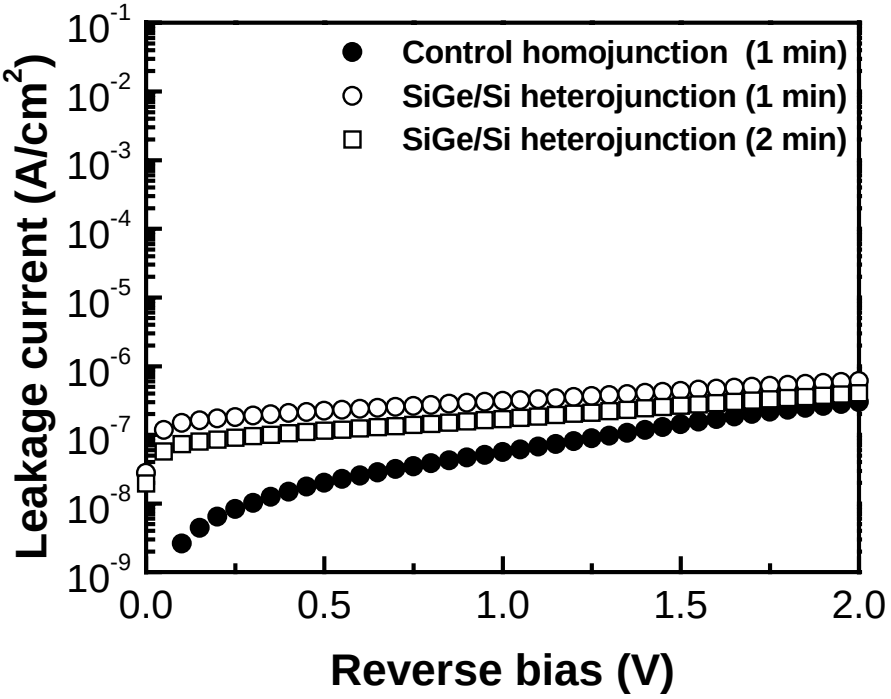


Figure 5.5 Comparison of diode reverse bias leakage currents for the heterojunction devices and homojunction (control) devices after annealing at 900°C. Heterojunction diode leakage is higher owing to a higher density of defects introduced at the $Si_{1-x}Ge_x$ interface

The next section illustrates the applicability of this approach in the fabrication of ultra-shallow S/D extension junctions for short channel MOSFET devices.

5.3 Elevated S/D PMOSFETs Formed by Selective Ge Deposition

The selective deposition of Ge on Si by LPCVD and the interdiffusion phenomena described earlier were used to fabricate sub-100 nm elevated S/D PMOSFETs. This section describes the processes used to fabricate these novel devices and their electrical performance.

5.3.1 Device Fabrication Process

The schematic of the process flow used for device fabrication is shown in Fig. 5.6. Detailed descriptions of individual processing steps have been provided in the Appendix. Local oxidation of silicon (LOCOS) was used to define active areas on Si substrates. After active-area definition by LOCOS, a retrograde channel doping profile was created by high-energy P^+ implantation (2×10^{13} cm^{-2}, 380 keV). This was followed by growth of 2.0 nm gate SiO_2 (by thermal oxidation) and deposition of ~150 nm undoped poly-$Si_{0.8}Ge_{0.2}$. $Si_{0.8}Ge_{0.2}$ was used as the gate material in this process in order to minimize gate depletion effects commonly observed with B doped poly-silicon electrodes. It has been shown earlier that poly-$Si_{0.8}Ge_{0.2}$ can successfully minimize gate depletion leading to enhanced short-channel performance [27]. A 20 nm thick layer of SiO_2 was then deposited as a hard mask (to prevent Ge deposition on the gate electrode), and electron-beam lithography was used to define gate electrode lines. Electron-beam lithography was performed using a Hitachi Nanowriter at the Center for Integrated Systems, Stanford University, Palo Alto, CA. This lithography tool is nominally capable of producing minimum line widths down to 80 nm using optimum combinations of pixel size and resist type/thickness. A pixel size of 40nm and a negative resist (UVN-30, 380nm thick) were used in this process.

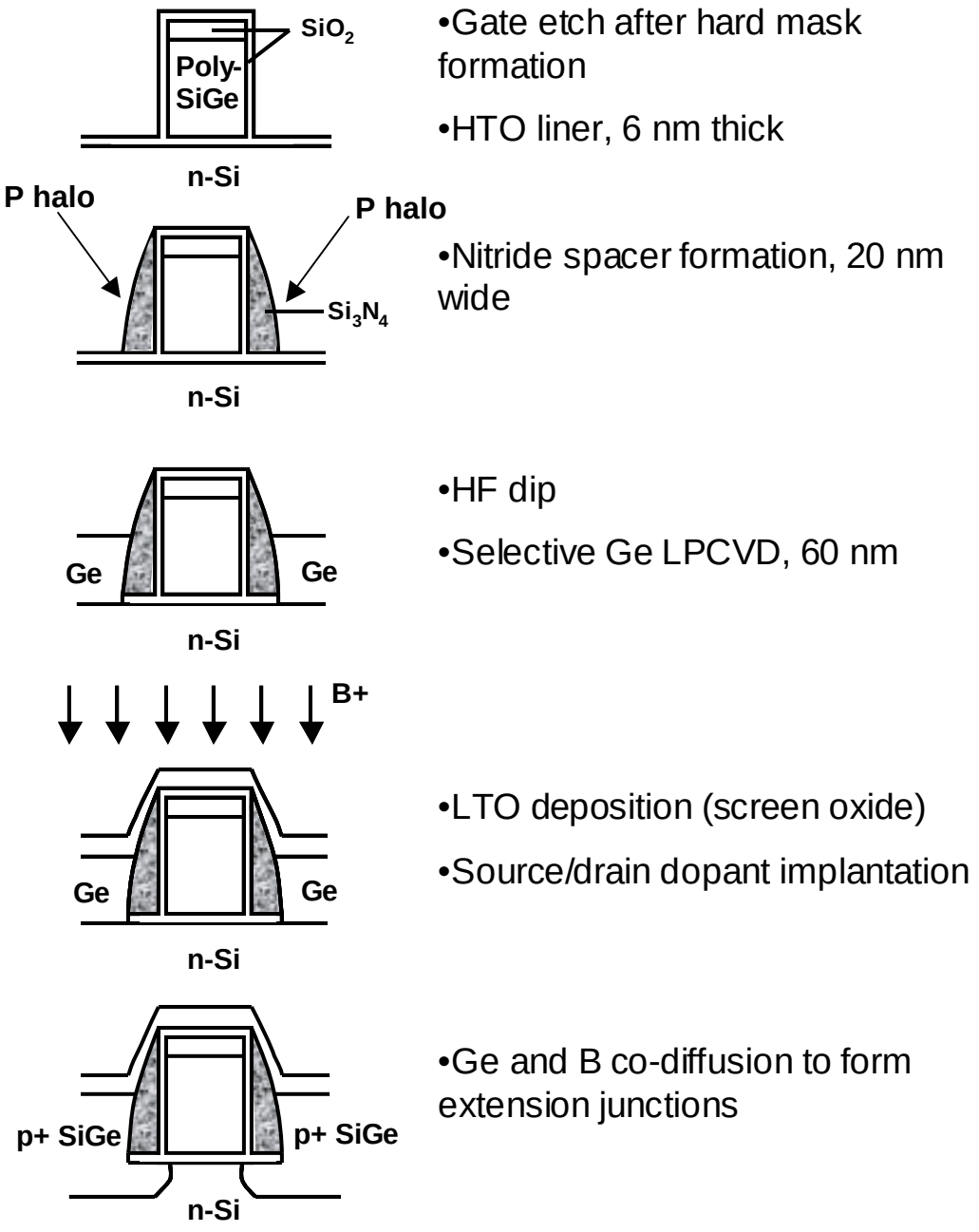


Figure 5.6 Front-end process sequence for elevated S/D PMOSFET fabrication using selective Ge deposition. The P halo implants were performed only on select die

Upon resist development, gate line widths were found to be slightly larger than their drawn dimensions. Gate lines with L_d = 60 nm were found to be ~80 nm using scanning electron

microscope (SEM) imaging. The wafers were then exposed to an O_2 reactive plasma in a Lam reactive-ion-etch (RIE) system to "ash" the lines, trimming ~10 nm from either side to give minimum line widths of ~60 nm. The same etch tool was then used to transfer the gate patterns from resist to the SiO_2/SiGe stack. After gate and hard-mask patterning, 25 nm wide sidewall spacers were formed by depositing and etching a dual-deposited layer consisting of 6 nm of high-temperature oxide (HTO) and 20 nm of silicon nitride. The HTO layer served as an etch stop for the subsequent nitride dry-etch process. At this step, several die on the wafer were selectively exposed using photoresist masking and were subsequently sent for large angle (30°) tilted ion implants designed to produce "halos" at the gate edge as shown in the schematic of Fig. 5.6. Two implant doses (P^+, 1×10^{13} cm^{-2} and $2\times 10^{13} cm^{-2}$) were used (on separate devices) for the halo implants. The halo implants were used to provide enhanced immunity from short-channel effects (by minimizing drain-induced barrier lowering (DIBL) and sub-surface punchthrough). Following selective halo implantation, remaining oxide over the S/D regions was removed using diluted HF, and 60 nm of Ge was selectively deposited in the S/D regions, in a conventional LPCVD furnace using GeH_4 as the gaseous source in an identical process as described earlier. A 20 nm capping layer of LPCVD SiO_2 (LTO) was then deposited over the wafers to prevent Ge evaporation during high temperature annealing and boron implantation (6×10^{15} cm^{-2} B^+ @ 5keV) was performed to dope the gate and S/D regions. 400 nm low temperature oxide (LTO) was subsequently deposited as a passivation layer, after which the devices were subjected to an RTA (900°C, 7 min) step. During this annealing step, Si atoms from the substrate diffuse into the Ge film while the Ge and B diffused vertically (as well as laterally under the sidewall spacer) into the Si, to form S/D extensions to the channel. Device fabrication was completed with standard contact and metallization processes using Al-2%Si metallurgy. A cross-sectional high-

resolution transmission-electron-microscope image through the gate stack is shown in Fig. 5.7 alongwith the details of the gate oxide and the gate edge.

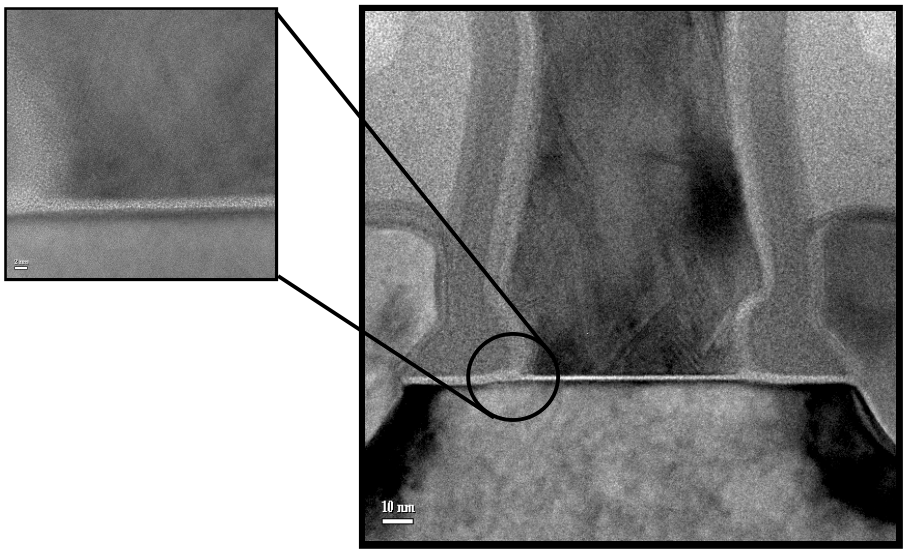


Figure 5.7 High resolution TEM image through a transistor gate stack fabricated using the process flow of Fig. 5.6. Non-optimal etch of the gate and the sidewall spacer is evident (manifested by the the T-shaped gate and the “foot” shaped protrusion of the spacers). The thermal oxide thickness is ~2 nm as desired. The gate length at the dielectric interface is 60 nm

5.3.2 Transistor Performance

The *I-V* characteristics for a transistor with L_{gate} = 80 nm are shown in Fig. 5.8. It can be seen that reasonable turn-off characteristics can be achieved for sub-100 nm gate length devices using the elevated S/D structure without the use of angled halo implants. With a conventional transistor structure (using implanted lightly doped drain (LDD) extensions), it is generally accepted that gate lengths under 100 nm will need angled halo implants to minimize encroachment of the drain field over the channel charge.

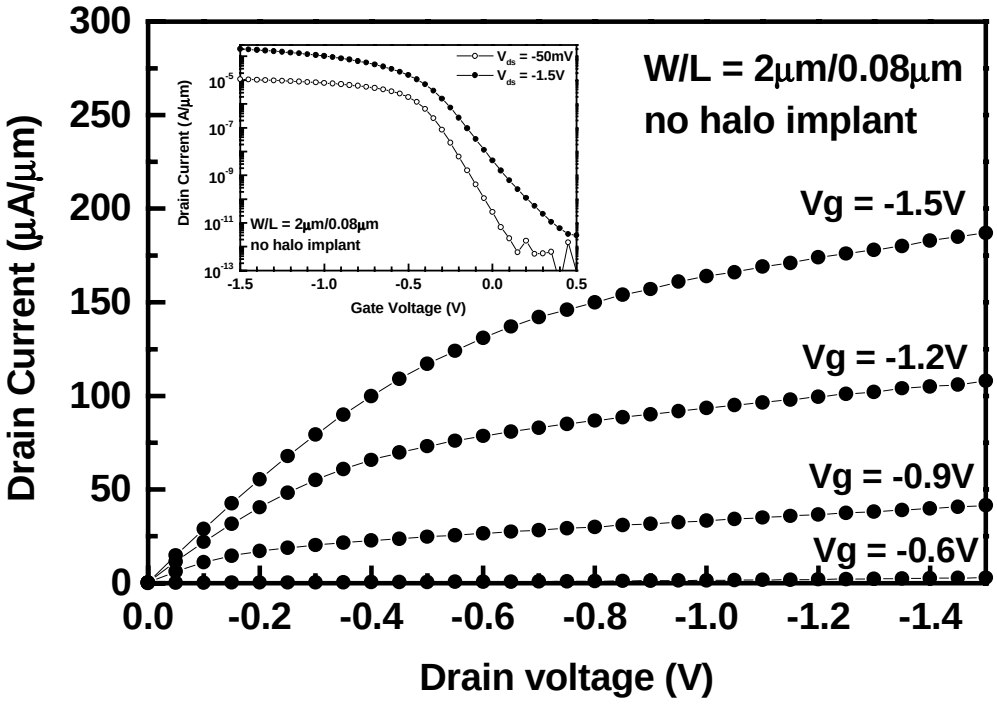


Figure 5.8 *I-V* Characteristics for an 80 nm gate length PMOSFET fabricated using the proposed elevated S/D process. I_D-V_D characteristics shown in the main figure as a function of gate bias, while I_D-V_G characteristics shown in inset. A reasonable sub-threshold swing of 82 mV/decade was observed.

One of the benefits of the proposed process is the enhanced immunity to short-channel effects without the use of halo implants up to 80 nm gate lengths. Figure 5.9 shows the *I-V* characteristics for a transistor with L_{gate} = 60 nm. The turn-off characteristics of this device indicate that with appropriate halo implants, the proposed elevated S/D structure with selective Ge deposition can be used to fabricate sub-70 nm physical gate length bulk PMOSFETs with acceptable short channel performance.

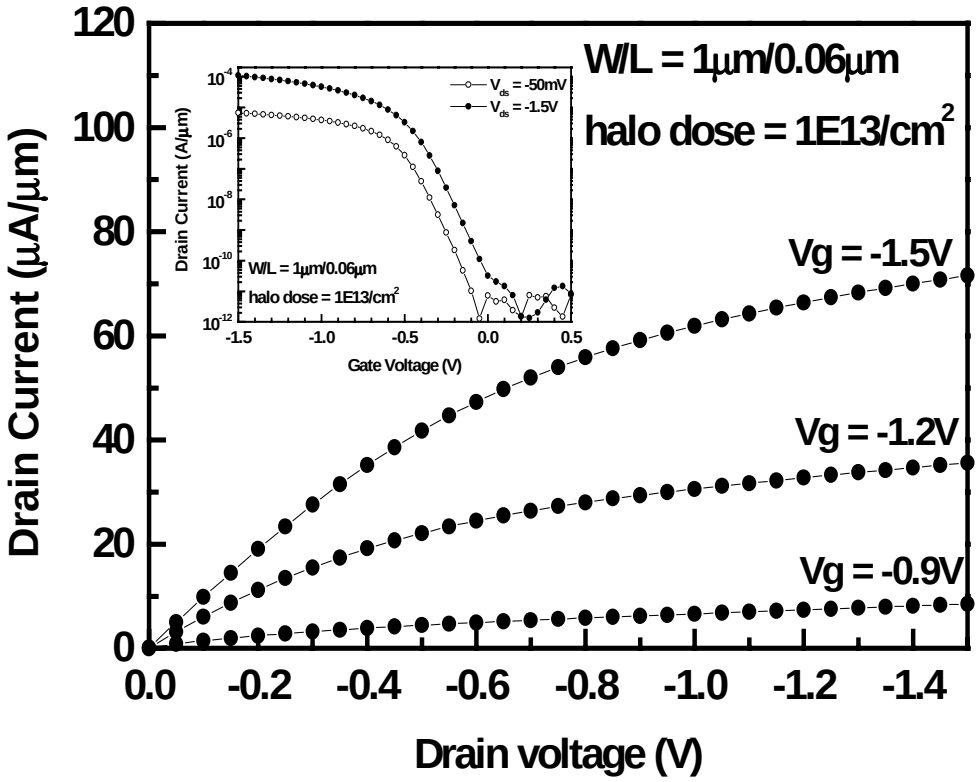


Figure 5.9 *I-V* Characteristics for a 60 nm gate length PMOSFET fabricated using the proposed elevated S/D process. I_D-V_D characteristics shown in the main figure as a function of gate bias, while I_D-V_G characteristics shown in inset. A reasonable sub-threshold swing of 84 mV/decade could be maintained with the use of a moderate halo implant.

Figures 5.10 (a-b) provide a comparison of V_{th} roll-off characteristics and sub-threshold swing variation with gate length for the elevated S/D devices and conventional devices with LDD extensions (with and without angled halo implants). The data show that the elevated S/D structure with $Si_{1-x}Ge_x$ S/D extensions effectively suppresses SCE for gate lengths down to 80 nm, whereas the conventional structure shows poor short-channel characteristics, (possibly due to the lack of a halo implant). For the same gate overdrive (V_g-V_t = -1.6 V), I_{Dsat} for the 80 nm gate length elevated S/D and control structures was 178 μA/μm and 128 μA/μm respectively, indicating a 40% improvement in drive current with the new device structure. It can also be seen that while the proposed structure precludes the need for halo implants for devices with gate lengths down to 80 nm, halo implants (or alternative channel engineering schemes) appear to be essential for devices with sub-80 nm gate lengths. The relatively low drive currents observed are attributed to non-ideal gate and sidewall spacer profiles and the resulting increase in channel series resistance. Optimization of the boron implant and gate and sidewall spacer dry etch processes are expected to lower the channel series resistance and further enhance the drive current.

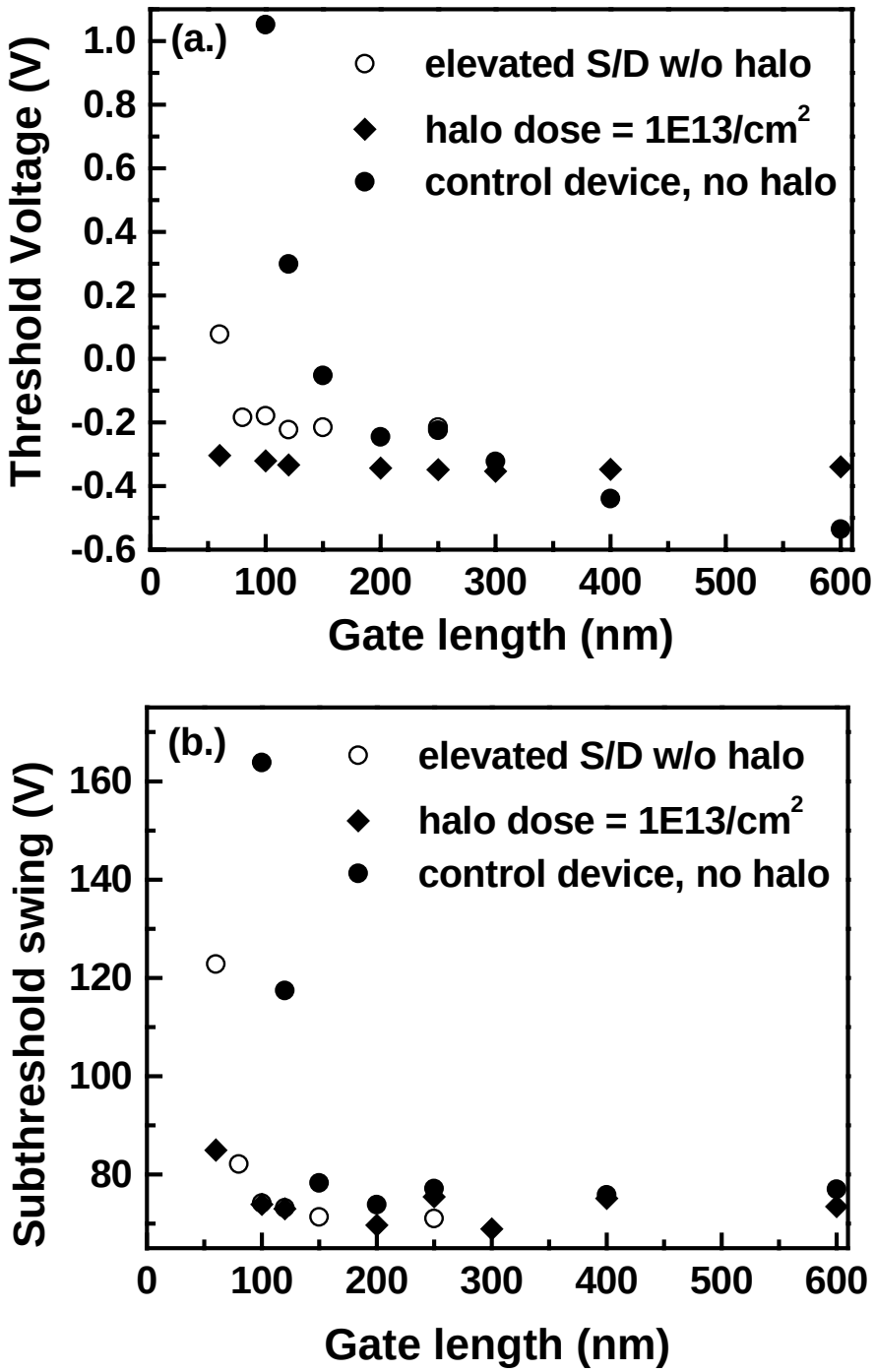


Figure 5.10 (a.) V_{TH} roll-off characteristics for the elevated S/D devices (with and without halo implants) and the control devices, (b.) Subthreshold swing variation with gate length for the three devices. Both figures indicate the superior short channel performance achieved with the elevated S/D structure

5.4 Additional Motivation for SiGe S/D MOSFETs

In addition to the enhanced immunity to short-channel effects descried earlier in this chapter, the use of SiGe alloys in the source and drain regions of a transistor are expected to provide a few additional advantages in terms of process integration and eventual device performance. These are highlighted in this section.

5.4.1 Low Temperature Dopant Activation

It is well known that the activation of B in Ge and SiGe alloys requires a smaller thermal budget than that required for dopant activation in Si [13,16]. This advantage of SiGe alloys can potentially be extremely useful in the development of "gate-first" CMOS process flows with advanced gate stack materials. The use of gate-first and gate-last processing schemes was discussed in an earlier chapter. It was noted that the use of high-k and metal gate stacks might impose severe thermal stability constraints and preclude the use of conventional gate-first process flows. This is because typical dopant activation temperatures in Si typically exceed 1000°C to ensure that the junctions do not add significantly to the channel series resistance. The use of an elevated S/D structure with selective Ge deposition as described earlier might provide a low temperature, gate-first compatible solution to this problem. This section presents preliminary results of experiments designed to ascertain the temperature limits of B, As and P activation in the selectively deposited and intermixed SiGe films.

Heterojunction diodes were fabricated using selective deposition of Ge onto single-crystalline Si (n or p type) exposed in windows of SiO_2 as described in Section 5.2.2. A typical cross-section of such a diode is shown in Fig. 5.11. After Ge deposition, dopant ions (B, P or As) were implanted into the wafers to form n+/p or p+/n diodes. Several different

thermal budgets were used to activate dopants in the Ge films in order to ascertain the lowest thermal budgets at which sufficient activation is achieved. Figure 5.12 (a-b) shows the variation of the sheet resistance with thermal budget for the B and As implanted diodes. It can be seen that a 30s anneal at 700°C is enough to obtain sufficiently low resistivity. It should be noted that on a homologous temperature scale (T/T_M), a temperature of 711°C for Ge corresponds to 1100°C for Si. It can also be seen that below a temperature of 600°C, furnace annealing (long time) is necessary for complete activation whereas comparable resistances can be obtained by rapid thermal annealing above 600°C. Figure 5.13 compares the sheet resistance for As and P doped Ge. It can be seen that for comparable thermal budgets, it is much easier to activate As than P.

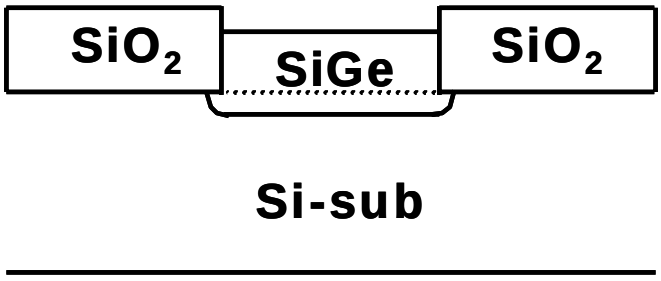


Fig. 5.11 Schematic cross-section through a heterojunction diode (after Ge-Si interdiffusion). The SiGe and the Si substrate have opposite conductivity

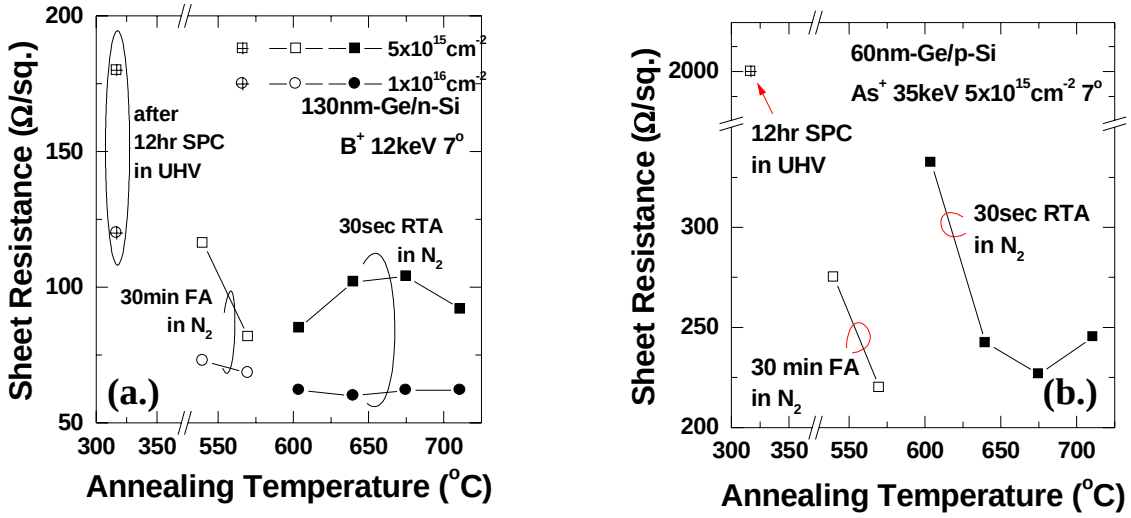


Fig. 5.12 (a-b) Variation of sheet resistance with thermal budget for B and As doped heterojunctions (From Ref. [28])

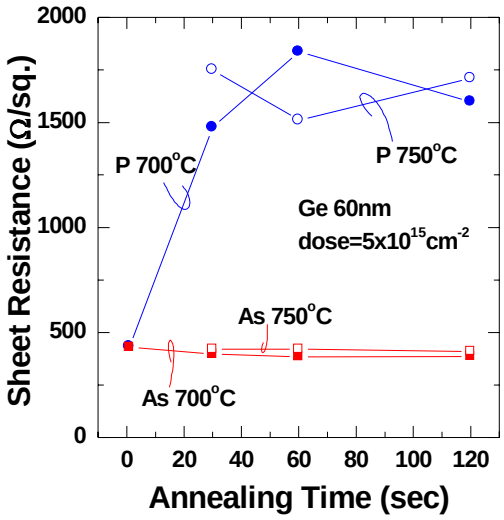


Fig. 5.13 Variation of sheet resistance with thermal budget for As and P doped heterojunctions. For comparable thermal budgets, As activation is significantly higher (From Ref. [28])

It can thus be seen that the elevated S/D process with selective Ge deposition can be an attractive integration route for advanced CMOS transistor fabrication when low thermal budget processing is desired. A potential problem with such an approach however, is that heterojunction leakage is much higher than the corresponding homo-junctions. This is believed to be due to the presence of defects at the interface owing to the lattice mismatch. As shown in Fig. 5.14, a 900°C annealing is needed to lower the junction leakage to acceptable levels.

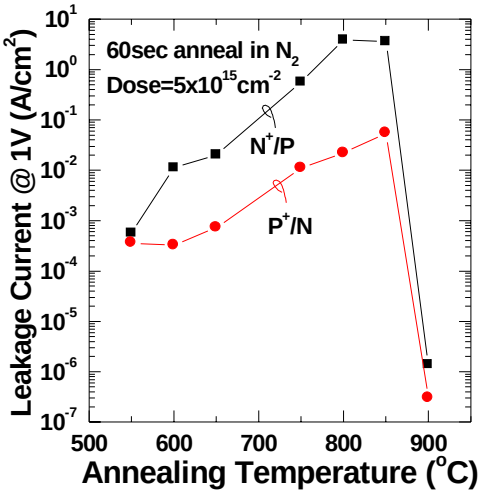


Fig. 5.14 Effect of annealing temperature on junction leakage for n+/p and p+/n diodes (As and B). It can be seen that the 900C anneal dramatically lowers the leakage possibly due to a reduction in the defect density at the heterojunction interface (From Ref. [28])

It should be noted however, that this is most likely not a fundamental barrier and several options to lower the junction leakage can be explored. One possibility is to use an extended time, low temperature solid-phase crystallization anneal before the higher temperature anneal as discussed in an earlier section. Another possibility is the implantation of Si (or Ge) ions in the junction area (after Ge deposition) followed by an annealing step to recrystallize the damage. The presence of additional Si atoms might serve to annihilate vacancies that might

have been injected during Ge deposition. Another approach could involve the amorphization of the Si lattice (with a high dose Ge or Si implant) prior to Ge deposition. This might prevent the injection of vacancies into the junction region during Ge deposition and lower the defect density arising from the mismatch. These possibilities need further investigation. It must be noted that the misfit dislocations under the lattice mismatched heterojunction interface are equilibrium defects that cannot be eliminated with additional processing. Further processing might however, eliminate threading dislocations and lower the stacking fault density and also spread out the dislocations in the direction of the Ge composition gradient. This might minimize the effect of the dislocations on the recombination properties of minority carriers near the junction. Another possibility is to lower the electrical activity of these defects by promoting segregation to the core of the dislocations.

5.4.2 Contact Resistance

An important component of source and drain junction design is the fabrication of low resistance contacts between the front-end (S/D regions) of the transistor and the subsequent back-end metallization schemes. In conventional CMOS devices, metal silicidation of the source, drain and poly-Si gate regions is performed simultaneously using either Ti or Co metals. The silicide regions significantly lower the contact resistance between the Si and the back-end metallurgy. Any change in the S/D materials as proposed in this chapter, will necessitate a re-design of the conventional silicidation processes. In particular, the introduction of Ge in the S/D regions will lead to the formation of ternary M-Si-Ge alloys instead of the conventional binary M-Si alloys. The electrical properties of these ternary metal germanosilicides need to be closely examined. Theoretically, the metal/p+SiGe

junction should provide a lower contact resistance owing to the reduced band gap. Since the S/D junctions are heavily doped, the contact between the silicide and the Si substrate is generally an ohmic (tunneling) contact. The tunneling contact resistance is exponentially dependent on the barrier height at the junction and has an inverse exponential dependence on the active dopant concentration in the Si. Since it is possible to achieve higher active dopant concentrations in SiGe (than in Si) and since the band gap of SiGe is smaller than that of Si (with most of the discontinuity being in the valence band), it is expected that the barrier height (and hence the contact resistance) across a metal/p+ SiGe junction will be much lower than the corresponding metal/Si junction. While a low contact resistance is achievable in principle, thermodynamic and kinetic constraints might prove to be limiting factors. It has been observed that during high temperature annealing of metal/SiGe interfaces, Ge atoms from the SiGe segregate at grain boundaries near the junction and this in turn leads to a dramatic increase in the contact resistance [29]. Innovative junction engineering schemes will need to be developed to counter this problem.

5.5 Summary

This work summarizes techniques to incorporate Ge in the source and drain regions of a conventional MOSFET and discusses the potential applications of $Si_{1-x}Ge_x$ in the fabrication of ultra-shallow S/D extensions in bulk Si PMOSFET devices. The use of $Si_{1-x}Ge_x$ in the S/D regions of conventional bulk-Si PMOSFETs is shown to result in excellent suppression of short channel effects for sub-100 nm gate length devices. Two alternative fabrication techniques are described which allow for the incorporation of Ge in the S/D regions of a bulk MOSFET. $Si_{1-x}Ge_x$ allows for the fabrication of shallow extension junctions with high concentrations of electrically active dopant atoms and low sheet resistance and can also

lower the contact resistance between the S/D regions and the back end metallization schemes. The relatively low activation annealing temperatures needed for dopant activation might prove advantageous for the eventual integration of high-k dielectrics and metal gate electrodes in CMOS gate stacks. These techniques can thus enable the scaling of bulk-Si transistor technology into the sub-50 nm gate length regimes.

5.4 References

1. International Technology Roadmap for Semiconductors, Semiconductor Industry Association, 2001
2. F. A. Trumbore, "Solid Solubilities of Impurity Elements in Germanium and Silicon", Bell Syst. Tech. Journal, 39, p.205, 1960
3. R. A. Colclaser, Microelectronics: Processing and Device Design, Wiley Interscience, New York, 1980
4. P. Stolk, H.-J. Gossmann, D. Eaglesham, D. Jacobson, C. Rafferty, G. Gilmer, M. Jaraiz, J. Poate, H. Luftman, T. Haynes, "Physical Mechanisms of Transient-Enhanced Dopant Diffusion in Ion Implanted Silicon", Journal of Applied Physics, v. 81, 9, p. 6031, May 1997
5. A. Agarwal et al, "Boron Enhanced Diffusion of Boron: The Limiting Factor for Ultra-Shallow Junctions", International Electron Devices Meeting 1997, Technical Digest, p. 467, Dec. 1997
6. M. Foad, A. Murrell, E.J.H. Collart, G. de Cock, D. Jennings, M. Current, "Practical aspects of forming ultra-shallow junctions by sub-keV boron implants"**,** Symp. Proceedings, Mater. Res. Soc Symposium, p. 55, April 1999
7. Murrell, E.J.H. Collart, M. Foad, G. de Cock, D. Wagner, Y. Matsunaga, "Production-worthy shallow junction formation for sub-micron technologies using electron-volt ion implantation in the Applied Materials XR LEAP™",1998 International Conference on Ion Implantation Technology Proceedings, v.1, p. 215, 1998
8. A. Agarwal et al, "Critical issues in ion implantation of silicon below 5 keV: defects and diffusion", Materials Science & Engineering A, v. A253, p. 269, September 1998
9. A. Nishida, E. Murakami and S. Kimura, "Characteristics of low-energy BF_2 or As-implanted layers and their effect on the electrical performance of 0.15 μm MOSFET's", IEEE Transactions on Electron Devices, Vol. 45, p. 701, 1998
10. S. Talwar, G. Verma and K. H. Weiner, "Ultra-shallow, abrupt, and highly-activated junctions by low-energy ion implantation and laser annealing", International Conference on Ion Implantation Technology, v. 2, p. 1171, 1998
11. K. S. Jones, et al, "Transient enhanced diffusion after laser thermal processing of ion implanted silicon", Applied Physics Letters, v. 75, 23, p. 3659, December 1999

12. K. Itoh, Y. Sasaki, T. Mitsuishi, M. Miyao, M. Tamura, “Thermal behavior of B, P and As atoms in supersaturated Si produced by ion implantation and pulsed-laser annealing”, Japanese Journal of Applied Physics, Part 2 (Letters), vol.21, 5, p. L245, May 1982
13. T.-J. King, J. P. McVittie, K. C. Saraswat, “Electrical Properties of Heavily Doped Poly-crystalline Silicon-Germanium Films”, IEEE Transactions on Electron Devices, vol. 41, 2, p. 228, Feb. 1994
14. H. Takeuchi, W.-C. Lee, P. Ranade, T.-J. King, “Improved PMOSFET short-channel performance using ultra-shallow $Si_{0.8}Ge_{0.2}$ source/drain extensions”, International Electron Devices Meeting 1999, Technical Digest, p. 501, Dec. 1999
15. P. Ranade, H. Takeuchi, V. Subramanian, T.-J. King, “Observation of Boron and Arsenic Mediated Diffusion Across Germanium/Silicon Interfaces”, Electrochemical and Solid State Letters, v. 5, no. 2, 2002, p. G5-G7, February 2002
16. K. Jones and E. Haller, “Ion Implantation of Boron in Germanium”, Journal of Applied Physics, 61, 7, p. 2469, April 1987
17. S. Gannavaram, N. Pesovic and M. Öztürk, “Low Temperature (≤800°C) Recessed Junction Selective Silicon-Germanium Source/Drain Technology for sub-70 nm CMOS”, International Electron Devices Meeting 2000, Technical Digest, p. 437, Dec. 2000
18. H. Huang, K. Chen, T. Huang, T. Chao, G. Huang, C. Chien, C. Chang, “Improved Low Temperature Characteristics of P-Channel MOSFETs with $Si_{1-x}Ge_x$ Raised Source and Drain”, IEEE Transactions on Electron Devices, v. 48, 8, p.1627, August 2001
19. A. Fukami, K. Shoji, T. Nagano, C. Y. Yang, “Characterization of SiGe/Si heterostructures formed by Ge^+ and C^+ implantation”, Applied Physics Letters, vol.57, 22, p. 2345, Nov. 1990
20. K. Shoji A. Fukami, T. Nagano, T. Tokuyama, C. Y. Yang, “Improved crystalline quality of $Si_{1-x}Ge_x$ formed by low temperature germanium ion implantation”, Applied Physics Letters, vol. 60, 4, p.451, Jan. 1992
21. M. C. Öztürk, et al,“Optimization of the germanium preamorphization conditions for shallow-junction formation”, IEEE Transactions on Electron Devices, vol.35, no.5, p. 659, May 1988
22. L. Csepregi, W. Chu, H. Muller, J. Mayer, T. Sigmon, “Influence of thermal history on the residual disorder in implanted (111) silicon”, Radiation Effects, v. 28, p. 227, 1976

23. S. Chang and K. Wang, Interdiffusion in a symmetrically strained Ge/Si superlattice", Applied Physics Letters, 54, 13, p. 1253, 1989
24. J.-M. Baribeau, R. Pascual, S. Saimoto, "Interdiffusion and strain relaxation in $(Si_mGe_n)_p$ superlattices", Applied Physics Letters, 57, 15, p. 1502, 1990
25. A. Csik, D. L. Beke, G. A. Langer, Z. Erdelyi, L. Daroczi, K. Kapta, M. Kis-Varga, "Non-linearity of diffusion in amorphous Si-Ge multilayers", Vacuum, vol.61, p. 297, 2001
26. A. Csik, G. A. Langer, D. L. Beke, Z. Erdelyi, M. Menyhard and A. Sulyok, Interdiffusion in amorphous Si/Ge multilayers by Auger depth profiling technique", Journal of Applied Physics, vol.89, p. 804, 2001
27. W.-C. Lee, T.-J. King and C. Hu, "Observation of reduced boron penetration and gate depletion for poly-$Si_{0.8}Ge_{0.2}$ gated PMOS devices", IEEE Electron Device Letters, vol.20, 1, p. 9, Jan. 1999
28. H. Takeuchi, P. Ranade and T.-J. King, "Low Temperature Dopant Activation Technology Using Elevated Ge S/D Structure", Proceedings of the 1st International SiGe Device and Technology Meeting (ISTDM), Nagoya, Japan, January 2003 (to be published)
29. J. Liu, H. Mo and M.C. Ozturk, "A study on solid phase reactions of Ni and Pt on Si-Ge alloys as contacts to ultra-shallow p/sup +/n junctions for CMOS technology nodes beyond 70nm", Gate Stack and Silicide Issues in Silicon Processing II, Materials Research Society Symposium Proceedings Spring 2001, Vol. 670, p. K7.2.1, April 2001

Chapter 6

Conclusion

6.1 Summary

This work has sought to address some of the most pressing issues in CMOS transistor scaling beyond the 90 nm technology node. As of the writing of this dissertation, the 90 nm CMOS technology node is under development at most leading semiconductor chip manufacturers and several companies are poised to introduce products based on this technology by the end of 2003. Concomitantly, some of the major players in the IC industry are initiating development efforts for the 65 nm technology node, expected to be ready for manufacturing in late 2005. Most leading IC manufacturers have indicated that devices based on the 90 nm technology node will not be significantly different from their predecessors in terms of gate stack materials and processes. However, after nearly 30 years of continued CMOS scaling using Si-based gate stack materials, the 65 nm node might very well represent an inflection point in terms of gate stack materials and front-end process integration techniques. In developing a 65 nm technology (minimum gate lengths expected to be < 35 nm), it is likely that the industry will have to address some of the materials and process integration challenges discussed herein.

6.1.1 Metal Gate Materials and Process Integration

As discussed in Chapter 2, the task for finding a replacement material for poly-silicon gate electrodes is by no means trivial. Numerous problems related to the selection of materials

and the identification and development of optimal processing techniques (deposition, lithography and etching) are needed before a suitable metal gate process can be transferred to manufacturing. In fact, the integration challenges are so intricate that the trade-off between process complexity and performance enhancement needs to be critically examined before a replacement technology can be identified. The goal of this thesis has been to identify materials and develop robust integration schemes that can be incorporated into manufacturing with as little disruption as possible. The development of ion implantation enabled metal work function engineering is one such integration scheme. As discussed in Chapters 3 and 4, molybdenum appears to be a promising material for gate work function engineering. However, the range of work functions observed with nitrogen implantation is insufficient for bulk-Si CMOS and more applicable for FDSOI-CMOS structures. Some of the distinct contributions made are highlighted in a subsequent section.

6.1.2 Ultra-Shallow Junction Formation

Chapter 5 discussed some of the challenges involved in scaling the transistor extension junction depths below 30 nm. This is another area where novel materials and processes are likely to be employed beyond the 65 nm technology node. Silicon-Germanium is a very promising material for use in the source and drain regions of the transistor. As discussed in Chapter 5, it can provide for a few distinct advantages over conventional Si source-drain processes. One of these is the high activation efficiency of dopants in SiGe as manifested in the improved sheet resistances over Si junctions for comparable thermal budgets. This also implies that SiGe source-drains can provide an easy pathway for the development of low temperature (< 600°C) transistor fabrication processes. This can be especially important for transistors employing advanced gate stack materials which otherwise might not be able to

withstand the high thermal budgets generally employed for dopant activation. SiGe also has a lower bandgap (than Si) and this is expected to provide for a lower contact resistance between the back-end metallurgy and the source-drain regions. Most published work on SiGe source-drain MOSFETs has discussed the use of epitaxially deposited SiGe. While very effective, this process significantly adds to the process cost and adversely affects throughput. The elevated Ge S/D MOSFET process described in Chapter 5 is a cost-effective alternative which can potentially provide for most of the process enhancements generally expected from the use of SiGe in the S/D regions. Coupled with metal gate electrodes and high-k dielectrics in the gate stack of the MOSFET, the proposed Ge elevated S/D process is a very attractive alternative for future device technology. Some of the distinct contributions arising from this work are highlighted in the next section.

6.2 Contributions

Some of the contributions of this work to the field of semiconductor device processing are summarized in this section. Over the last four years, the results of the research described herein have been published in numerous technical journals and presented at leading technology conferences. This thesis has drawn heavily from these prior publications and a complete bibliography has been provided at the end of this chapter for interested readers.

In the area of alternative gate stack materials, a detailed survey of refractory metals has led to the identification of molybdenum as an attractive gate metal for PMOS transistors. The thermal stability of molybdenum was evaluated on several gate dielectrics and a high level of compatibility was observed on most dielectrics of interest. The mid-gap work function of tungsten and its applicability to fully-depleted SOI transistor technology has also been identified. In arriving at these conclusions, metals like Ti and Ni were also evaluated. Ti

was successfully used as a NMOS gate metal on Si_3N_4 gate dielectric in a gate-first process, while it was used on SiO_2 gate dielectric in a gate-last process. Ni was used as a PMOS gate metal in a gate-last process and its thermal stability on SiO_2 was also evaluated. The demonstration of Ti and Mo gate NMOS and PMOS transistors on a single Si substrate was the first demonstration of a dual-metal gate CMOS process.

One of the seminal contributions of this work has been the development of a framework for correlating the work function of a metal film to its structural and chemical morphology, particularly when applied to transistor gate electrodes. The use of this framework was demonstrated through experiments with molybdenum and tungsten gate electrodes. The implantation of nitrogen to modify thin film chemistry was identified as an easily integrable approach to alter the work function of Mo gate electrodes. The applicability of this process was demonstrated through the fabrication of bulk-Si CMOSFETs, ultra-thin body (UTB) FETs and FinFETs. The demonstration of tunable work function metal gate bulk-Si CMOS devices was the first of its kind, while the demonstration of tunable work function metal gate FinFETs was among the first to be reported. Some of the ideas described herein have formed the basis of subsequent work by other research groups that optimized the nitrogen implantation process and suggested an alternative process to introduce N into Mo films (See Ref. [8] of Chapter 3 and Ref. [20] of Chapter 4)

The development of the elevated Ge source-drain structure forms another significant contribution of this work. The novel elevated S/D process described in Chapter 5 provides a cost-effective approach to source-drain engineering. It was shown that this process can lead to excellent short-channel effects in sub-100 nm bulk-Si PMOSFETs without the use of any additional channel engineering schemes like halo or pocket implantation. In addition, the scalability of this approach to 60 nm gate lengths was verified and it was shown that with the

use of a medium halo implant, this approach can be effective beyond the 90 nm technology node. In subsequent work, this approach was also used in the fabrication of ultra-thin body FET fabrication and FinFET fabrication. In the former, where an elevated S/D process is necessary given the ultra-thin Si body, this approach gave excellent results, while the relative performance of elevated S/D FinFETs was much higher than their non-elevated counterparts. Another distinct advantage of the elevated Ge process is the ability to obtain low resistance source and drain regions at significantly reduced temepratures (<600°C) as discussed in Chapter 5. This technique might allow for a gate-first approach to be used with advanced high-k gate dielectrics and gate electrode materials that ordinarily would not be able to withstand the high temepratures used for dopant activation in silicon.

6.3 Suggestions for Future Work

In general, this work has been very exploratory in nature. Some of the ideas described herein have been demonstrated in devices and preliminary performance has been verfied. A more rigorous characterization of devices will however be necessary for some of the ideas described to be used in volume production. Suggestions for future work directly and indirectly related to the concepts described earlier are highlighted in this section.

In the case of molybdenum gate technology, although initial results have been encouraging, a detailed examination of the long-term reliability of MOSFETs made using the tunable gate process needs to be performed. Such an analysis should focus on the effects of unintentional nitrogen penetration (implant channeling) into the gate dielectric and potential impact on the time-dependent-dielectric breakdown (TDDB) of the gate dielectric. Another area that needs investigation is the potential penetration of nitrogen into the transistor

channel and impact on transistor performance (subthreshold swing, carrier mobility in the channel, etc.).

The optimization of a number of variables including the metal film thickness, implant dose, implant energy, tilt angle and rotation needs to be performed. The impact of metal thickness variation across a 300 mm Si substrate and the resulting variation in the implant profiles on the device threshold voltage needs to be verified.

For a viable metal gate technology to be developed, it will be essential to optimize gate metal reactive ion etch (RIE) processes so that ideal gate profiles (vertical sidewalls) can be obtained. In addition, it is essential to have high selectivity etch processes with respect to the gate dielectric (although the selectivity of the Cl_2/O_2 Mo RIE process to SiO_2 has been verfied, its applicability to high-k dielectrics needs to be demonstrated).

The impact of Ar implantation on the work function of Mo gates was described in Chapter 3. It was shown that although Ar implantation leads to amorphization of the metal film, the change is temporary and the film crystallizes upon high temperature annealing. As such, the crystallization process will not allow this approach to be used in a CMOS process (given the high thermal budgets involved). However, potential techniques to increase the recrystallization temperature can be investigated. One approach would be to implant additional species (eg. co-implantation of Si and N to form ternary M-Si-N alloys) that might keep the film in the amorphous state over wider thermal budgets.

The work described in this thesis employed physical vapor deposition (sputtering) of metal films as gate electrodes. However, from a manufacturing perspective, chemical vapor deposition is believed to be a more robust alternative. Enhanced control over deposition parameters and hence microstructure, more conformal deposition over vertical sidewalls, reduced physical damage to the underlying dielectric film (improved reliability) are some of

the advantages of CVD over PVD. Metal CVD is however a challenging area in its own right and the development of CVD processes for some of the metals discussed earlier presents many research opportunities.

While most of this work has focused on the development of metal gate materials, it is important to continue to explore opportunities with Si based gate electrodes. The use of SiGe as a gate electrode has already been documented and has been shown to be less prone to gate depletion and dopant penetration. Recent investigations into the compatibility of SiGe with high-k dielectrics like HfO_2 have also been encouraging. Ways to achieve super-saturated active gate dopant concentrations in Si or SiGe gates can also be explored. Laser annealing can provide such a possibility by raising the local temperature in the gate film to near the melting point over nano-second pulses (thus minimizing possible chemical reaction between the gate electrode and the gate dielectric while maximizing dopant diffusivity and activation). In this regard, SiGe is more attractive since it has a lower melting point and would hence require a smaller laser intensity or shorter annealing times. This would also mean enhanced stability of the underlying gate dielectric.

Recently, the use of doped metal silicides for CMOS gate applications has also been reported (See Ref. [13] of Chapter 4). It was observed that n+ or p+ doped Ni-silicide work functions are quite different and span a range in the lower half of the Si band gap depending upon the dopant concentration used. The exact mechanism behind this observation and the extension of this concept to other metal systems warrants further investigation.

In the area of shallow junction fabrication, the application of the elevated source-drain approach for short-channel PMOSFET devices has been demonstrated. The application to NMOSFETs using As or P dopant atoms needs to be investigated. Also, while n+/p and p+/n heterojunction diodes have been fabricated and demonstrate very low sheet

resistance, the junction leakage currents are much higher than the specified limit. Further optimization needs to be performed to lower the junction leakage currents. It is surmised that the leakage current is high because of a high density of point defects and misfit dislocations at the SiGe/Si interface after annealing. Ways to lower the defect density could include using an amorphization implant (Si or Ge) in the active regions before Ge deposition. This would ensure Ge growth on amorphous Si at the surface and might lessen the injection of point defects into the substrate. The crystalline Si substrate beneath the amorphized region can then serve as a template to recrystallize the SiGe regions upon subsequent annealing.

One of the most important components of S/D design however is the formation of low specific resistance contacts between the junctions and the back-end metallization schemes. The use of SiGe in the S/D regions provides the opportunity to lower the contact resistance (owing to the smaller band gap of SiGe). However, there are several thermodynamic constraints that need to be addressed. Germanium atoms in the SiGe films are known to segregate out at grain boundaries near the metal/SiGe interface and this in turn increases the contact resistance significantly. Possible solutions to this problem include the insertion of a intermediate layer (Zr or Ti) between the SiGe and the silicidation metal (Ni, Ti or Co) (See Ref. [28] of Chapter 5). It was observed that the Zr forms a stable germanosilicide (i.e. no observed loss of Ge) however the use of Zr alone leads to diffusion of B or P to the metal/germanosilicide interface). The use of thin Zr (5 nm) intermediate layers between Pt and the SiGe were observed to prevent dopant redistribution and Ge loss from the germanosilicide. Such a technology can be attractive and needs to be explored further.

6.4 Bibliography

1. H. Takeuchi, P. Ranade and T.-J. King, “Low Temperature Dopant Activation Technology Using Elevated Ge S/D Structure”, Proceedings of the 1st International SiGe Device and Technology Meeting (ISTDM), Nagoya, Japan, January 2003 (to be published)
2. H. Takeuchi, P. Ranade and T.-J. King, “Suppression of Boron TED by Low Temperature Solid Phase Crystallization Anneal Prior to Dopant Activation”, IEEE Transactions on Electron Devices (to be published)
3. P. Ranade, Y.-K. Choi, D. Ha, A. Agarwal, M. Ameen and T.-J. King, “Tunable Work Function Metal Gate Technology for FDSOI-CMOS”, IEEE International Electron Devices Meeting, Technical Digest, p. 363, Dec. 2002
4. Y.-K. Choi, L. Chang, P. Ranade, J.-S. Lee, D. Ha, S. Balasubramanian, A. Agarwal, M. Ameen, T.-J. King and J. Bokor, “FinFET Process Refinements for Improved Mobility and Gate Work Function Engineering”, IEEE International Electron Devices Meeting, Technical Digest, p. 259, Dec. 2002
5. D. Ha, P. Ranade, Y.-K. Choi, J.-S. Lee, T.-J. King and C. Hu, “Ultra-Thin Body Silicon-on-Insulator (UTB-SOI) MOSFET with Metal Gate Work-Function Engineering for sub-70 nm Technology Node”, Extended Abstracts of the 2002 International Conference on Solid State Devices and Materials, p. 782, Sep. 2002
6. P. Ranade, H. Takeuchi, W.-C. Lee, V. Subramanian and T. -J. King, “Application of Silicon-Germanium in the Fabrication of Ultra-Shallow Extension Junctions for sub-100nm PMOSFETs”, IEEE Transactions on Electron Devices, v. 49, no. 8, Aug. 2002, p. 1436
7. H. Takeuchi, P. Ranade, V. Subramanian, T.-J. King, “Observation of Dopant-Mediated Intermixing at Ge/Si Interface”, Applied Physics Letters, v. 80, no. 20, May 2002, p. 3706
8. Y.-C. Yeo, P. Ranade, T.-J. King, and C. Hu, “Effects of High-k Gate Dielectric Materials on Metal and Silicon Gate Workfunctions", IEEE Electron Device Letters, vol. 23, no. 6, Jun. 2002, p. 342

9. P. Ranade, H. Takeuchi, V. Subramanian and T. -J. King, “A Novel Elevated Source/Drain PMOSFET Formed by Ge-B/Si Intermixing”, IEEE Electron Device Letters, v. 23, no. 4, April 2002, p. 218
10. I. Polishchuk, P. Ranade, T.-J. King, C. Hu, “Dual Work Function CMOS Transistors Fabricated by Ni-Ti Interdiffusion”, IEEE Electron Device Letters, v. 23, no. 4, April 2002, p. 200
11. P. Ranade, Q. Lu, I. Polishchuk, H. Takeuchi, C. Hu and T. -J. King, “Dual Work Function Metal Gate Technology for Future CMOS Devices”, American Vacuum Society, 3rd International Conference on Microelectronics and Interfaces, Symposium Proceedings, February 2002, p. 154
12. P. Ranade, H. Takeuchi, V. Subramanian and T. -J. King, “Observation of Boron and Arsenic Mediated Interdiffusion Across Germanium/Silicon Interfaces”, Electrochemical and Solid State Letters, v. 5, no. 2, February 2002, p. G5
13. R. Lin, Q. Lu, P. Ranade, T. -J. King, C. Hu, “An Adjustable Work Function Technology Using Mo Gate for CMOS Devices”, IEEE Electron Device Letters, v. 23, no. 1, January 2002, p. 49
14. P. Ranade, H. Takeuchi, T. -J. King and C. Hu, “Work Function Engineering of Metal Gate Electrodes with Nitrogen Implantation”, Electrochemical and Solid State Letters, v. 4, no. 11, November 2001, p. G85
15. I. Polishchuk, P. Ranade, T.-J. King and C. Hu, “Dual Work Function Metal Gate CMOS Gate Technology Using Metal Interdiffusion”, IEEE Electron Device Letters, v. 22, no. 9, September 2001, p. 444
16. Y.-C. Yeo, Q. Lu, P. Ranade, H. Takeuchi, K. J. Yang, I. Polishchuk, T. -J. King, C. Hu, S.C. Song, H. F. Luan, D. -L. Kwong, “Dual-Metal Gate CMOS Technology with Ultra-Thin Silicon Nitride Gate Dielectric”, IEEE Electron Device Letters, v. 22, no. 5, May 2001, p. 227
17. P. Ranade, R. Lin, Q.Lu, Y. -C. Yeo, P. Ranade, H. Takeuchi, T. -J. King, C. Hu, “Molybdenum Gate Electrode Technology for Deep Sub-micron CMOS Generations”, Materials Research Society Symposium Proceedings, v. 670, K5.2.1, April 2001
18. I. Polishchuk, P. Ranade, T.-J. King and C. Hu, “Dual Work Function CMOS Gate Technology Based on Metal Interdiffusion”, Materials Research Society Symposium Proceedings, v. 670, K5.1.1, April 2001

19. Q. Lu, R. Lin, P. Ranade, T. -J. King, C. Hu, “Metal Work Function Adjustment for Future CMOS Technology”, Symposium Proceedings, IEEE VLSI Technology Symposium, p. 45, June 2001
20. Y.-C. Yeo, P. Ranade, Q. Lu, R. Lin, H. Takeuchi, T. -J. King, C. Hu, “Gate Work Function Engineering for CMOS Transistors with High-k Dielectrics”, Symposium Proceedings, IEEE VLSI Technology Symposium, p. 49, June 2001
21. Q. Lu, R. Lin, P. Ranade, Y.-C. Yeo, X. Meng, T.-J. King, C. Hu, H. Luan, S. Lee, W. Bai, C.-H. Lee, D. -L. Kwong, X. Guo, X. Wang, T.-P. Ma, “Molybdenum Metal Gate MOS Technology for Post-SiO_2 Gate Dielectrics”, IEEE International Electron Devices Meeting, Technical Digest, p. 641, December 2000
22. Q. Lu, Y. -C. Yeo, P. Ranade, H. Takeuchi, T. -J. King, C. Hu, “Dual-Metal Gate CMOS Fabrication”, Symposium Proceedings, IEEE VLSI Technology Symposium, p. 72, June 2000
23. P. Ranade, Y. -C. Yeo, Q. Lu, H. Takeuchi, T. -J. King, C. Hu, “Molybdenum as a Gate Electrode for Deep Sub-micron CMOS Processing”, Materials Research Society Symposium Proceedings, v. 611, C3.2.1, April 2000
24. H. Takeuchi, W.-C. Lee, P. Ranade, T. -J. King “Improved PMOSFET Short-Channel Performance Using Ultra-Shallow $Si_{0.8}Ge_{0.2}$ Source/Drain Extensions”, IEEE International Electron Devices Meeting, Technical Digest, p. 501, Dec. 1999
25. Y.-K. Choi, Y.-C. Jeon, P. Ranade, H. Takeuchi, T. -J. King, J. Bokor, C. Hu, “Ultra-thin body SOI Device with 30nm Gate Length and Selectively Raised Ge S/D”, IEEE 58th Device Research Conference, Conference Proceedings, p. 23, June 2000

Appendix

Process Flow for SiGe Elevated S/D PMOSFET Fabrication

STEP	PROCESS	PROCESS SPECIFICATION	EQUIPMENT	COMMENT
1.0	Si wafers	4" p-type (Resistivity = 38-63 Ω·cm)		
2.0	**Pad oxide / Nitride**			
2.1	Pre-clean	Pirahna clean 10 mins + 25:1 HF 20sec	Sink6	
2.2	Initial oxidation	1000°C, 21min O_2, 20min N_2	Tylan5	T_{ox}≈ 250 Å
2.3	Nitride deposition	800°C	Tylan9	T_{SiN} ≈ 1300 Å
3.0	**Active definition**			
3.1	Prebake	Hard bake and HMDS coat	Primeoven	
3.2	PR coating	New OCG I-line	SVGCOAT	T_{PR}= 1.1μm
3.3	Exposure	Use posted exposure time and focus	GCAWS	
3.4	Development	120°C, 60sec, I-line developer	SVGdev	
3.5	Descum	O_2 ashing for 1min, 50W	Technics-c	
3.6	Hard bake	120°C, 20min	VWR oven	
4.0	**Nitride etch**			
4.1	Dry etching	150W, Overetch: 10%	Lam1	
4.2	PR stripping	O_2 ashing 7min., 300W, 230 mTorr	Technics-c	
5.0	**LOCOS formation**			
5.1	Cleaning	Pirahna clean 10 min, 25:1 HF 20sec	Sink6	
5.2	Field oxidation	1000°C, 2hrs	Tylan2	T_{ox}=5800 Å
5.3	Oxide wet etching	25:1 HF 45sec.	Sink6	
5.4	Si_3N_4 wet etching	H_3PO_4, 170°C, 2.5hrs (fresh solution)	Sink7	
6.0	**Sacrificial oxide**			
6.1	Cleaning	Pirahna clean 10 min, 10:1 HF 1.5min	Sink6	Check dewet
6.2	Pre-oxidation	900°C, 17min	Tylan1	T_{ox}=500 Å
6.3	Oxide wet etching	10:1 HF 3 min.	Sink6	Check dewet
7.0	**Retrograde well formation**			
7.1	Pre-oxidation	950°C, 45min. O_2, 20min. N_2	Tylan6	T_{ox}= 250 Å
7.2	Nitride deposition	800°C, 75 min	Tylan9	T_{SiN}= 3000 Å
7.3	Lithography	Open up n-well		T_{PR}= 1.1μ
7.4	Si_3N_4 etching	Power: 150W, Timed Overetch: 20s	Lam1	
7.5	N-well I/I	31 P^+, 380 keV, $2x10^{13}cm^{-2}$		R_p = 0.57μm

7.6	P channel implant	3 splits (based upon simulation)		
7.7	PR stripping	O_2 ashing for 6 min, 300W	Technics-c	
7.8	Wet etching	H_3PO_4 acid, 160°C, 5hr	Sink7	Check oxide thickness
7.9	Cleaning	Pirahna clean, no HF dip	Sink8 + Sink6	
7.10	Oxidation	850°C, 45 min O_2, 5s N_2	Sink6, Tylan5	
8.0	**Drive-in of E-beam Alignment Keys**			
8.1	Prebake	Hard bake and HMDS coat	Primeoven	
8.2	PR coating	New OCG I-line	SVGcoat1	T_{PR}=1.1 μm
8.3	Exposure	Use posted exposure time and focus	GCAWS	
8.4	PEB,development	120°C, 60sec, I-line developer	SVGdev	
8.6	Descum	O_2 ashing for 1min, 50W	Technics-c	
8.7	Hard bake	120°C, 20min.	VWR oven	
8.9	Dry etching	Breakthrough: TCP=200W, BP=40W, CF4=100cc, 13mT, 20sec, ME: RF = 300W/150W, Cl_2:HBr=50sccm:150sccm, 15mT, 1 min.	Lam5	
8.10	PR stripping	PRS3000 for 20 min	Sink5, Spindryer3	Inspect PR residue
9.0	**Gate stack formation**			
9.1	Cleaning	Pirahna clean + 10:1 HF 1min.	Sink6	Check dewet
9.2	Oxide growth	750°C, 15min O_2, 900°C, 20 min N_2 anneal	Tylan6	TCA clean before oxdn, 20 Å SiO_2
9.3	SiGe deposition	300mT, 550°C, SiH_4: GeH_4 =200sccm:19sccm	Tylan19	1500 Å (1 min Si nucleation)
9.4	LTO deposition	450°C, 2 min		250 Å, hard mask
10.0	**Gate lithography**			
10.1	Prebake	Hard bake and HMDS coat	SVGcoat	Remove moisture, HMDS
10.2	PR coat + bake	UVN30 + 120°C, 2min.	SVGcoat	T_{PR}=0.34μm
10.3	Exposure	Use calibrated condition	Hitachi ebeam	At Stanford
10.4	PEB	110°C, 2min.	Hot plate	At Stanford
10.5	Development	CD14, 40 sec. + DI water rinse		At Stanford
10.6	Ashing	O_2 ashing for ~3min, 30W	Technics-c	
10.7	Hard bake	120°C, 20 min	Oven-VWR	
11.0	**Gate etching**			
11.1	Dry etching	Breakthrough: RF = 450W/50W, CF_4=50sccm, Main Etch: RF=300W/150W, HBr:Cl_2:He =150:50:4sccm, over-etch: RF=250W/120W, HBr:O_2:He = 200:5:4sccm		Breakthrough also etches through hard mask
11.2	PR stripping Characterize linewidths	(100:1 HF 30s) + (50W 3min O_2 ashing)	Sink7, Technics-c	Inspect PR residue. Characterize L_{phys} extensively
11.3	Cleaning	10 min pirahna clean	Sink8, Sink6	

11.4	HTO deposition	800°C, 10 min., 10 sccm SiH_2Cl_2, 100 sccm N_2O	Tylan9	T_{ox}=50Å, liner for spacer etch stop
12.0	**Halo implantation (optional, performed on select die after drop-out lithography)**			
12.1	Phosphorus II	120keV, $1x10^{15}cm^{-2}$ (or $2x10^{15}cm^{-2}$), 30° tilt		
13.0	**Si_3N_4 spacer formation**			
13.1	Cleaning	Pirahna clean	Sink6	
13.2	Nitride deposition	800°C	Tylan9	T_{SiN}=200Å
13.3	Dry etching	RF=300W/150W, HBr:Cl_2:He =150:50:4 sccm	Lam5	
14.0	**Selective Ge deposition and Hard Mask Formation**			
14.1	Cleaning	Pirahna clean, followed by 25:1 Hf dip until dewet observed on active	Sink6	Remove oxide from active, but keep hard mask over gate
14.2	Ge deposition	600 mT, GeH_4, 219 sccm, 12 min, 340°C	Tylan19	
14.3	LTO deposition	450°C, 2 min	Tylan12	250 Å, mask for II screen
15.0	**P+ Gate & S/D I.I.**			
15.1	Boron II	5 keV, $6x10^{15}cm^{-2}$		
16.0	**Activation anneal**			
16.1	Solid-phase crystallization	300°C, 6 hrs.	Tylan19	
16.2	Gate & S/D annealing	900°C, 7 mins, N_2	Heatpulse3	
17.0	**Passivation layer**			
17.1	LTO deposition	450°C, 30min., no PH_3	Tylan12	t_{ox}=4300Å
18.0	**Contact definition**			
18.1	PR coating	New OCG I-line	Primeoven, Svgcoat2	
18.2	Exposure	Posted exposure time and focus	GCAWS	
18.3	Development,	Standard PEB and develop	SVGdev	
19.0	**Contact etching**			
19.1	Oxide dry etching	RF = 450W/50W, CF_4=50sccm	Lam5	Check oxide thickness in contact hole
19.2	PR strip + clean	PRS3000 spin, Pirahna+10:1 HF 2min	Sink8, Sink6	Inspect PR residue
20.0	**Metal deposition**			
20.1	Cleaning	25:1 HF 15s	Sink6	
20.2	Ti TiN Al sputtering	20mT Ar, 1 kW, Ti, 80cm/min, 1 pass 20mT (Ar:N = 3:1), 1.5kW, Ti, 40cm/min, 1pass 6mT, 4.5 kW, Al, 28.8cm/min, 2 passes	CPA	t_{Ti}=50 Å t_{TiN}=200 Å t_{Al}=5000 Å
21.0	**Metal lithography**			
21.1	PR coating	New OCG I-line	SVGcoat2	T_{PR}=1.1μm
21.2	Exposure	Posted exposure time and focus	GCAWS	

21.3	PEB, Development	120^{o}C, 1 min, I-line developer	SVGdev	
21.4	Hard bake	120^{o}C, 30min.	Oven-VWR	
22.0	**Metal etching and final sintering**			
22.1	Al etching	Wet etch in Al etchant	Sink8	
22.2	TiN and Ti etching	RF=300W/150W, HBr:Cl_2:He =150:50:4sccm	Lam5	
22.3	PR stripping	PRS3000 spin	Spindryer3	
22.4	Cleaning	DI water rinse	Sink8	
22.5	Al sintering	400^{o}C, 1hr	Tylan13	