Low Power Band to Band Tunnel Transistors

Anupama Bowonder

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2010-154
http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-154.html

December 15, 2010

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
15 DEC 2010		**00-00-2010 to 00-00-2010**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Low Power Band to Band Tunnel Transistors	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **University of California at Berkeley,Department of Electrical Engineering and Computer Sciences,Berkeley,CA,94720**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

As scaling continues, the number of transistors per unit area and power density are both on the rise. A reduction in Vdd is highly desirable to reduce power consumption. For MOSFETs however, this would mean scaling the threshold voltage to maintain performance and thereby enhancing the off current and static power consumption since MOSFETs are limited to a swing of 60mV/decade at best. A low voltage transistor that allows Vdd scaling to 0.5V and below is highly desirable. In this thesis, gate induced band-to-band tunneling transistors are explored as a low voltage alternative because of their potential to achieve lower than 60mV/decade turn-off. Since BTBT is strongly dependant on the band gap of the semiconductor, moving from Silicon to Germanium to lower band gap materials can help scale Vdd. Biaxially strained Si1-xGex based heterostructures can provide ultra low effective band gaps. Strain is used to engineer complimentary Si1-xGex heterostructures with low effective band gap for both N and P type transistors. The design and fabrication of heterostructure based tunnel transistors is explored to help scale Vdd to 0.5V and below. Dopant engineering techniques to enhance the electric field are also explored both with simulations and experiments.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**105**	

Low Power Band to Band Tunnel Transistors

By

Anupama Bowonder

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Chenming Hu
Professor Sayeef Salahuddin
Professor John Morris

Fall 2010

The dissertation of Anupama Bowonder, titled Low Power Band to Band Tunnel Transistors, is approved:

Chair _____ Date _____

_____ Date _____

_____ Date _____

University of California, Berkeley

Fall 2010

Abstract

Low Power Band to Band Tunnel Transistors

by

Anupama Bowonder

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Chenming Hu, Chair

As scaling continues, the number of transistors per unit area and power density are both on the rise. A reduction in V_{dd} is highly desirable to reduce power consumption. For MOSFETs however, this would mean scaling the threshold voltage to maintain performance and thereby enhancing the off current and static power consumption since MOSFETs are limited to a swing of 60mV/decade at best. A low voltage transistor that allows V_{dd} scaling to 0.5V and below is highly desirable.

In this thesis, gate induced band-to-band tunneling transistors are explored as a low voltage alternative because of their potential to achieve lower than 60mV/decade turn-off. Since BTBT is strongly dependant on the band gap of the semiconductor, moving from Silicon to Germanium to lower band gap materials can help scale V_{dd}. Biaxially strained $Si_{1-x}Ge_x$ based heterostructures can provide ultra low effective band gaps. Strain is used to engineer complimentary $Si_{1-x}Ge_x$ heterostructures with low effective band gap for both N and P type transistors. The design and fabrication of heterostructure based tunnel transistors is explored to help scale V_{dd} to 0.5V and below. Dopant engineering techniques to enhance the electric field are also explored both with simulations and experiments.

To my most beloved father, for his unselfish love and encouragement

Table of Contents

Acknowledgements

Firstly I would like to express my deepest gratitude to my advisor Professor Chenming Hu for all his guidance, support and encouragement through my entire graduate career at UC Berkeley. I thank him for helping me become a better researcher and for me helping through my interaction with him in group meetings understand how to question and constantly challenge yourself so no detail is missed. I would also like to thank Professor John Morris for serving both on my qualifying exam committee and my dissertation committee. I would like to thank Professor Sayeef Salahuddin for serving on my dissertation committee and for his valuable insights in our weekly group meetings. I would like to thank Professor Tsu Jae King for her invaluable processing advice and insights during our weekly group meetings and for supporting Microlab related requests. I would like to thank Professor Vivek Subramaniam for serving as the chain on my qualifying exam committee.

I would like to express sincere thanks to all the Microlab staff without whom none of my experiments would have been possible. In particular I would like to thank Sia Parsa and Bob Hamilton for always listening to my complaints and trying to help with keeping the lab MOS clean. I would like to thank Joe Donnelly for all his support with Novellus. I will always remember the wonderful conversations with Joe while running Novellus. I would like to thank Evan Stateler for all his invaluable support of Picosun, ASML and Centura. Only Joe and Evan can keep Novellus and ASML running along and still smile when we make annoying requests.

I would like to thank Dr. Prashant Majhi and Dr. H.H. Tseng for giving me the opportunity to work at SEMATECH and make TFETs when this project was still very young.

I would like to thank all the device group members for making device group so incredibly awesome. In particular I would like to thank Pratik Patel and Cheuk Chi Lo. We started graduate school together in the device group, we studied for our prelims together and we have been through the ups and downs of Microlab and graduate school together. I will forever be thankful for their friendship and support through all the failures and successes of graduate school. I would also like to thank Kanghoon Jeon, Chun Wing Yeung and Jack Yaung for their support, help and friendship and for being great group-mates. I would like to thank Zach Jacobson for being a good friend and for always helping to keep things funny and interesting in my life. I would like to thank all the senior students (Joanna Lai, Sriram Balasubramaniam, Donovan Lee, Vidya Varadarajan, Chung Hsun Lin) who helped welcome me and teach me so much about fabrication in the lab.

I would like to thank DARPA STEEP and MARCO MSD for financial support for this project.

I would like to thank my parents for anything and everything that I am today. Without their love, support and encouragement I would not be who I am today. They always pushed me to do better and push the limit and try and achieve even the hardest goal. I would like to thank my father for supporting me to move from home and come to the US to get an education at an early age. Without his support I would never had made it. My dad has taught me the importance of hard work, sincerity, honesty and not taking anything for granted. I will always remember how incredibly proud he was of me getting a PhD from UC Berkeley and even though he is not with me to see me go through this huge step in life, I know I have made him proud. I would like to thank my mom for being there for me through everything in life and for making me a stronger person. Without their infinite love and support I would never have made it this far.

Lastly I would like to thank Pratik for being so incredible! I can't imagine getting through graduate school without his companionship, support, friendship and love.

Chapter 1

Introduction

1.1 Scaling and Power Density

Generations of CMOS technologies have thrived from scaling transistor dimensions. While scaling primarily drives cheaper and denser integrated circuits because of the reduced area, it also drives faster circuits. The increase in circuit density and functionality yields higher computing power at the cost of increased power consumption per chip. As the number of transistors per unit area increases the rising power density leads to severe packaging/thermal management concerns. There is also the issue of increased leakage power and its impact on the battery life of electronic equipment.

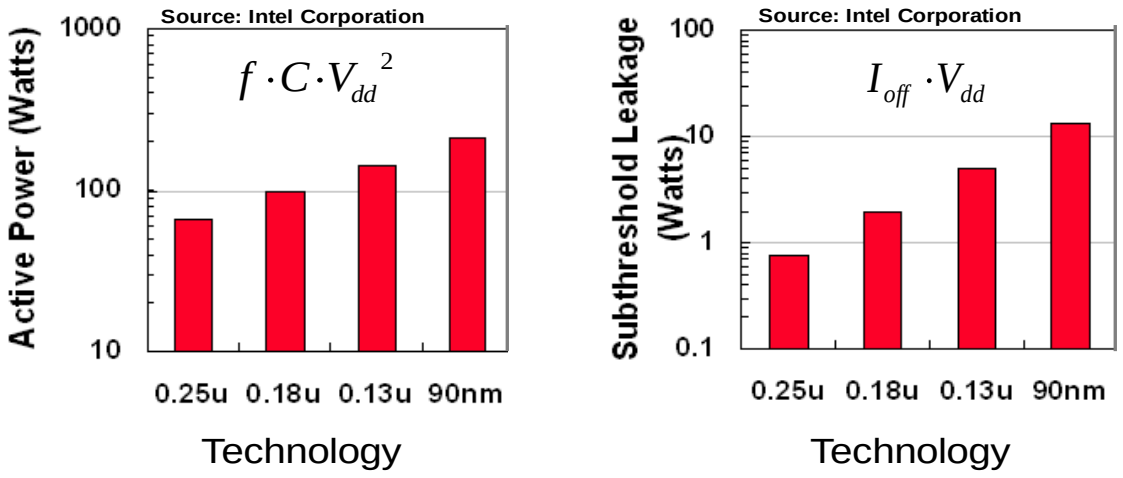


Figure 1.1(a) Active power consumption has been increasing with shrinking technology nodes (b) Standby leakage power also increasing with shrinking technology nodes [1.1]

Figure 1.1(a) and 1.1(b) illustrate the increase in active power consumption and standby leakage (subthreshold leakage) power consumption for various CMOS technology nodes [1.1]. The active and standby power is seen to increase steadily with scaling transistor dimensions. As shown by the equations embedded in the figures both active and standby power scale with the operation voltage (V_{dd}) and can therefore be reduced by scaling V_{dd}. Figure 1.2 shows that V_{dd} scaling has however remained stagnant at ~ 1V for several technology generations now.

1.2 LIMITATIONS OF MOSFET VOLTAGE SCALING

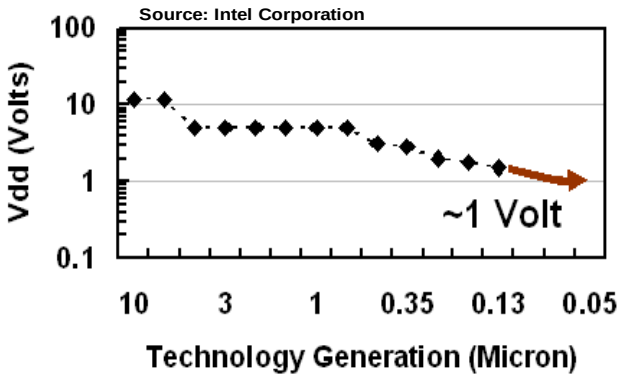


Figure 1.2 V_{dd} scaling has however remained stagnant at ~ 1V for several technology generations now [1.1]

The primary challenge in scaling V_{dd} comes from a basic limitation in the MOSFET operation mechanism. A MOSFET relies on the injection of carriers over a potential barrier which is modulated by the gate. Therefore the MOSFET current is exponentially dependant on the gate voltage as seen in 1.1.

$$I_d \propto n_s \propto e^{(\frac{q\varphi_s}{kT})} \propto e^{(\frac{qV_g}{\eta kT})} \quad 1.1$$

Since $e^{(\frac{qV_g}{\eta kT})}$ increases by 10 for every η x 60mV of gate voltage, the MOSFET turn on (subthreshold swing, S) is limited to 60mV/decade increase in current. To maintain high performance, V_{dd} scaling would require threshold voltage (V_t) scaling as seen in 1.2.

$$I_{on} \propto (V_{dd} - V_t)^{\alpha} \quad 1.2$$

Since the subthreshold swing (S) is 60mV/decade at best, scaling the threshold voltage would lead to an exponential rise in standby leakage current of the transistor.

$$I_{off}(V_g = 0) = I_d(V_g = V_t)e^{-\frac{Vt}{S}} \quad 1.3$$

Figure 1.3 demonstrates this increase in leakage power with scaling technology nodes [1.1].

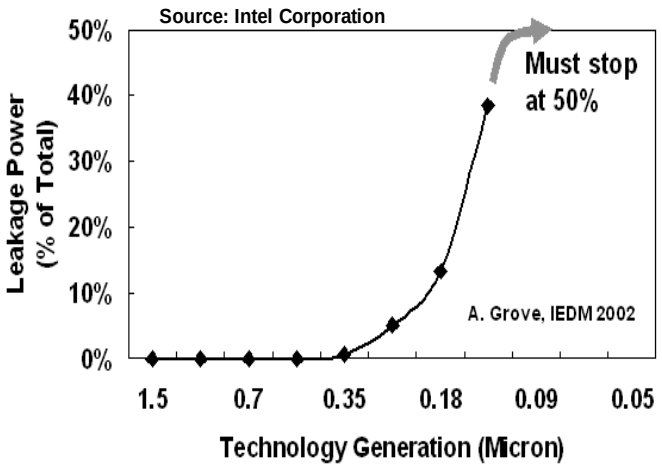


Figure 1.3 Leakage power as a percentage of total power consumption has been on the rise with shrinking technology nodes [1.1].

The most obvious means to enable V_{dd} scaling without a hit in performance would involve transistors with smaller than 60mV/ decade turn off. Transistors which inject carriers through the potential barrier (quantum mechanical tunneling) instead of over the barrier and mechanical switched could overcome the 60mV/decade turn off and allow for low voltage operation.

1.3 RESEARCH OBJECTIVES AND DISSERTATION OUTLINE

This dissertation deals with the exploration and design of novel tunneling based transistors (TFET) [1.2-1.7] and the pursuit of the most optimal TFET structure with steepest turn on characteristic and highest I_{on}/ I_{off} at a low operating voltage.

Chapter 2 deals with the structure and operation of the simplest TFET design – a gated PN diode. Through the basics of tunneling physics the key parameters that affect TFET performance are determined – electric field and effective bandgap. This chapter also deals with experimental work on bulk silicon (Si) TFETs with high-k and metal gate and the process development required to tune an existing CMOS baseline process at SEMATECH to fabricate TFETs. Electrical characterization and detailed analysis of the devices is also presented. Finally experimental work on ultra thin body silicon TFETs with raised germanium (Ge) source and drain is also presented.

Chapter 3 deals with the use of dopant engineering to enhance the electric field [1.8,1.9] and therefore TFET performance enabling V_{dd} scaling to below 0.5V. Dopant pockets are thin regions of dopant opposite in type to the source dopant that can be incorporated in addition to low bandgap materials for even higher performance at ultra low voltages. Simulations are used to understand the operation and optimize a silicon pocket TFET design. Experimental data which confirms the superior behavior of the pocket TFET over a simple PIN TFET is presented. Finally the formation of accidental MOSFETs when creating pocket TFETs is also discussed.

Chapter 4 deals with the use of biaxial strain engineering to obtain the lowest possible effective bandgap for tunneling and therefore enhance TFET performance while still

maintaining a high I_{on}/I_{off}. Since TFET performance is limited by the BTBT generation rate which is exponentially dependant on the effective tunneling bandgap of the semiconductor, the obvious means to enhance TFET performance and scale V_{dd} is by scaling this effective tunneling bandgap. While this can be achieved by moving from silicon to ultra low bandgap III-V materials, these materials with issues like poor dielectric interface quality and low density of states[1.10]. Further homo junction TFETs in these ultra low bandgap materials exhibit degraded I_{on}/I_{off} despite the I_{on} enhancement because of enhanced I_{off}. Simulations are used to explore biaxial strain engineering of silicon and germanium based heterostructures to achieve ultra low effective tunneling bandgap without a severe I_{off} penalty. Further strain engineering is also explored to achieve symmetric N and P hetero TFETs which can enable V_{dd} scaling to below 0.4V.

Biaxial strain has been researched greatly for enhanced mobility and performance in MOSFETs [1.11-1.13]. Chapter 5 explores electrical measurements of gate induced band to band tunneling (GIDL) from biaxial strained heterostructure MOSFETs as preliminary verification of enhanced tunneling across heterostructures with reduced effective bandgap. The temperature dependence of the GIDL current and impact of mechanical strain on the GIDL current from such biaxial strained heterostructures is also presented.

As seen in chapter 5 when working with relaxed germanium layers on a silicon substrate to create heterostructure TFETs, the junction leakage is very high and overshadows the tunneling behavior close to turn on. This high leakage when using relaxed epitaxial layers is largely attributed to the large dislocation density caused during strain relaxation. Bulk germanium substrates could however help solve this problem as long as implant damage is annealed sufficiently. Chapter 6 explores the use of bulk germanium substrates to fabricate homojunction and biaxial strain heterostructure TFETs. Several essential modules required to fabricate hetero TFETs on germanium substrates are implemented and explored in detail. Chapter 7 provides an overall conclusion as well as possibilities for future work.

1.4 REFERENCES

[1.1] S. Borkar, “Getting gigascale chips: Challenges and opportunities in continuing Moore’s law”, ACM Queue, 1(7):26–33, Oct. 2003.

[1.2] T. Y. Chan, J. Chen, P. Ko, C. Hu, “The impact of gate-induced drain leakage current on MOSFET scaling”, International Electron Devices Meeting, Vol.33, pp: 718-721, 1987

[1.3] W. M. Reddick, G. A. Amaratunga, "Silicon surface tunnel transistor," Applied Physics Letters, vol.67, no.4, pp.494-496, Jul 1995

[1.4] C. Aydin, A. Zaslavsky, S. Luryi, S. Cristoloveanu, D. Mariolle, D. Fraboulet, S. Deleonibus, “Lateral interband tunneling transistor in silicon-on-insulator," Applied Physics Letters , vol.84, no.10, pp.1780-1782, Mar 2004

[1.5] K. K. Buwalka, M. Born, M. Schindler, M. Schmidt, T. Sulima and I. Eisele, “ P-Channel Tunnel Field-Effect Transistors down to Sub-50nm Channel Lengths”, Japanese Journal of Applied Physics, 45, pp. 3106 -3109, 2006

[1.6] Chenming Hu, "Green transistor as a solution to the IC power crisis," Solid-State and Integrated-Circuit Technology, 2008, ICSICT 2008. 9th International Conference on, pp.16-20, 2008
[1.7] A. Bowonder, P. Patel, K. Jeon, J. Oh, P. Majhi, H. H. Tseng, C. Hu; , "Low-voltage green transistor using ultra shallow junction and hetero-tunneling," Junction Technology, 2008. IWJT '08, pp.93-96, 15-16 May 2008
[1.8] V. Nagavarapu, R. Jhaveri, J.C.S. Woo, "The Tunnel Source (PNPN) n-MOSFET: A Novel High Performance Transistor," Electron Devices, IEEE Transactions on, vol.55, no.4, pp.1013-1019, April 2008
[1.9] P, Patel, K. Jeon, A. Bowonder, C. Hu, "A Low Voltage Steep Turn-Off Tunnel Transistor Design", Simulation of Semiconductor Processes and Devices, 2009, SISPAD '09. International Conference on, vol., no., pp.1-4, 9-11 Sept. 2009.
[1.10] J. A. del Alamo, D. H. Kim, "The prospects for 10 nm III-V CMOS," VLSI Technology Systems and Applications (VLSI-TSA), 2010 International Symposium on, pp.166-167, 26-28 April 2010
[1.11] K. Uchida, R. Zednik, C. H. Lu, H. Jagannathan, J. McVittie, P. C. McIntyre, Y. Nishi, "Experimental study of biaxial and uniaxial strain effects on carrier mobility in bulk and ultrathin-body SOI MOSFETs", Electron Devices Meeting, 2004. IEDM Technical Digest. IEEE International, pp. 229- 232, 13-15 Dec. 2004
[1.12] M. L. Lee, E. A. Fitzgerald, M. T. Bulsara, M. T. Currie and A. Lochtefeld, "Strained Si, SiGe, and Ge channels for high-mobility metal-oxide-semiconductor field-effect transistors", J. Appl. Phys. **97**, pp.011101-011101-28, Jan 2005
[1.13] J. Welser, J. L. Hoyt, S. Takagi, J. F. Gibbons, "Strain dependence of the performance enhancement in strained-Si n-MOSFETs," Electron Devices Meeting, 1994. IEDM '94. Technical Digest, International, pp.373-376, 11-14 Dec 1994

Chapter 2

Homojunction PIN Tunnel FETs

2.1 Structure and Operation of BTBT Based FETs

A low voltage transistor that allows V_{dd} scaling to below 0.5V is highly desirable for reducing the power consumption of future ICs. Gated PN diode transistors based on band-to-band tunneling (BTBT) [2.1-2.8], are not limited by 60 mV/decade turn-off, and are being heavily researched as a low voltage CMOS alternative.

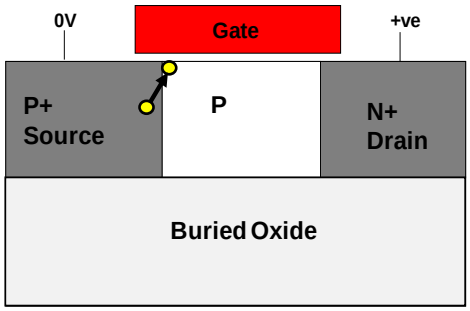


Figure 2.1 Illustration of the simplest PIN N-type Tunnel FET. The TFET is asymmetric and is a three terminal device. In the on-state the drain and gate are biased positive.

Figure 2.1 is an illustration of the simplest N-type BTBT based transistor. It is a three terminal device with a P+ doped source and N+ doped drain. A positive gate bias bends the bands down in the P+ source. With sufficient band bending, overlap occurs between filled valence band states and empty conduction band states. BTBT occurs in the source near the surface generating electron and holes. A positive V_{ds} reverse biases the source-drain junction and the drain collects the BTBT generated electrons. The arrow in Figure 2.1 indicates the path along which BTBT occurs generating electrons and holes at the ends of the tunnel path.

When the valence band and conduction band extrema of a material are located at the same point in k space, direct BTBT occurs between the two extrema. BTBT can be thought of as an electron penetrating the forbidden gap along the imaginary k axis and making a smooth transition from one band into the other band [2.9]. While in classical mechanics an electron of energy E cannot enter into a region with a potential energy U > E because it would have to possess a negative kinetic energy, in quantum mechanics the negative kinetic energy represents a wave with attenuating amplitude [2.10]. If for example the wave function is

$$\psi = Ae^{(\pm ikx)} \qquad 2.1$$

k becomes imaginary or $ik = \alpha$ being real in the potential barrier region. Thus in quantum mechanics even if an electron has E < U the wave function of the electron attenuates as it propagates but is not zero within the barrier and the electron has a finite probability of getting through the barrier region. The transmission probability through the barrier depends on both the width of the barrier and the height of the barrier.

2.2 WKB Approximation and Tunneling Probability

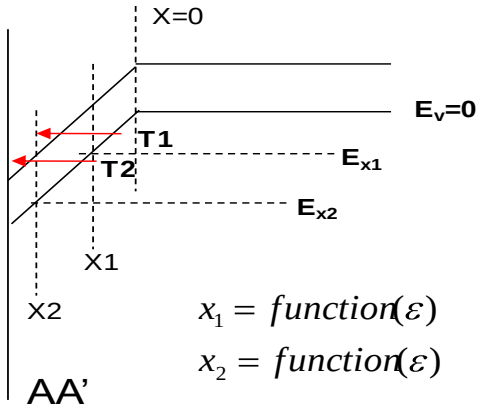


Figure 2.2 Simplified illustration of a band diagram along a cutline in the P+ source when a positive gate voltage causes the bands to bend down. Several different tunneling paths exist with overlap of valence and conduction bands. x1 and x2 are starting and ending points of a tunneling path for a given energy and vary with the specific path.

Figure 2.2 is a simplistic illustration of the band diagram along a cutline in the P+ source bending down with positive gate voltage. The scenario presented shows sufficient overlap of valence and conduction bands and two tunneling paths at two different energies (E_{x1} and E_{x2}) are illustrated. x_1 and x_2 represent the starting and ending points of a tunneling path for a given energy and therefore vary with the specific tunneling path. In order to calculate the number of carriers generated by BTBT along the various tunneling paths, the tunneling probability along each specific tunneling path needs to be computed. The following section explains the details of such a calculation. If the potential varies slowly with distance the WKB (Wentzel-Kramers-Brillouin) approximation can be used to find the solution of the electron wave function [2.10]. When making the WKB approximation, if the change in potential energy over the decay length is smaller than the magnitude of the kinetic energy (E – U) then the transmission probability or tunneling probability through the barrier region

$$T = \exp(-2i \int_{x1}^{x2} K dx) \tag{2.2}$$

k being the wave vector or dispersion relation within the band gap and x1 and x2 are starting and ending points of the tunneling path.

The dispersion relation within the bandgap can be calculated in many different ways. The simplest dispersion relation, the parabolic one band relation [2.11]

$$k = \frac{i\sqrt{2m_r}}{hbar} \cdot \sqrt{Eg - qFx} \tag{2.3}$$

$$\frac{1}{m_r} = \frac{1}{m_c} + \frac{1}{m_v} \tag{2.4}$$

is obtained by conservation of energy of the carrier tunneling from the valence band to the conduction band. Momentum conservation is implicit in this assumption.

$$\mathrm{E}_v - \frac{hbar^2 \cdot k^2}{2m_v} + qFx = \mathrm{E}_c + \frac{hbar^2 \cdot k^2}{2m_c} \tag{2.5}$$

It is well known that this dispersion relation is most appropriate when the carrier is close to the conduction band edge or valence band edge. For carriers tunneling from the valence band to the conduction band, it is more accurate to use a dispersion relation like *Franz' Symmetric Two Band Relation.* This is a theoretically proven dispersion relation [2.11] which, is applicable to simple two band semiconductors and can take into account unequal effective masses near the conduction and valence band edges.

$$K = \frac{K_c K_v}{\sqrt{K_c^2 + K_v^2}} \tag{2.6}$$

$$K_c(x,\varepsilon) = \frac{\sqrt{2.m_c(x).|E_c(x) - \varepsilon|}}{h} \tag{2.7}$$

$$K_v(x,\varepsilon) = \frac{\sqrt{2.m_v(x).|\varepsilon - E_v(x)|}}{h} \tag{2.8}$$

x is the position and ε is the energy for a specific tunneling path.

Figure 2.3 is a comparison between single and two band dispersion relations illustrating why the two band relation is more accurate when considering tunneling within the bandgap of the semiconductor from the valence band to the conduction band.

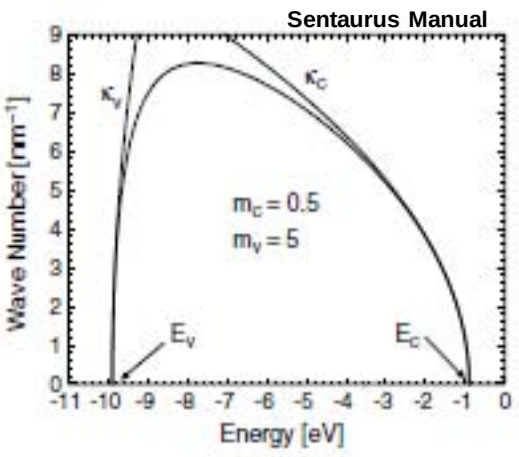


Figure 2.3 A comparison of single and two band dispersion relations illustrating that the two band relation is more accurate when considering tunneling within the bandgap of the semiconductor from the valence band to the conduction band.

Assuming continuity in all 3 directions the tunneling current is a product of 3d density of states and the velocity in the tunneling direction weighted by the transmission probability in the tunneling direction [2.10].

$$J = q\iiint \frac{2}{h^3} T \cdot v_x \cdot (f_v - f_c) \cdot dp_x dp_y dp_z \qquad 2.9$$

Assuming that the valence band is fully occupied by carriers and the conduction band states are fully unoccupied and available to receive states, $f_v - f_c = 1$.

$$J = q\iiint \frac{2}{h^3} T \cdot v_x \cdot dp_x dp_y dp_z \qquad 3.0$$

If

$$E_x = \frac{p_x^2}{2m_r};\;\; E_{perpen} = \frac{p_y^2 + p_z^2}{2m_v} = \frac{p_{perpen}^2}{2m_v} \qquad 3.1$$

We can do a transformation from momentum space to energy space as shown below

$$dp_y dp_z = 2 \cdot \pi \cdot p_{perpen} \cdot dp_{perpen} \qquad 3.2$$

$$2 \cdot \pi \cdot m_v dE_{perpen} = 2 \cdot \pi \cdot p_{perpen} \cdot dp_{perpen} \qquad 3.3$$

$$dE_x = v_x \cdot dp \tag{3.4}$$

and the current can be expressed as

$$J_{BTBT} = q \iint \frac{2}{h^3}.2.\pi.m_r.T.dE_x dE_{perpen} \tag{3.5}$$

Both E_{perpen} and E_x cannot exceed the total overlap energy (overlap of empty conduction band states with filled valence band states). From this the BTBT generation rate per unit volume can be computed as shown below.

$$G_{rate} = \int_0^{E_{max}} q.\frac{4}{h^3}.\pi.m_r.T.\frac{dE}{dx}dE_{perpen} \tag{3.6}$$

This generation rate can be simplified by assuming a constant E field to give a relationship which clearly details the parameters that affect it

$$G_{BTBT} = A_{BTBT}\frac{E^2}{\sqrt{E_g}}\exp(-B_{BTBT}\frac{E_g^{3/2}}{E}) \tag{3.7}$$

$$A_{BTBT} = \frac{\pi.m_r^{1/2}.q^2}{9h^2} \qquad B_{BTBT} = \frac{\pi^2 m_r^{1/2}}{qh} \tag{3.8}$$

From (3.6) the tunneling design parameters that can be tuned for better tunneling performance are the electric field and the bandgap of the semiconductor. Lower bandgap and larger electric field are desired for enhanced generation rate. The electric field in the source can be enhanced by doping the source heavier

$$E_{Si} = -\frac{q \cdot Na \cdot W_{dep}}{C_{ox}} \tag{3.9}$$

Thin heavily doped pockets doped opposite to the source and either on top of the source or adjacent to the source can also be used to the enhance the electric field further. Since BTBT is strongly dependant on the bandgap (E_g) of the semiconductor, moving from silicon (E_g=1.12eV) to germanium (E_g=0.67eV) to InAs (E_g=0.36eV) and further to heterostructures [2.6] of even lower effective E_g can help lower V_{dd} to below 0.5V [2.8, 2.12].

2.3 TCAD SIMULATIONS OF TFETS

TCAD simulators such as Synopsis' Sentaurus are equipped to self consistently compute

BTBT current in TFETs. A specified TFET structure is overlapped with a meshing specification so as to create numerous grid points within the entire structure. When calculating BTBT current, the simulator dynamically searches for overlap between valence and conduction bands and then computes the generation rate along various tunneling paths.

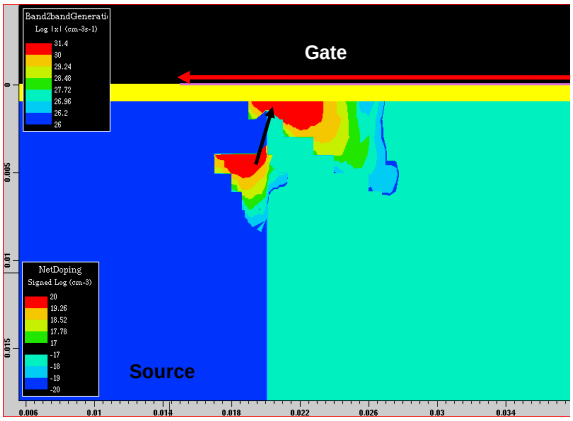


Figure 2.4 TCAD simulation output illustrating contours of BTBT generated holes and electrons. The arrow indicates the direction of tunneling which is not entirely vertical or lateral because both a vertical and lateral field exists. Tunneling originates a few nanometers below the oxide interface entirely within the source of the TFET.

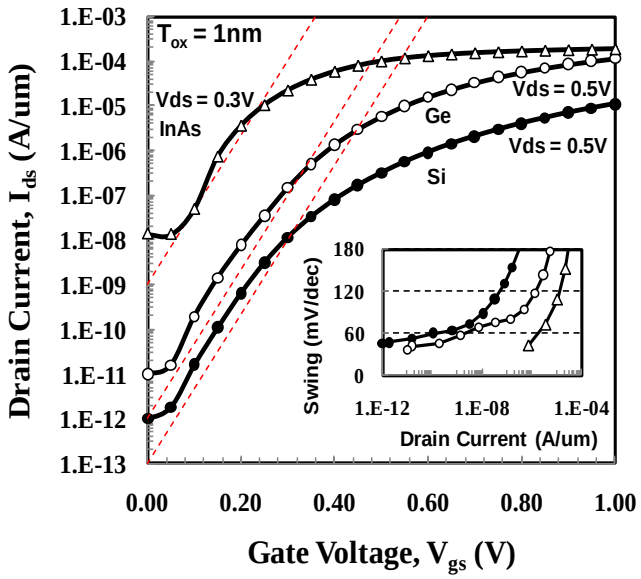


Figure 2.5 Simulations of the I_d-V_g characteristics of Si, Ge and InAs homojunction TFETs. The I_{on} increases with lower E_g, but so does the I_{off} (junction leakage).

Figure 2.4 is a TCAD simulation output illustrating contours of BTBT generated holes and electrons. The arrow indicates the direction of tunneling which is not entirely vertical or lateral

because both a vertical and lateral field exists. The figure also demonstrates that tunneling originates a few nanometers below the oxide interface entirely within the source of the TFET.

Figure 2.5 summarizes TCAD simulations comparing the I_d-V_g characteristics of silicon, germanium and InAs homojunction TFETs. The I_{on} increases with lower bandgap because of the increased generation rate but so does the I_{off} (junction leakage). The I_{on}/I_{off} for a purely InAs TFET is worse than that of a purely silicon or germanium TFET. The dotted lines indicate 60mV/decade. Ideally TFET I_{off} is entirely due to reverse bias junction leakage. For materials like silicon and germanium TFET I_{off} is in the 1-10pA current range making these TFETs ideal from a low standby power perspective. Also since the generation rate is exponentially dependant on the electric field, the instantaneous swing of the TFETs is constantly varying and is less than 60mV/decade only over a certain current range. The challenge is in keeping the low I_{off} and improving the TFET design to increase the I_{on}/I_{off} at low voltages. Future chapters explore dopant pocket engineering as well as biaxial strained silicon and germanium heterostructures to improve on the simplest homojunction TFET design and enable V_{dd} scaling to below 0.5V.

2.4 High-K Metal Gate Silicon TFETs

Using an existing High-K metal gate CMOS baseline process flow to fabricate silicon TFETs required the development of a half mask which would align to the gate. This half mask would help with creating asymmetrically doped source and drain regions.

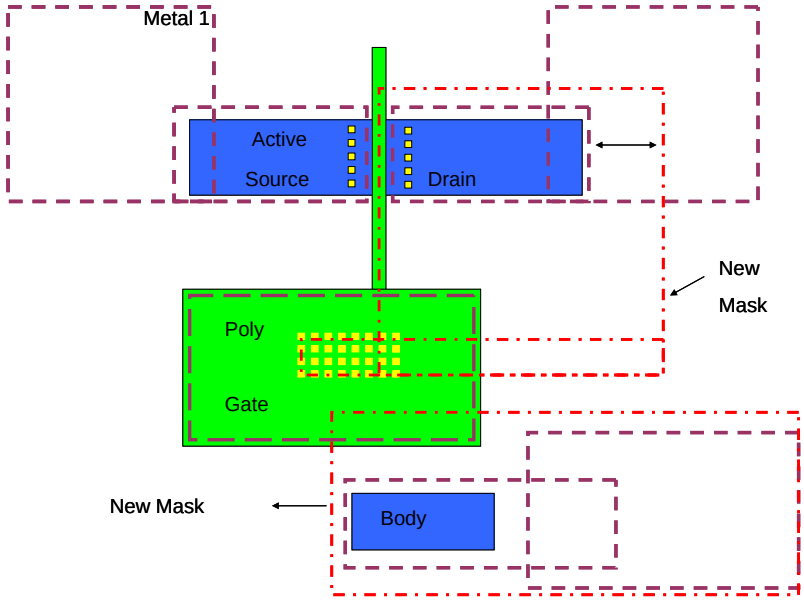


Figure 2.6 Illustration of the new half mask layout designed to be compatible with the existing CMOS mask set at SEMATECH.

Figure 2.6 is an illustration of the new half mask designed to be compatible with the existing CMOS mask set. This mask was designed to cover the drain while creating the source. Since a reverse half mask was not initially created to minimize cost, the process flow involved a drain implant high enough to create good ohmic contact but low enough to keep compensation of the source dopants minimal. Aligning the half mask to sub 100nm gate lengths required the use of a 193nm DUV stepper with an overlay tolerance specification of +/- 40nm. Figure 2.7 is an SEM image of the half mask aligned to a 50nm gate.

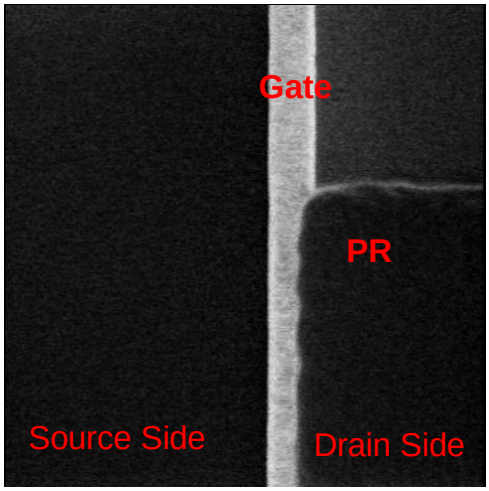
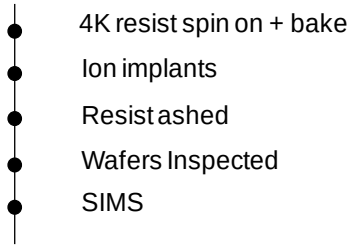


Figure 2.7 Top down SEM image of the half mask aligned to a 50nm gate

The 4000Å thick DUV resist had previously only been used to resolve gates. Using this photoresist to effectively block ion implanted source side dopants from entering the drain side had to be tested. It was also required to test if this resist could after the ion implantation be successfully ashed and cleaned off the wafer. These concerns were tested by a quick experiment shown below in Figure 2.8.



Wafer Split	Species	Implant Dose	Implant Energy
1	As	4e15	35KeV
2	B	8e15	10KeV
3	As	4e15	35KeV
		2e15	10KeV

Figure 2.8 Process steps to determine if the resist effectively blocks ion implants. The table shows the various implants tested with the resist.

Figure 2.9 shows the concentration of Arsenic and Boron in the silicon substrate from SIMS analysis. The large spike in concentration seen within the first few nanometers at wafer surface is due to an artifact and is well known to be present in all SIMS analayis data. The concentration

deeper in the substrate is negligible indicating that the photoresist does act effectively to block implants from entering the silicon substrate.

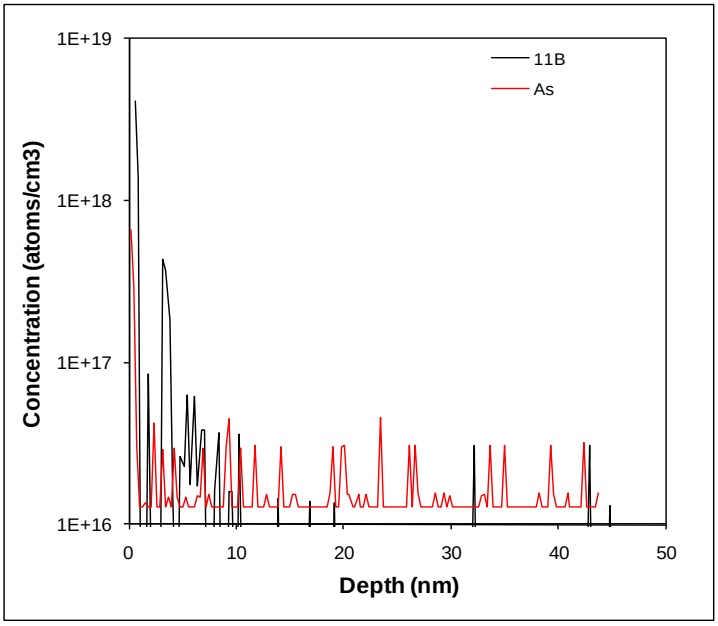
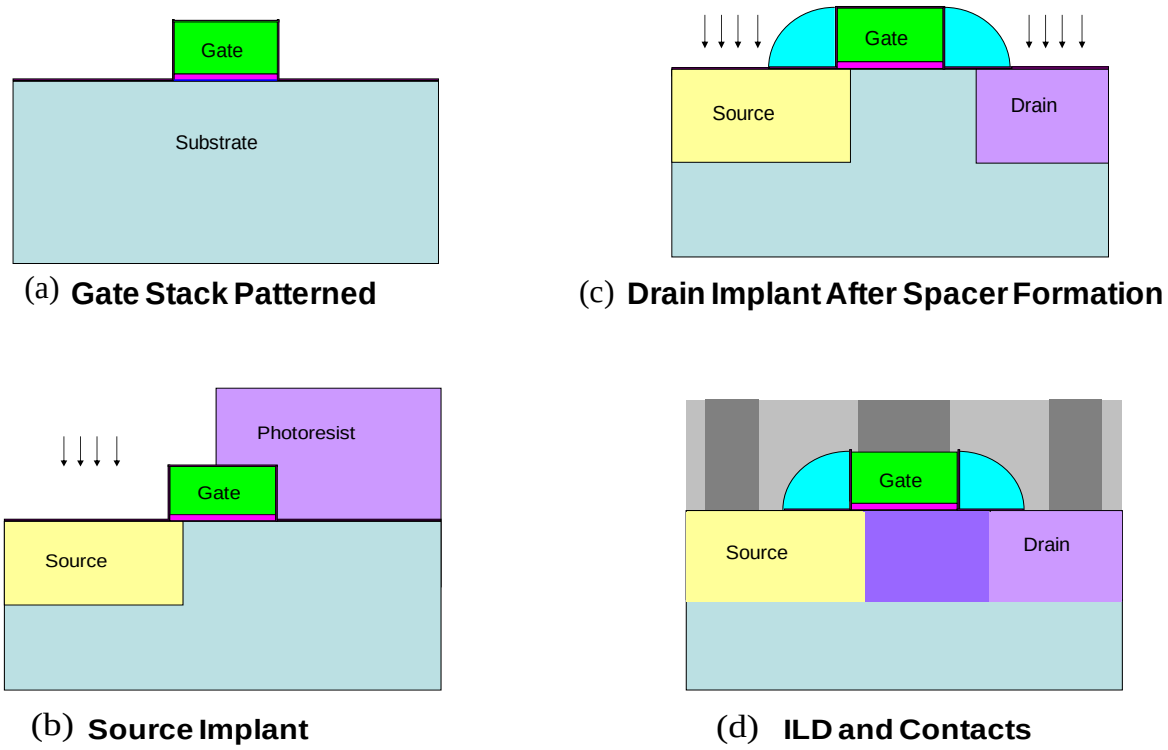


Figure 2.9 Concentration of Arsenic and Boron in the silicon substrate from SIMS analysis. The concentration in the substrate is negligible indicating that the photoresist effectively blocks implants from entering the silicon substrate. The large spike in concentration within the first few nanometers of the wafer surface is an artifact well known to be present in all SIMS analysis data.

The gate first process flow illustrated in Figure 2.10 involved using the basic MOSFET baseline. The $HfSiO_x$ high- K dielectric was deposited using ALD at 300 $^{\circ}$C and was then subjected to a nitridation anneal to incorporate N and form HfSiON. HfSiON has a higher crystallization temperature [2.13] which is desired to keep the bulk gate leakage low even after a high temperature spike anneal required to activate dopants and remove ion implant induced damage. This is followed by deposition of a 200Å thick TiN layer, a 1000Å thick polysilicon layer and a 600Å thick oxide layer. The oxide layer acts as a hard mask for the gate during gate etching. After the gate stack is patterned and etched, the residual high K dielectric is etched in a buffered HF solution. This is followed by the deposition of a 5nm SiN layer. The SiN is a seal layer which protects the high K dielectric at the gate edges from implant damage. The half mask is then used to protect the drain side with 4000Å of photoresist while the source side is left exposed. For the p-type TFETs a low energy high dose Arsenic implant is carried out to ensure that the source is heavily doped close to the dielectric interface. A 70nm nitride spacer is created and the drain implant is performed with the source exposed. The Bf_2 implant dose is chosen to be high enough to form ohmic contact to the drain with minimal compensation to the source. The activation anneal is performed at 1070 $^{\circ}$C for 3s. This is followed by TEOS deposition, contact formation and forming gas anneal.

Figure 2.10 Illustration of a gate first process flow to create silicon homojunction TFETs.



N-type wafer splits were also processed and a Lanthanum capping layer [2.14] was deposited above the $HfSiO_x$ and before the TiN to achieve N+ band edge work function. Inclusion of the cap however posed a severe problem during the gate etch. As seen from top down SEM shown in Figure 2.11, the standard gate etch cleared the polysilicon and TiN but left La particles close to the gate edges.

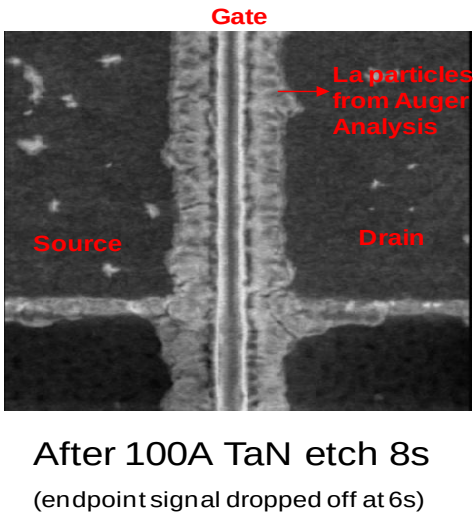


Figure 2.11 Top down SEM after the gate etch of N-type wafer splits. The Polysilicon and TiN were cleared but La particles were left behind close to the gate edges.

The wafers were soaked in SC1 solution for 5minutes to try and remove the particles since SC1 is used to etch metals gates [2.15]. As seen in Figure 2.12, the SC1 etch was not sufficient. The wafers were then soaked in SC2 solution for 3 minutes and this was found to remove the residual particles as shown in Figure 2.12.

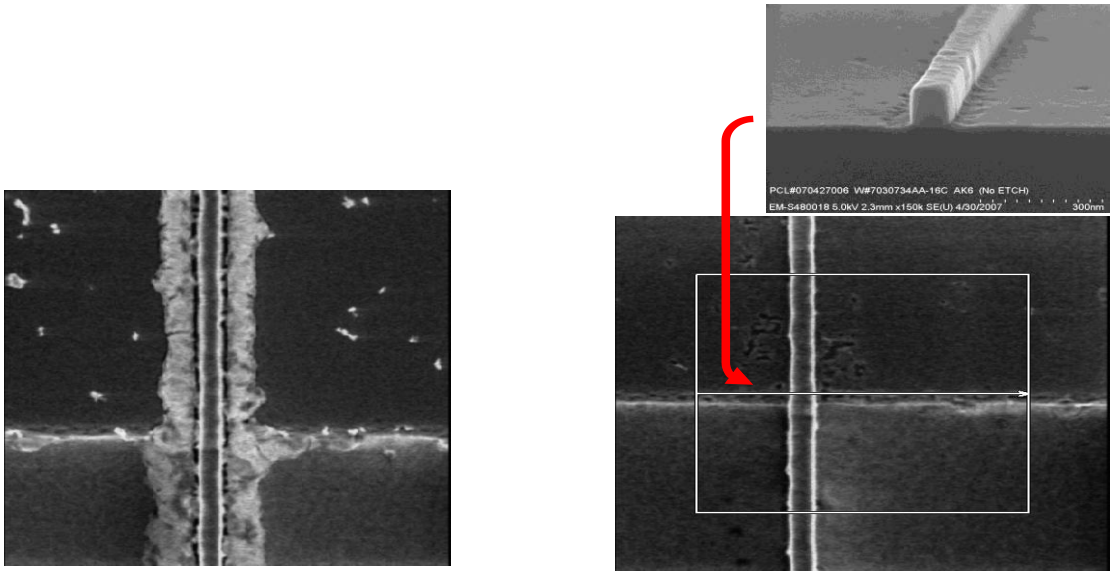


Figure 2.12 Top down SEM image after a 5 minute soak in SC1. This etch was not sufficient. The wafers were then soaked in SC2 solution for 3 minutes. This was found to remove the residual particles as shown in the top down SEM image on the right.

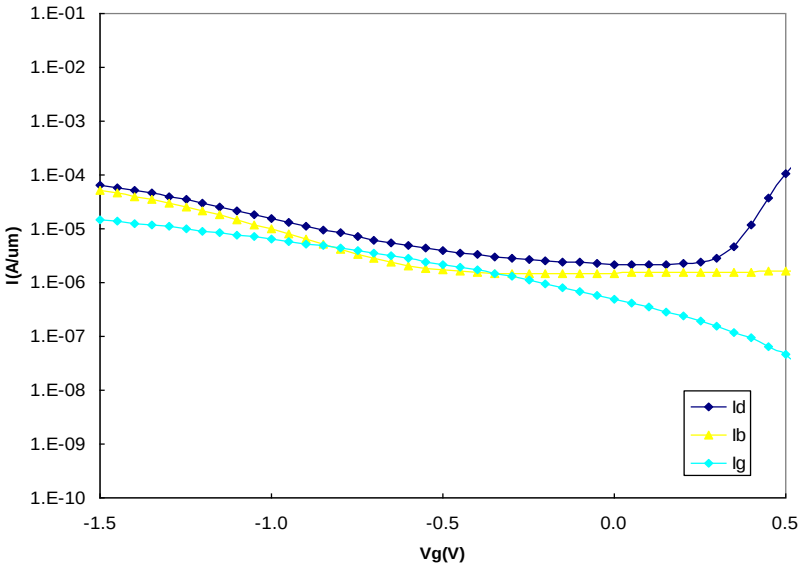


Figure 2.13 The P-type transistor I_d-V_g shows evidence of BTBT current but the very high junction leakage and gate leakage mask the true nature of tunneling characteristics.

The N-type transistor splits when measured exhibited very high gate leakage current which masked any possible tunneling characteristics. As seen in Figure 2.13, the P-type transistors showed evidence of BTBT current but the very high junction leakage and gate leakage masked the true nature of tunneling characteristics. The sign of the current flowing out of the source, drain and gate is carefully noted and found to indicate that the drain collects holes while the source collects electrons as expected from normal P-type TFET operation.

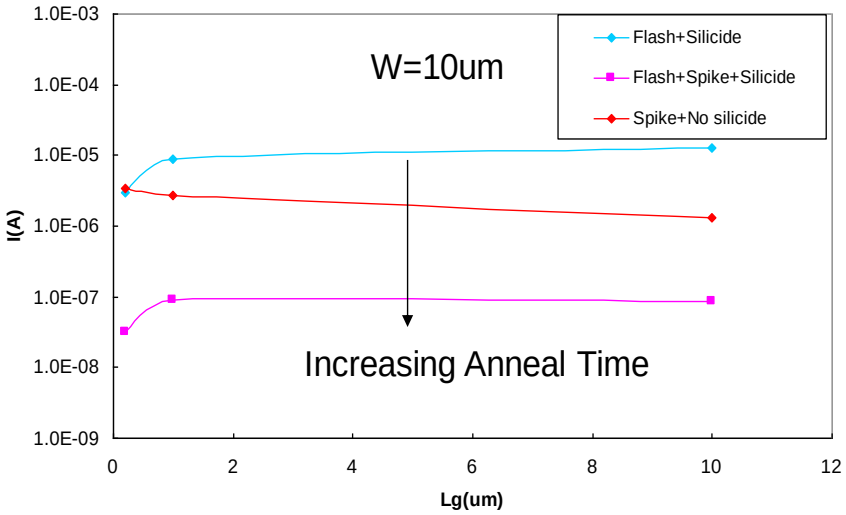


Figure 2.14 TFET I_{off} plotted vs. transistor gate length for three different TFET splits. The splits involved different dopant anneals – flash annealing, flash + spike annealing and spike annealing.

The TFET I_{off} is plotted for various gate length transistors and is found to be fairly independent of gate length. Three different TFET splits are presented in Figure 2.14. The TFET I_{off} is found to be lowest for the split which received both spike and flash anneal. This seems to indicate that the I_{off} is due to insufficient damage annealing. The TFET split with flash anneal is found to have larger I_{off} than the split with spike anneal, again indicating that the high I_{off} is due to un-annealed damage.

Figure 2.15 combines TCAD simulations of a P-type TFET with measured gate leakage current from a capacitor and a transistor. For the same area, the transistor gate current is much higher than the capacitor gate current. From Figure 2.16 it is clear that the gate leakage current density increases with decreasing gate length indicating that the dominant transistor gate leakage current is from the gate edges rather than the bulk.

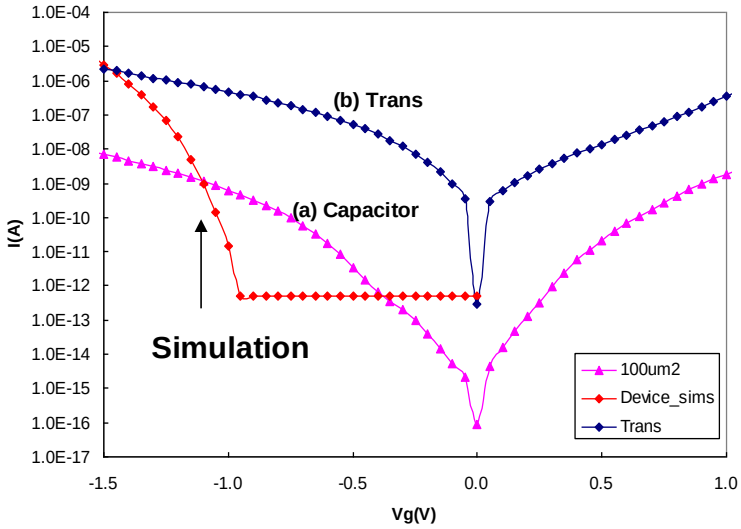


Figure 2.15 TCAD simulations of a P-type TFET overlaid with measured gate leakage current from a capacitor and a TFET. For the same area, the transistor gate current is much higher than the capacitor gate current.

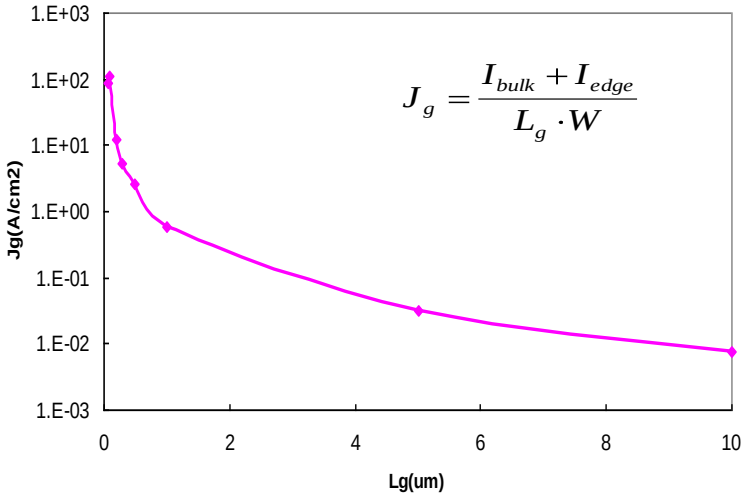


Figure 2.16 Gate current density increases with decreasing gate length. This indicates that the dominant transistor gate leakage current is from the gate edges rather than the bulk.

Figure 2.17 indicates that the gate leakage current density increases with decreasing gate length regardless of the activation anneal mechanism.

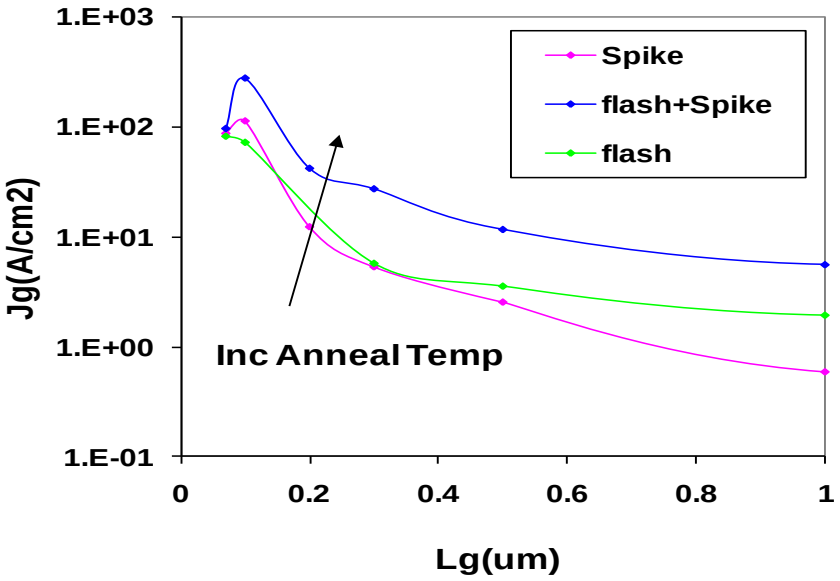


Figure 2.17 The gate leakage current density vs. transistor gate length for varying anneal mechanisms - spike, flash + spike and flash

SEM images (Figure 2.18) of the drain side and source side contact indicate that the contact etch etched into the silicon substrate to a deeper extent in the source than in the drain. The etch rate of heavily N-doped silicon is known to be higher and could explain this.

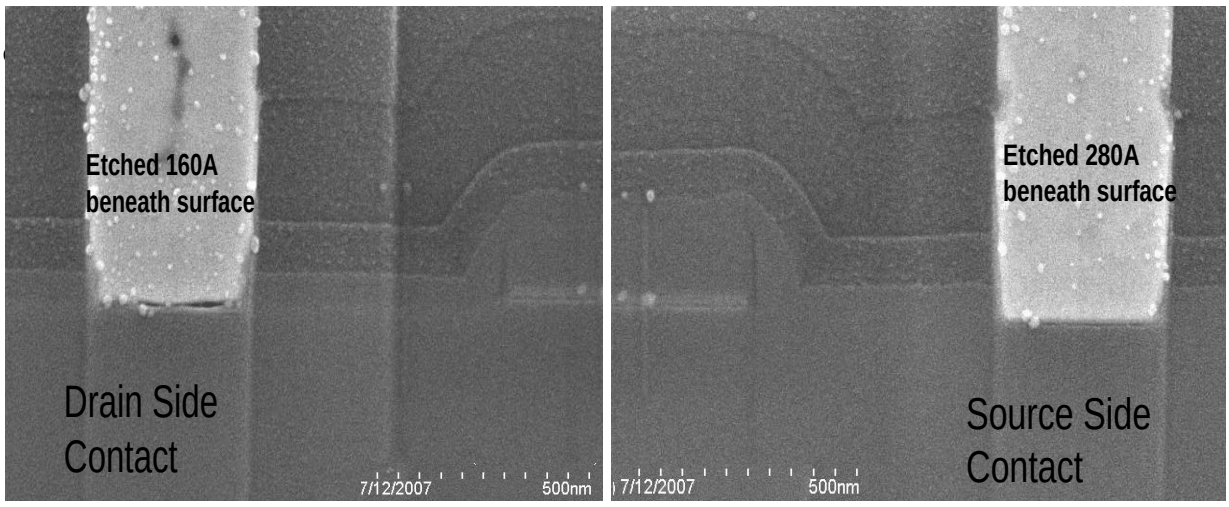


Figure 2.18 SEM images of the drain side and source side contact

In conclusion from the electrical characterization it is clear that in order to see the true tunneling characteristics at the onset of tunneling, the gate leakage current and junction leakage current need to be minimized. The high temperature anneal needs to be optimized for low

junction leakage and the gate etch need needs to be optimized to reduce edge related gate leakage.

2.5 ULTRA THIN BODY SILICON TFETS

Ultra thin body MOSFETs [2.16, 2.17, 2.18] have been successfully fabricated at UC Berkeley in the past. UTB MOSFETs help achieve better short channel control by eliminating the silicon in the channel which is least effectively modulated by the gate. In UTB MOSFETs the aim is for the potential in the channel to be controlled entirely by the gate rather than by the drain. In TFETs the benefit of the UTB comes from the body being thin enough such that when overlap of conduction and valence bands first occurs, the electric field is large enough that there is a sudden jump in transistor current from the junction leakage (I_{off}) to a high tunneling current value. This sudden jump or steep turn on characteristic would occur at a larger voltage than in a thicker body TFET.

UTB MOSFETs do however suffer from large source and drain series resistance because of the thin body. A simple solution to this is using elevated S/D processes. A selectively deposited raised germanium S/D process was used successfully for UTB MOSFET fabrication at UC Berkeley [2.16]. This previously tested process was used to try and fabricate UTB silicon TFETs with raised germanium S/D.

The key steps of the process flow are as shown below

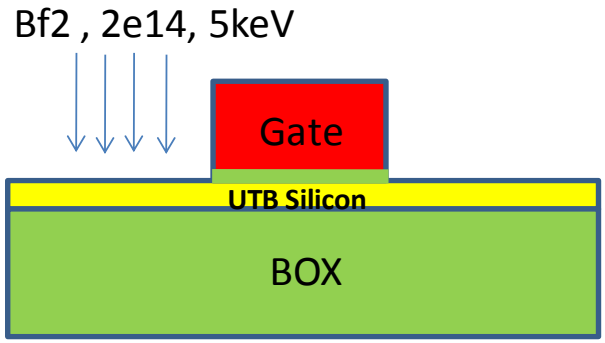


Figure 2.19 (a) Source side dopant implant into UTB silicon after etching the gate stack

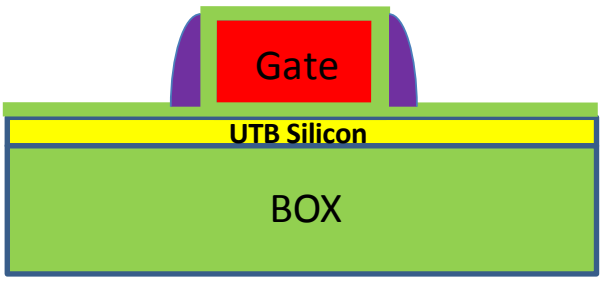


Figure 2.19 (b) Bi-layer spacer formation with nitride on top of HTO.

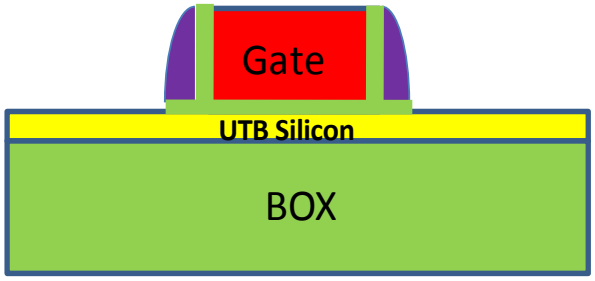


Figure 2.19 (c) Removal of HTO over source and drain region by wet etch

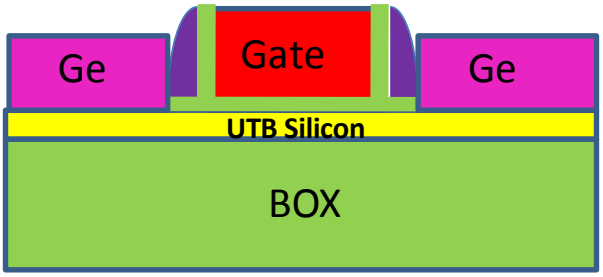


Figure 2.19 (d) Selective deposition of LPCVD Poly germanium on the source and drain

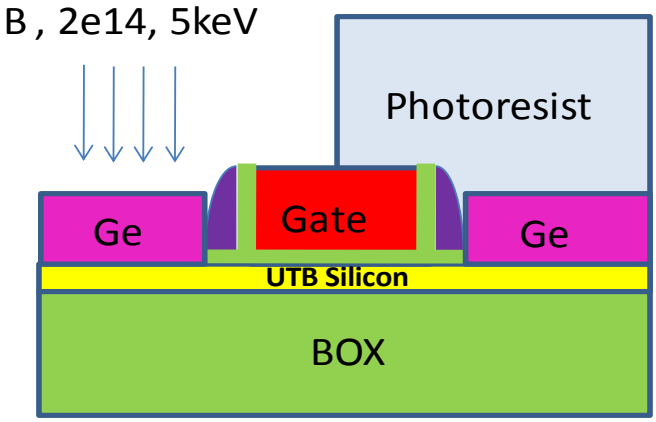


Figure 2.19 (e) Photoresist mask to protect the drain side and source dopant implant into the raised germanium source

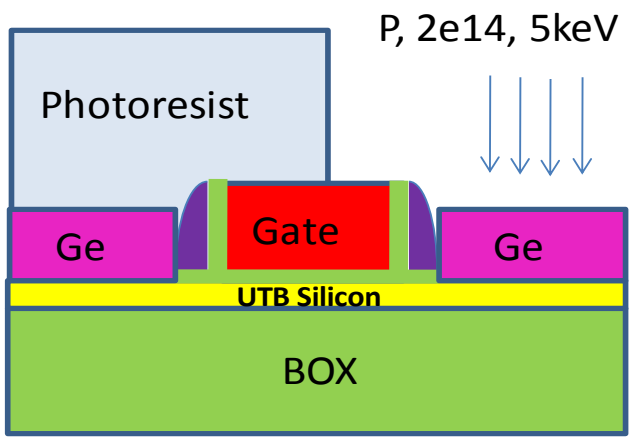


Figure 2.19(f) Photoresist mask to protect the source side and drain dopant implant into the raised germanium drain

Thinning the silicon on insulator down to 5nm was done in several cycles of dry oxidation followed by wet etch and removal of the oxide. Because of variation in the initial 100nm SOI substrate thickness, the thinned down silicon also had a variation in thickness. The center of the wafer was the thinnest (6.5nm) with rings of thicker silicon (8nm and 11nm) as you move outwards. After gate oxidation the thinnest body measured 5nm in the center of the wafer. Oxidation was followed by in-situ doped polysilicon deposition, gate lithography and etch. This was followed by source implant into the UTB aligned to the source edge. This was a variation from the UTB MOSFET process flow where no ion implantation was performed into the ultra thin body. For TFETs since the source needs to be overlapped by the gate in order to enable gate induced BTBT and therefore it was necessary to introduce dopants close to the gate edge.

A novel process used to fabricate UTB MOSFETs involved a bi-layer spacer consisting of HTO and Nitride. The HTO layer acts as a stopping layer for the nitride spacer etch eliminating the chance of etching into the silicon substrate and disconnecting the channel from the source and drain. In the UTB TFET a bilayer spacer with 20nm of HTO following by 20nm Nitride was used. After the nitride etch the exposed HTO is then cleaned off in HF to leave the silicon substrate exposed in the S/D region. Selective deposition of undoped germanium is then carried out in a LPCVD furnace. The germanium is capped with a LTO layer right away and then half masks are used to implant the source and drain while covering the other electrode. Finally dopant activation and FGA are performed before electrically testing the devices. YK Choi when fabricating UTB MOSFETs had noticed that annealing beyond 900 $^{\circ}$C lead to bursting of the channel region. To avoid this and still achieve dopant activation and damage annealing in the silicon the devices were annealed at 800 $^{\circ}$C for 30minutes.

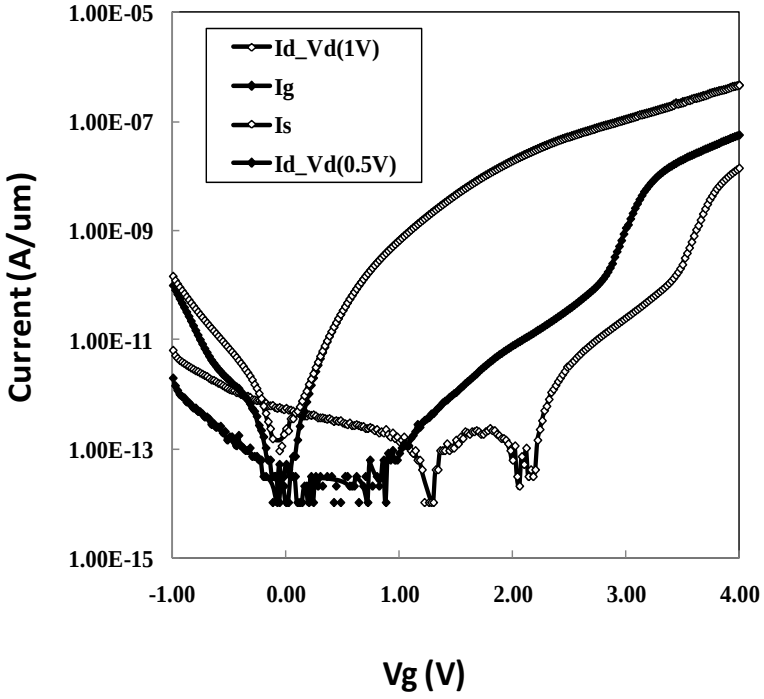


Figure 2.20 I_d-V_g measurements from an N-type UTB TFET. The gate current and source current are identical indicating a gate to source leakage mechanism.

Measurements from an N-type UTB TFET are shown in Figure 2.20. The gate current and source current are identical indicating a gate to source leakage mechanism. Ion implanting source dopants into the ultra thin body aligned to the gate edge is possibly the source of this large leakage current. A very thin ALD layer or LTO layer deposited to protect the gate edges before the ion implant could have possibly eliminated this gate edge damage. Typical MOSFET processing involves a re-oxidation step after the gate etch to heal any etch related damage to the gate edges. YK CHOI found during UTB MOSFET fabrication that the oxidation rate of SOI film less than 10nm was much higher than the normal oxidation rate of bulk silicon. In order to avoid rapid oxidation of the source and drain regions of the UTB TFETs, thermal re-oxidation after the gate etch was avoided which would have left any gate etch related damage un-healed.

In an N-type TFET a positive gate voltage allows for overlap and BTBT to occur in the source. A positive drain voltage reverse biases the PN junction and collects the BTBT generated electrons, while the BTBT generated holes flow to the source contact. Therefore the drain and source currents should be opposite in sign. In the figure shown above, the gate current is positive while both the drain and the source current are negative indicating that both source and drain terminals are collecting the same type of carrier. It appears that the positive gate voltage allows for electrons to be collected by the gate from both the source and the drain presenting no evidence of BTBT in these UTB TFETs. The I_d-V_g characteristic shown above is representative of most of the devices from the central dies of the wafer.

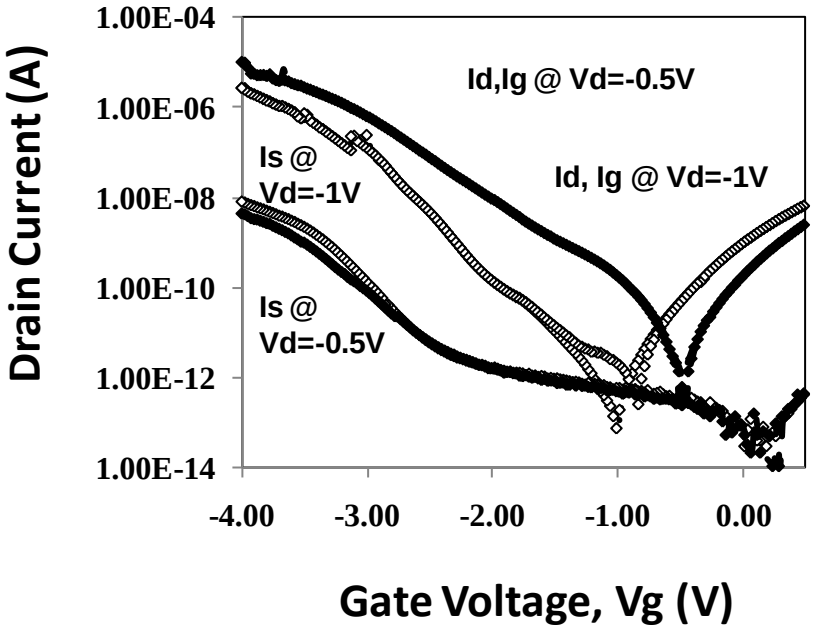


Figure 2.21 Id-Vg measurements from an N-type UTB TFET operated as a P-type TFET by grounding the N^+ terminal and applying a negative bias to the gate and P^+ terminal.

An N-type TFET measurement involves grounding the P+ terminal and applying a positive bias to the gate and N+ terminal to induce BTBT in the P+ source. The same transistor can however also operate as a P-type TFET by grounding the N+ terminal and applying a negative bias to the gate and P+ terminal. Here the N+ region is the source of BTBT while the P^+ acts as the drain.

The I_d-V_g characteristic from such a measurement is presented in Figure 2.21. The fact that in this measurement setup the drain to gate leakage dominates both at V_{ds} of -0.5V and -1V, can be explained as follows. The UTB TFET process flow designed to create N-type TFETs involved P+ source implants into the UTB aligned to the gate edge and N+ drain implants aligned to the 20nm bi-layer spacer. For a P-type measurement, the P+ region implanted at the gate edge with gate edge damage now acts as the drain and the N+ region not implanted at the gate edge acts as the source. Since the N+ implant was not aligned to the gate edge, dopant diffusion has to be relied on for overlap of the gate over the N+ region. This would lead to the gate overlapping lighter doped N+ regions which is not the ideal design for a TFET source.

The polarity of the source current in the Id-Vg measurement indicates collection of electrons. The increase in source current with the drain bias seems to indicate that the electrons are not entirely due to tunneling from the negatively biased gate into the source. An alternate generation mechanism of electrons exists, BTBT being a possibility. The I_d-V_d characteristic of the same P-type biased device is shown in Figure 2.22. For two different negative gate biases, the source current is seen to increase with increasing drain voltage and saturate just as expected from the I_d-V_d characteristic of a TFET. Again if the source purely collected electrons tunneling from the gate into the source, the source current wouldn't show an increase and saturation with

drain voltage again confirming an alternate generation mechanism for electrons collected by the source.

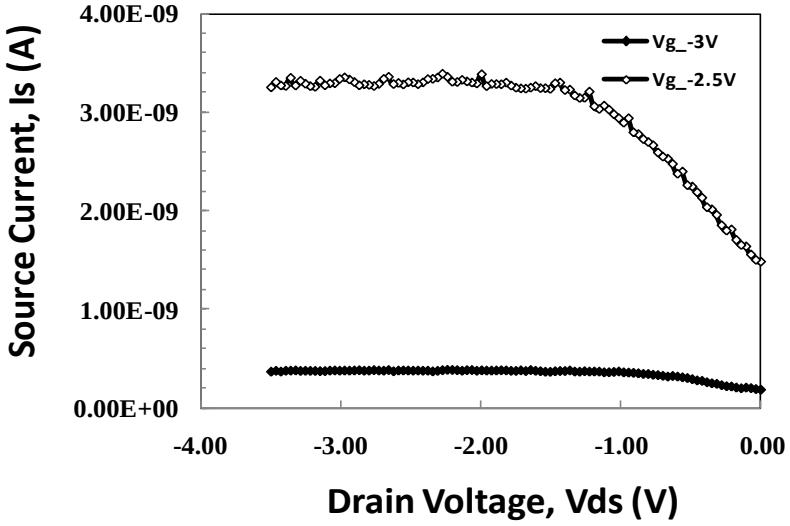


Figure 2.22 I_d-V_g measurements from an N-type UTB TFET operated as a P-type TFET by grounding the N+ terminal and applying a negative bias to the gate and P+ terminal.

Further annealing of the devices at 800 °C caused even larger increase in gate to drain leakage current, making it difficult to get further meaningful measurements from these devices.

In conclusion use of a protective layer to protect the gate edge before performing the source implant could eliminate or reduce the gate to source/drain leakage allowing us to truly observe and explore UTB TFET behavior. Solid phase epitaxy or annealing the devices at a low temperature like 400 °C for many hours could be explored to achieve better dopant activation in these devices while keeping the thermal budget low. Lastly using solid source diffusion to introduce dopants into the source and drain without damaging the silicon or gate edges could also be a possible direction for improved UTB TFET experiments.

2.6 REFERENCES

[2.1] T. Y. Chan, J. Chen, P. Ko, C. Hu, “The impact of gate-induced drain leakage current on MOSFET scaling”, International Electron Devices Meeting, Vol.33, pp: 718-721, 1987

[2.2] W. M. Reddick, G. A. Amaratunga, "Silicon surface tunnel transistor," Applied Physics Letters, vol.67, no.4, pp.494-496, Jul 1995

[2.3] C. Aydin, A. Zaslavsky, S. Luryi, S. Cristoloveanu, D. Mariolle, D. Fraboulet, S. Deleonibus, "Lateral interband tunneling transistor in silicon-on-insulator," Applied Physics Letters , vol.84, no.10, pp.1780-1782, Mar 2004

[2.4] K. K. Buwalka, M. Born, M. Schindler, M. Schmidt, T. Sulima and I. Eisele, " P-Channel Tunnel Field-Effect Transistors down to Sub-50nm Channel Lengths", Japanese Journal of Applied Physics, 45, pp. 3106 -3109, 2006
[2.5] C. Hu, "Green transistor as a solution to the IC power crisis," Solid-State and Integrated-Circuit Technology, 2008. ICSICT 2008. 9th International Conference on, pp.16-20, 20-23 Oct. 2008
[2.6] A. Bowonder, P. Patel, K. Jeon, J. Oh, P. Majhi, H. H. Tseng, C. Hu; , "Low-voltage green transistor using ultra shallow junction and hetero-tunneling," Junction Technology, 2008. IWJT '08, pp.93-96, 15-16 May 2008
[2.7] V. Nagavarapu, R. Jhaveri, J.C.S. Woo, "The Tunnel Source (PNPN) n-MOSFET: A Novel High Performance Transistor," Electron Devices, IEEE Transactions on, vol.55, no.4, pp.1013-1019, April 2008
[2.8] P, Patel, K. Jeon, A. Bowonder, C. Hu, "A Low Voltage Steep Turn-Off Tunnel Transistor Design," Simulation of Semiconductor Processes and Devices, 2009. SISPAD '09. International Conference on, vol., no., pp.1-4, 9-11 Sept. 2009.
[2.9] E.O.Kane, "Theory of Tunneling", JAP, Vol. 23, No.1, Jan, 1961
[2.10] S.Wang, Fundamentals of Semiconductor Theory and Device Physics, 1989
[2.11] Z.A.Weinberg, "On Tunneling in Metal-Oxide-Silicon Structures", JAP, Vol. 53, No.7, July 1982
[2.12] C. Hu, D. Chou, P. Patel, A. Bowonder, "Green Transistor – A V_{DD} Scaling Path for Future Low Power ICs", VLSI Technology, Systems and Applications, 2008. VLSI-TSA 2008. International Symposium on. April 2008, pp. 14-15.
[2.13] H. C. H. Wang, C. W. Tsai, S. Chen, C. T. Chan, H. J. Lin, Y. Jin, H. J. Tao, S. C. Chen, C. H. Diaz, T. Ong, A. S. Oates, M. S. Liang, M. H. Chi, "Reliability of HfSiON as gate dielectric for advanced CMOS technology," VLSI Technology, 2005. Digest of Technical Papers. 2005 Symposium on, pp. 170- 171, 14-16 June 2005
[2.14] C. Y. Kang, C. D. Young, J. Huang, P. Kirsch, D. Heh, P. Sivasubramani, H. K. Park, G. Bersuker, B. H. Lee, H. S. Choi, K. T. Lee, Y. H. Jeong, J. Lichtenwalner, A. I. Kingon, H. H. Tseng, R. Jammy, "The impact of la-doping on the reliability of low V_{th} high-k/metal gate nMOSFETs under various gate stress conditions," Electron Devices Meeting, 2008. IEDM 2008. IEEE International, vol., no., pp.1-4, 15-17 Dec. 2008
[2.15] M. M. Hussain, N. Moumen, J. Barnett, J. Saulters, D. Baker and Z. Zhang, "Metal Wet Etch Process Development for Dual Metal Gate CMOS", Electrochem. Solid-State Lett., Volume 8, Issue 12, pp. G333-G336 (2005)
[2.16] Y. K. Choi, A. K, N. Lindert, V. Subramanian, T. J. King, J. Bokor,C. Hu; , "Ultra-thin body SOI MOSFET for deep-sub-tenth micron era," Electron Devices Meeting, 1999. IEDM Technical Digest. International, pp.919-921, 1999
[2.17] Y. C. Yeo, V. Subramanian, J. Kedzierski, X. Peiqi, T. J. King, J. Bokor, C. Hu, "Nanoscale ultra-thin-body silicon-on-insulator P-MOSFET with a SiGe/Si heterostructure channel," Electron Device Letters, IEEE , vol.21, no.4, pp.161-163, Apr 2000
[2.18] P. Xuan, J. Kedzierski, V. Subranmanian, J. Bokor, T. J. King, C. Hu; , "60 nm planarized ultra-thin body solid phase epitaxy MOSFETs," Device Research Conference, 2000. Conference Digest. 58th DRC, pp.67-68, 2000

Chapter 3

Lateral Pocket TFETs as a Path to Scale V_{dd}

3.1 INTRODUCTION

The previous chapter explored simple PIN silicon TFET designs. In these PIN transistors, the tunneling current is exponentially dependant on the field set up by the gate and therefore the instantaneous swing constantly varies and gradually increases with current. Since low I_{off} is required to keep the standby power consumption low, the transistor with lowest I_{off} and largest I_{on}/I_{off} at a reduced V_{dd} is desired for ultra low voltage operation. The simple PIN design (Figure 3.1a) can be further enhanced by using well engineered dopant pockets [3.1, 3.2] to enhance the I_{on}/I_{off} at ultra low V_{dd} (V_{dd} < 0.3V). These dopant pockets are thin regions of dopant opposite in type to the source dopant. Dopant pockets can be incorporated in addition to low bandgap materials for even higher performance at ultra low voltages. This chapter presents an overview of the operation and optimization of a silicon pocket TFET design. Further, experimental data which confirms the superior behavior of the pocket TFET over a simple PIN TFET is also presented.

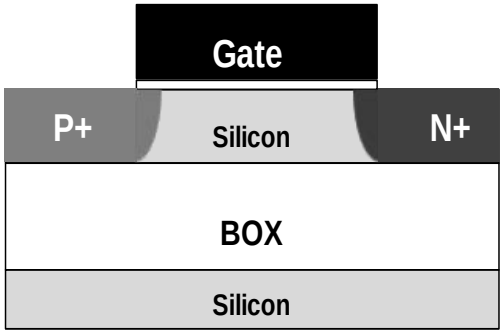


Figure 3.1a. Si PIN TFET with P+ source and N+ drain. The TFETs in this work were fabricated on SOI, 40nm Silicon on 100nm Box.

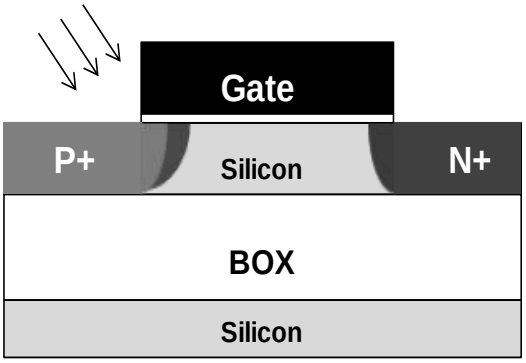


Figure 3.1b. Si PIN TFET with angled implant to create a lateral N doped pocket adjacent to the P+ source. LPTFETs were also fabricated on 40 nm Silicon on 100nm Box.

3.2 LATERAL POCKET TFET DESIGN

The pocket regions can be either perpendicular to the gate dielectric, adjacent to the source (lateral pocket-LPTFET -Figure 3.1b) [3.1] or parallel to the dielectric interface, above the source (vertical pocket-VPTFET [3.2]). The lateral pocket TFET design (Figure 3.1b) is

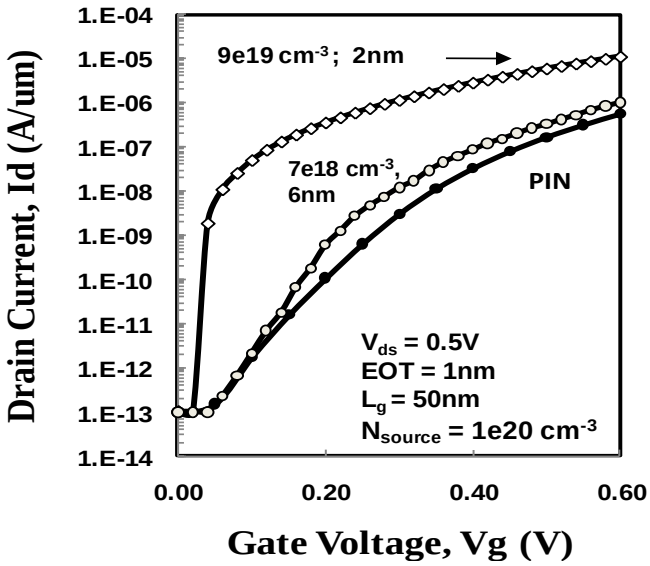


Figure 3.2 TCAD simulations of N-type lateral pocket TFETs and PIN TFET. An ideal heavily doped, fully depleted N type pocket (9e19 cm^{-3}, 2nm) adjacent to the source leads to significant enhancement in performance over PIN TFET, while a not so optimal N doped pocket shows minimal enhancement over a PIN TFET.

referred to as tunnel source PNPN MOSFET [3.1] and is explored using SILVACO's device simulator, ATLAS, coupled with analytical tunneling calculations by Nagavarapu et.al. Some experimental data is also shown in [3.1]. In this work we add to the previous study by presenting self consistently done TCAD simulations and very detailed experimental characterization. The simulations were done with Sentaurus' dynamic non local tunneling model with the default silicon tunneling parameters (obtained from calibration with experimental GIDL current of MOSFETs).

Shown in Figure 3.2 are I_d-V_g simulations of an ideal lateral pocket TFET and simple PIN TFET. For the same gate WF the LPTFET turns on earlier than the PIN TFET but for a fair comparison (same V_{gs} -V_t), it is assumed that WF engineering is used to achieve identical turn on voltage for both transistors. For a PIN TFET, even at 1nm EOT, the sub 60mV/decade swing is only seen up to 10nA of current. The ideal lateral pocket TFET however can show steep switching over a much larger current range (sub 60mV/decade swing up to 5uA). For the same I_{off} the ideal LPTFET also shows enhanced I_{on}/ I_{off} over that of an ideal MOSFET (60mV/decade swing) at low V_{dds} below 0.4V. Well implemented pocket TFETs in silicon can therefore help lower V_{dd} below 0.4V. For larger I_{on} and enhanced I_{on}/I_{off} at V_{dds} below 0.4V, these lateral pockets can further be implemented in Ge, low bandgap III V materials and heterostructures with even smaller effective bandgap.

Figure 3.2 also shows a non optimal lateral pocket design where the enhancement over a PIN TFET is minimal. The ideal pocket is fully depleted and works by enhancing the E field at the source-pocket junction at the point of turn on of BTBT, while suppressing less efficient tunneling paths which would have resulted in a worse swing. This enhanced E field when BTBT first occurs allows for a sudden jump from the off state (junction leakage) to the enhanced BTBT

current value as seen in Figure 3.2. The no. of decades over which the current will jump from off state to on state will depend on the electric field at the point of overlap for band to band tunneling to occur.

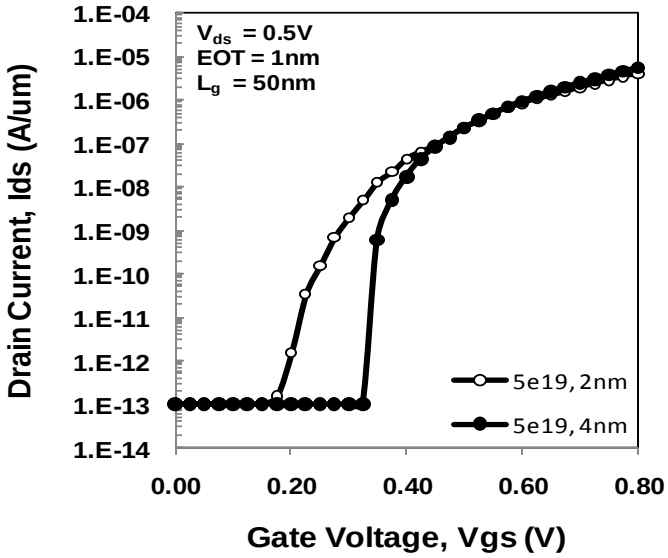


Figure 3.3 LPTFET with varied pocket dose (5e19 cm^{-3}, 4nm and 5e19 cm^{-3}, 2nm). In the 5e19 cm^{-3} N doped, 4nm pocket, the onset of BTBT occurs in the pocket and other less efficient tunneling paths are cut off.

The pocket dose thickness and doping therefore needs to be carefully engineered so that the onset of tunneling occurs within the high field pocket region.

Figure 3.3 and figure 3.5 compare LPTFETs with varied pocket dose. For a steep sudden turn on, the pocket width for a fixed doping concentration needs to be optimized to ensure the pocket is fully depleted and the onset of BTBT is in the high field region. This is also emphasized in [3.1]. Figure 3.3 compares pockets of same doping but varied thickness. In the wider pocket (5e19 cm^{-3}, N- doped, 4nm) the onset of BTBT occurs in the high field pocket region and other less efficient tunneling paths are cut off. The wider pocket therefore shows a steeper turn on. This is confirmed in Figure 3.4a and 3.4b, TCAD outputs showing contours of the BTBT generated carriers at the onset of tunneling for the 2nm pocket and 4nm pocket.

Figure 3.5 compares 2nm pockets of varied doping. We see that for a given pocket thickness increasing the doping of the pocket (net dose of the pocket) increases the E field at the source-pocket junction and therefore increases the tunneling current when the conduction and valence bands first overlap. This allows for a sudden jump from off state to a larger current value in the on state and therefore steeper turn on over a larger current range.

In addition to an overview of the optimization and design of the LPTFET, this chapter deals with characterization and analysis of prototype PIN TFETs with and without lateral pocket with 2.5nm EOT. While the use of a conservative dielectric prevents sub 60mV/decade swing and limits the on current of the TFETs, the prototypes demonstrate experimental verification of the enhancement in swing and on current due to lateral pocket design.

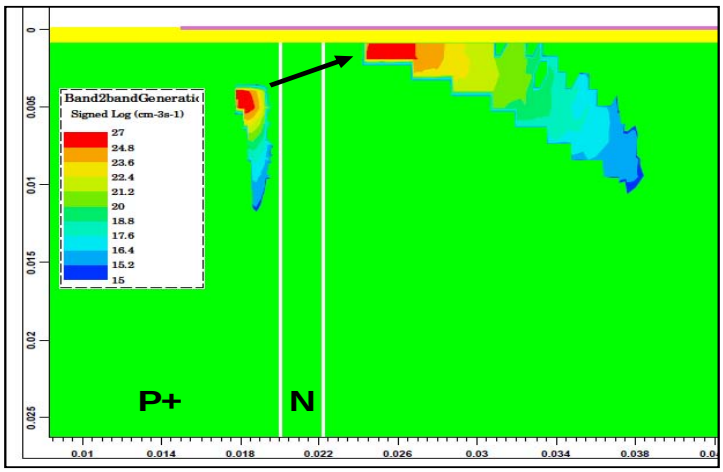


Figure 3.4a TCAD output of electrons and holes generated by BTBT for LPTFET with 5e19cm^{-3}, N-doped, 2nm pocket. This is the output at 0.05V above turn on. Onset of tunneling does not occur in the high field pocket region. This leads to lower tunneling current at the onset of tunneling.

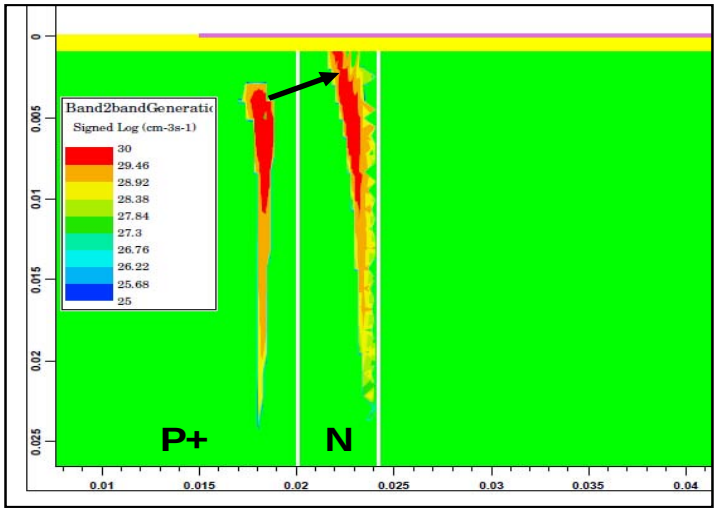


Figure 3.4b TCAD output of electrons and holes generated by BTBT for LPTFET with 5e19cm^{-3} N doped, 4nm pocket. This is the output at 0.05V above turn on (the sudden jump to on current). Onset of tunneling occurs in the high field pocket region. This leads to a sudden jump from off state to a high current in the on state. The higher the field at the onset of tunneling, the higher the jump to on current value will be (shown in Fig. 4).

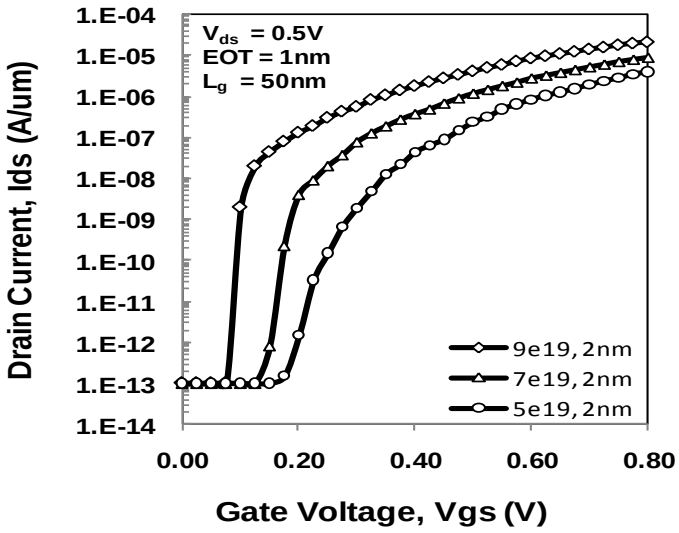


Figure 3.5 LPTFETs with fixed pocket width and varied pocket dose (5e19 cm^{-3}, 2nm; 7e19 cm^{-3}, 2nm and 9e19 cm^{-3}, 2nm). The E field at the source-pocket junction increases with heavier pocket doping allowing for larger tunneling current at the onset of BTBT and steeper turn on.

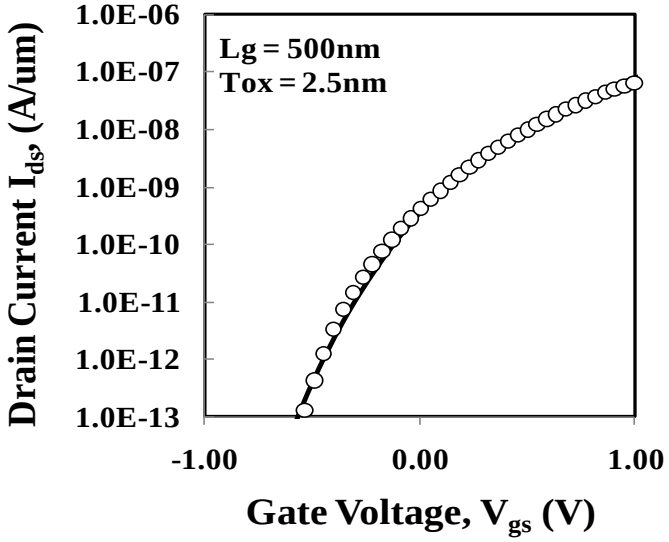


Figure 3.6 Comparison of TCAD simulation with experimental data of Si PIN TFET. TCAD simulations involve process simulation emulating fabrication followed by device simulation. Sub 60mV/dec cannot be achieved from a silicon PIN with 2.5nm EOT.

Compared to a Si PIN TFET, the LPTFET presented in this work has 14x larger current at $V_{ds} = V_{gs} - V_t = 0.5V$, 10x larger current at $V_{ds} = Vgs - V_t = 1V$ and reduction in minimum swing from 87mV/decade to 63mV/decade. The control TFET's I_{on} is comparable to simple silicon PIN TFETs from recent literature [3.3]. The experimental data also shows a good fit with TCAD simulations of the PIN TFET (Figure 3.6). Enhancements in the simple PIN TFET I_{on} by use of lower bandgap materials has already been demonstrated by several groups. LPTFETs can be implemented in these materials to further enhance performance and swing just as in the case of silicon.

The possible formation of a parasitic MOSFET during LPTFET fabrication by angled implant and anneal is also discussed in detail along with experimental results and analysis.

3.3 EXPERIMENTAL DESIGN

Almost all studies of TFETs to date have involved implanted source and drain regions [3.1, 3.3,-3.5], and avoided the integration of selective Epitaxy of the source and drain. Given the significant process complexity of fabricating LPTFETs using selective epitaxial growth of the pocket and source regions, in this initial study, a controlled angled implant aligned to the gate edge followed by a spike anneal is used to create the pocket. The enhancement due to pocket presented here will be greatly enhanced in TFETs with abrupt, epitaxial pocket and source.

PIN TFETs and LPTFETs (inset in Fig3.1b) were fabricated on SOI wafers with 45nm body thickness. The SOI wafers with 100nm box had 100nm body originally and were thinned down to 45nm by consecutive dry oxidation and etch of the oxide grown. Alignment marks are patterned onto the wafer and etched 1200Å deep into the wafer using a CF_4 based dry etch. The gate stack is then created. The 2.5nm gate oxide was grown by dry oxidation at 850 $^{\circ}$C for 1min and capped with 1000Å of in situ N+ doped poly-silicon deposited at 615 $^{\circ}$C. The drain is covered with a photoresist half mask aligned to the center of the gate and the P+ source is created by a vertical BF_2 implant ($5x10^{15}cm^{-2}$, 10keV) aligned to the gate edge. The SEM is used to ensure that the half mask is aligned appropriately and covers the drain. In case of alignment inaccuracies an intrafield offset can be specified in the ASML stepper to help align the mask to the center of the gate. With the source covered by a photoresist half mask the N+ drain is created by a vertical As implant ($4x10^{15}cm^{-2}$, 10keV). The wafer splits with lateral N+ pockets received an additional angled As implant ($2x10^{14}$, 20keV, 20^{o}) aligned to the gate edge on the source side. Implant conditions were based on TCAD process simulations, followed by device simulations to ensure optimal TFET design. This was followed by deposition of 1000Å of low temperature oxide deposited at 450 $^{\circ}$C. The oxide is deposited before dopant activation to prevent dopant loss during high temperature activation. The silicon PIN control wafer without lateral pocket implant was annealed at 1000 $^{\circ}$C for 1s using an RTA tool while the LPTFET splits received dopant activation anneals at 1000 $^{\circ}$C for 1s and at 1000 ^{0}C for 5s. In the RTA tool the anneal is ramped up to 450 $^{\circ}$C and held steady for 30s after which it is ramped up-to 1000 ^{0}C in 5s. Contact lithography is followed by a CF_4 and CHF_3 based anisotropic etch with 9:1 selectivity to the silicon substrate. Endpoint detection on a dummy with identical ILD thickness is used to estimate the time for contact etch and a 30% overetch is added in. A quick 3s HF dip is performed before the wafers are put into the sputter tools loadlock. A 10s sputter etch is first performed in the sputter tool followed by deposition of 1000Å of Al/Si 2%. The metal is

patterned and etched with a BCl_3 based anisotropic etch and the wafers receive a FGA at 450 °C for 30minutes before testing.

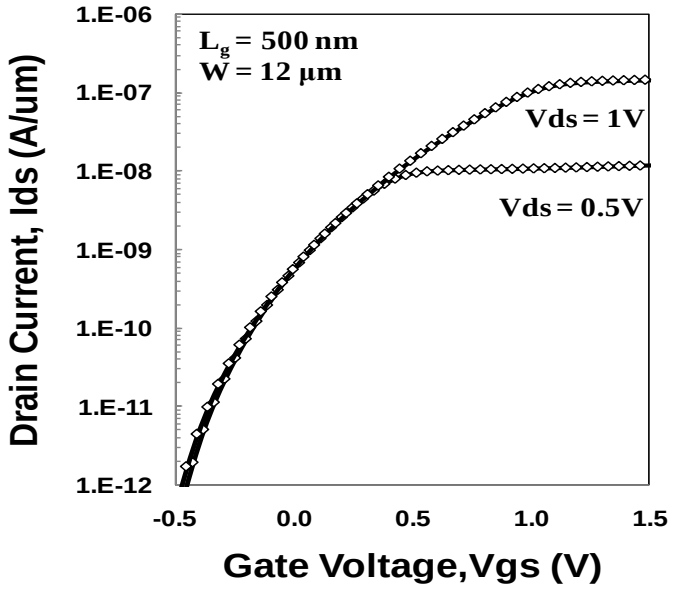


Figure 3.7 Experimental Id-Vg characteristics of a PIN TFET without pocket.

3.4 EXPERIMENTAL RESULTS

Figure 3.7 shows experimental I_d-V_g data from the PIN control TFET split without a pocket. The data fits very well to an analytical TFET framework based on the charge sheet model [3.6: to be published elsewhere]. While all the equations of the analytical framework are to be published in detail in [3.6], a very brief summary is provided here. The shape and saturation of the TFET I_d-V_g data presented in Figure 3.7 are dependent entirely on the electrostatics as follows. The tunneling current is determined by the BTBT generation rate, which is exponentially dependent on the E-field at dielectric-silicon interface (E-field at the interface has a square root dependence on the surface potential). For a given V_{ds}, the surface potential increases linearly with V_{gs} at low gate voltages and then saturates with V_{gs} at high gate voltages [3.6]. The point at which the surface potential saturates with V_{gs} increases with V_{ds} [3.6].

Figure 3.8a and Figure 3.8b show the experimental data from two different LPTFET splits with varied annealing conditions (1s, 1000 °C and 5s, 1000 °C). The control PIN TFET split without pocket is also shown in both figures. Measured transistor currents are as low as $1x10^{-14}$ A/µm, but $1x10^{-12}$A/µm is assumed as a realistic I_{off} for all transistors. Also for all transistors the point of onset of tunneling current above this I_{off} is defined as the turn on voltage (V_t). From the TCAD simulations (Figure 3.6), Si PIN TFETs with 2.5nm EOT exhibit minimum swing ~ 90mV/decade at pA. A thinner T_{ox} (~1nm EOT) is required for < 60mV/decade swing but was not used in the prototypes to eliminate potentially enhanced gate leakage after angled pocket

implant. The results and conclusions presented here with a conservative dielectric do however experimentally verify the benefit of and the enhancement in swing and on current with lateral pocket design.

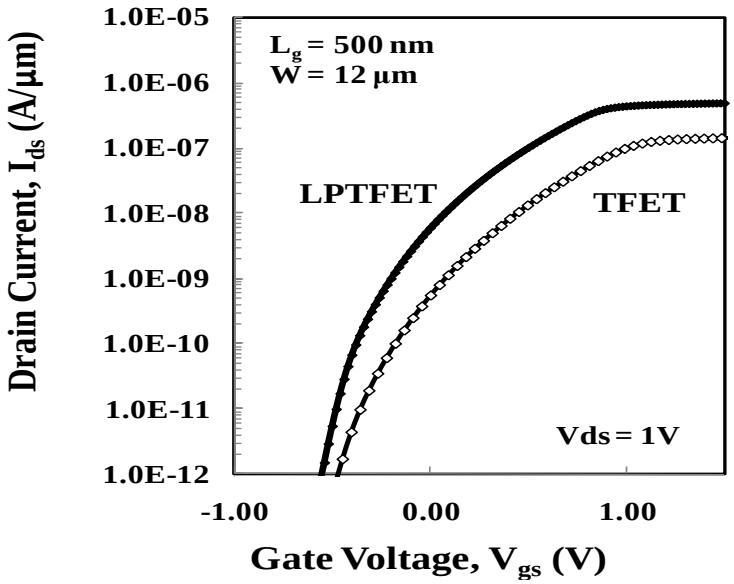


Figure 3.8a Experimental Id-Vg characteristics of a PIN TFET and a LPTFET (angled N implant, 1050 °C ,1s anneal) with identical 2.5nm EOT gate stacks. The LPTFET shows improved performance over PIN TFET. The minimum swing is reduced from 89mV/decade for the TFET to 63mV/decade for the LPTFET.

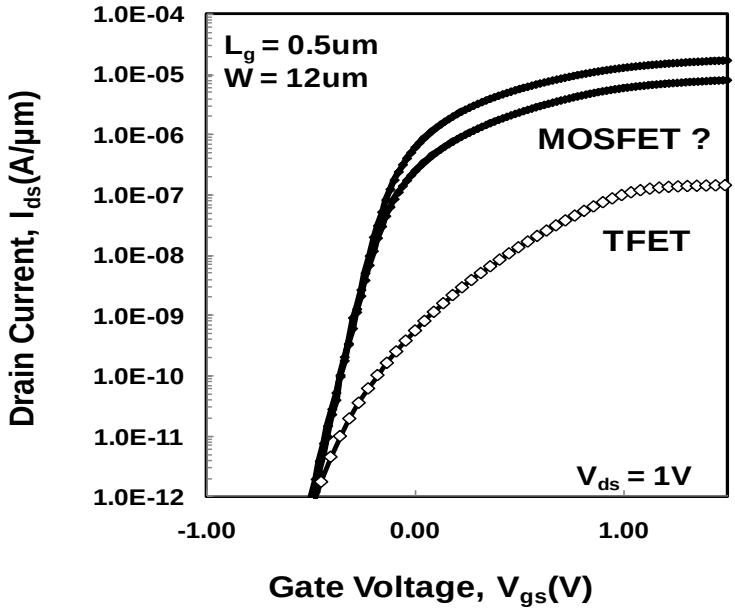


Figure 3.8b Experimental Id-Vg characteristics of a PIN TFET and LPTFETs (angled N implant, 1050 °C 5s anneal) with identical 2.5nm EOT gate stacks. Constant swing over several decades of current seems indicative of accidentally formed MOSFETs rather than LPTFETs.

The LPTFET in Figure 3.8a compared to the Si PIN TFET has 14x larger and 10x larger current at $V_{ds} = V_{gs} - V_t = 0.5V$ and 1V respectively. As seen in Figure 3.2, a wide lightly doped pocket, as would be expected after angled implant and anneal, does not show as much enhancement over a PIN TFET as a thin, heavily doped pocket with the optimal pocket dose. Therefore to see this steep turn on a thin, heavily doped pocket formed by epitaxial silicon growth needs to be employed.

The instantaneous swing of the LPTFET in Figure 3.8a is lower than the PIN TFET over all current ranges and the minimum swing is reduced from 89mV/decade to 63mV/decade. Since ellipsometry during fabrication was used to ensure identical T_{ox} on all splits and all other processing was identical for the two splits, the enhanced swing is attributed to the presence of a lateral pocket. The LPTFET data in Figure 3.8a is characteristic of the devices in that particular wafer split with the minimum swing of the devices in that split varying between 63mV/decade and 69mV/decade.

Since the LPTFET if formed by angled implant and anneal it is very challenging to achieve a thin uniformly doped, fully depleted pocket. Figure 3.8b shows the LPTFET split with a longer anneal time (1000 °C, 5s). All working transistors from this split demonstrated constant swing over several decades of current and ~100x greater current than the control PIN TFET split with no pocket. Nagavarapu et.al explained in detail the possible formation of a parasitic MOSFET if the N doped pocket is very wide and not fully depleted [3.1]. In other words instead of creating a thin depleted pocket of charge to enhance the E-field and tunneling at the source-pocket junction you form a parasitic NPN+ transistor and the injection mechanism of carriers into the channel is no longer tunneling from the source but is injection over a barrier. The formation of such a wide un-depleted pocket in the longer anneal split LPTFETs can be confirmed by verifying the carrier injection mechanism. If carriers are injected thermally over a barrier, the swing of these transistors will show a KT dependence. Figure 3.10 and Figure 3.11 explore the temperature dependence of the two LPTFET splits in Figure 3.8a and Figure 3.8b.

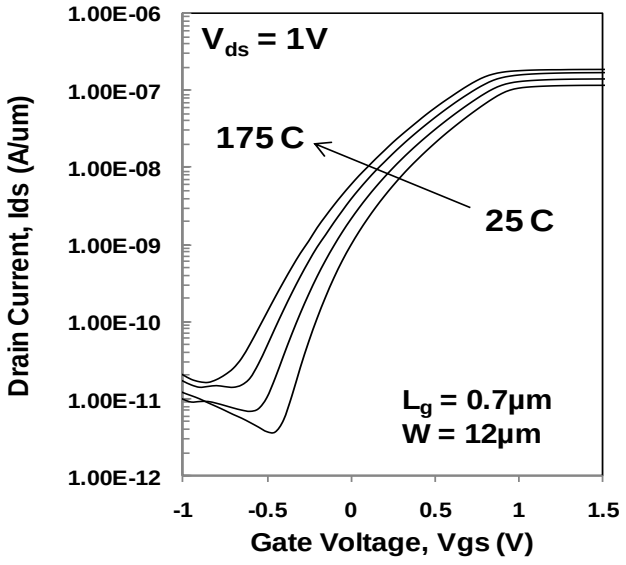


Fig. 3.9 Effect of temperature on Id-Vg characteristics of the PIN TFET

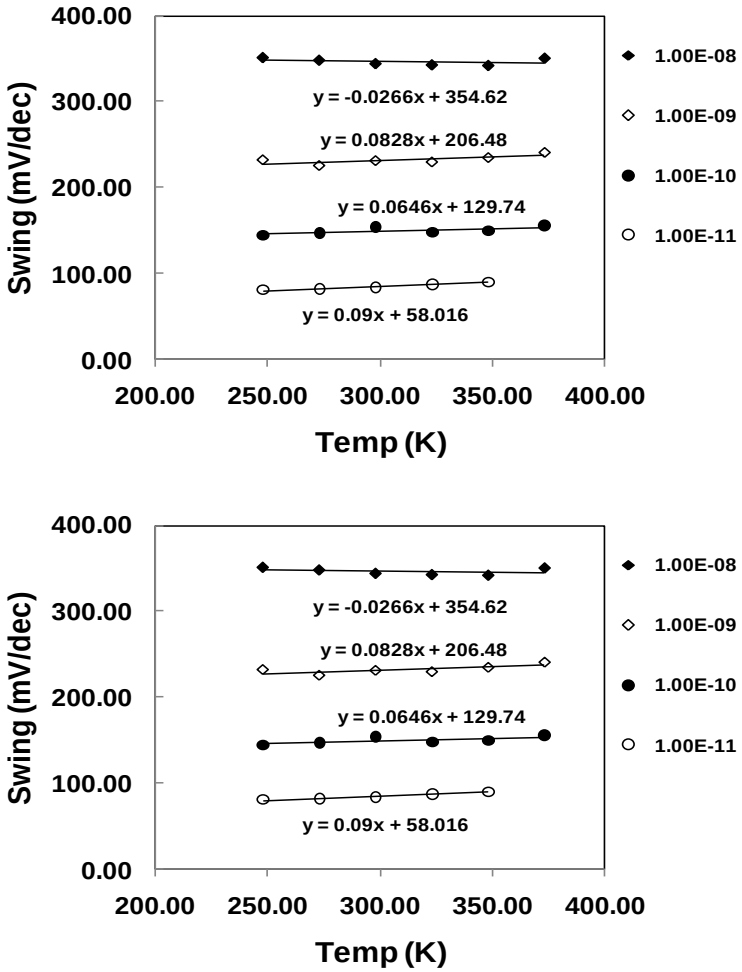


Figure 3.10 Effect of temperature on Id-Vg characteristics of the LPTFET shown in Figure 3.8a for 50°C steps in temperature. Swing of the LPTFET at various current ranges shows no kT dependence on absolute temperature (slope of the trendline).

In Figure 3.10, the LPTFET I_d-V_g exhibits parallel shifts with temperature. As expected from theory, the tunneling $V_{turn,on}$ decreases with increase in temperature because the bandgap decreases and therefore the V_{gs} to achieve sufficient band-bending for BTBT decreases with temperature. The swing at various current levels shows no KT dependence with absolute temperature, confirming a tunneling based injection of carriers into the channel. In Figure 3.11, the LPTFET I_d-V_g exhibits varying slope with temperature. The swing shows a KT dependence on absolute temperature confirming the presence of carrier injection over a barrier. This confirms the hypothesis that the 1000 °C, 5s anneal split lead to a very wide pocket and the accidental formation of a NPN+ transistor, while the 1000 °C, 1s anneal while not forming the ideal pocket, allowed the formation of a fully depleted N pocket with better performance than the control PIN TFET without pocket.

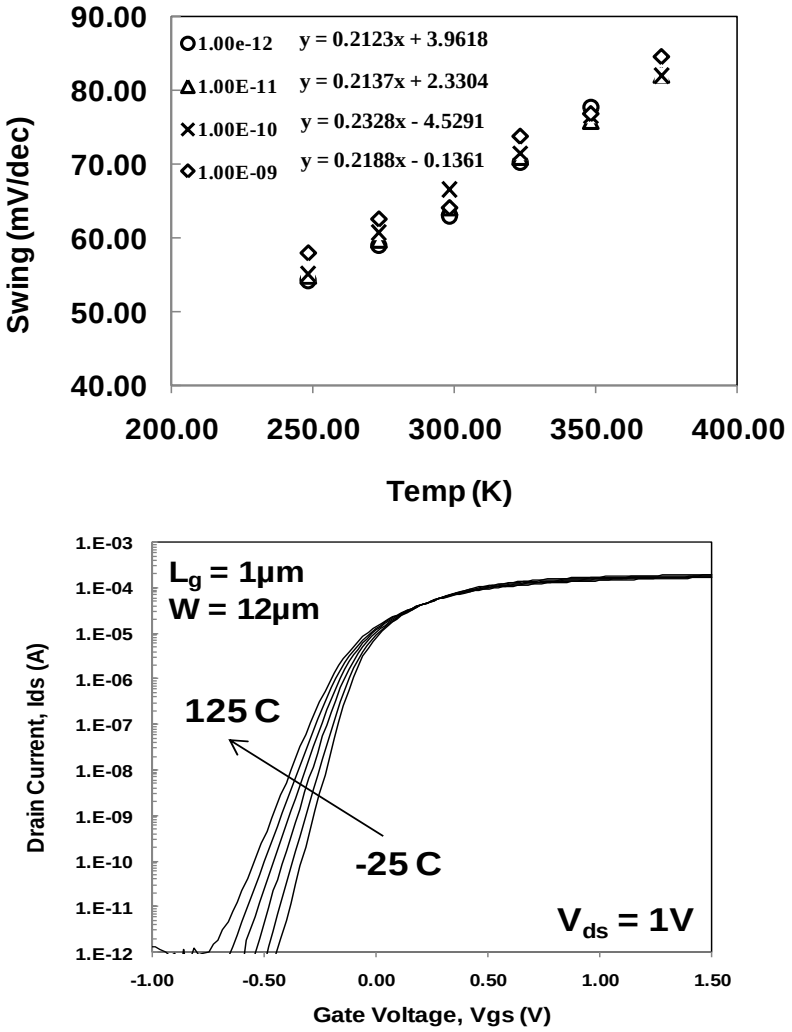


Figure 3.11 Effect of temperature on Id-Vg characteristics of the accidental MOSFET shown in Figure 3.8b for 25 °C steps in temperature. Swing at various current ranges shows a KT dependence on absolute temperature (slope of the trendlines) confirming the formation of an accidental MOSFET.

The I_d-V_d characteristics (Figure 3.12a, 3.12b) of the two LPTFET splits are also presented as further verification. The output characteristics of the LPTFET exhibit non-linearity at low V_{ds} as expected from analytical theory [3.6]. When tunneling is the carrier injection mechanism, the current (generation rate) is exponentially dependent on E-field (surface potential) at the interface. For a given V_{gs}, the surface potential increases linearly with V_{ds} until it finally becomes independent of the V_{ds} [3.7]. The tunneling current therefore increases exponentially with V_{ds} and eventually saturates at high drain bias. Complete details and fit of the I_d-V_d data with theory will be published in [3.6]. When the carriers are injected over a barrier, no such non-linearity is expected in the output characteristics. This is seen in Figure 3.12b, again confirming the hypothesis that the 1000 °C, 5s anneal lead to formation of an accidental MOSFET.

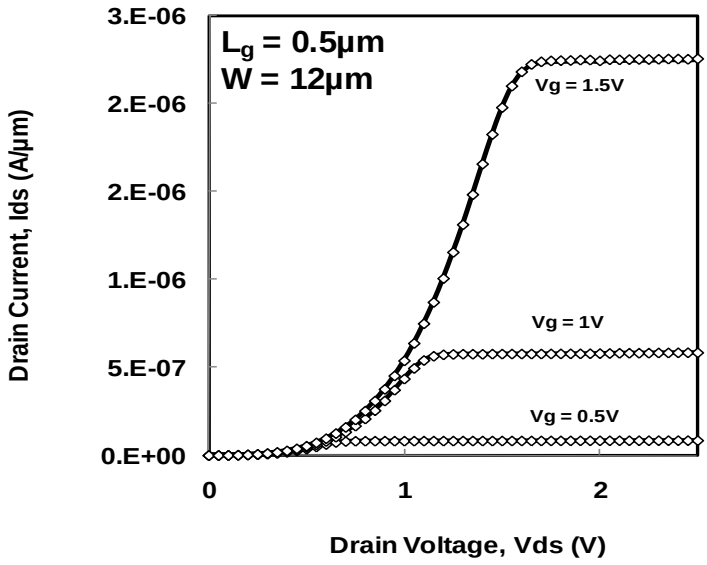


Figure 3.12a I_d-V_d characteristics of the LPTFET shown in Fig.7a. The non-linearity at low V_{ds} is because I_{ds} is exponentially dependant on E-field at the surface (the surface potential). For a given V_{gs} surface potential increases linearly with V_{ds} until it becomes independent of V_{ds} [3.6, 3.7].

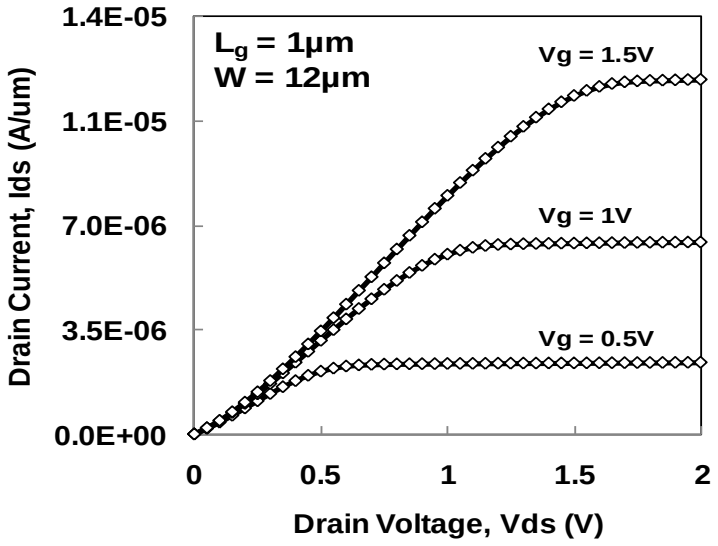


Figure 3.12b I_d-V_d characteristics of the accidental MOSFET split.

3.5 CONCLUSIONS

LPTFETs with optimized pocket dose and width can enhance performance of a PIN TFET greatly and enable V_{dd} scaling to below 0.5V. An angled N pocket implant is used to fabricate LPTFETs in silicon. An optimized anneal condition while not forming an ideal pocket yielded a LPTFET with enhanced performance over the control PIN TFET, while with a longer anneal condition a wide, not fully depleted N pocket formed causing carrier injection over a barrier instead of tunneling from the source. The dependence of swing on absolute temperature was used to confirm carrier injection mechanism in the varied LPTFET splits.

The use of lateral pocket to enhance performance is verified experimentally in this work. Further performance enhancement than demonstrated here can be achieved with better control of the source and lateral pocket profiles and thinner gate dielectric. If implantation is used, flash annealing or other diffusion-less annealing methods can help realize more abrupt source-pocket junctions than possible by RTA. Selective epitaxial growth of source and pocket regions would be ideal to provide maximum enhancement in LPTFET performance and enable V_{dd} scaling. The Enhancement in swing and on current due to lateral pocket opens up a path to lower TFET operating voltage which can be combined with bandgap scaling to achieve high I_{on} and high I_{on}/I_{off} at ultra low V_{dd}.

3.6 REFERENCES

[3.1] V. Nagavarapu, R. Jhaveri and J. C. S. Woo, "The Tunnel Source (PNPN) n-MOSFET: A Novel High Performance Transistor," IEEE Transaction on Electron Devices, Vol. 55, No.4, April 2008, pp.1013-1019.

[3.2] P. Patel, K. Jeon, A. Bowonder, C. Hu, "A Low Voltage Steep Turn-off Tunnel Transistor Design," SISPAD Sept 2009.

[3.3] F. Mayer, C. L. Royer, J. F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali and S. Deleonibus, "Impact of SOI, $Si_{1-x}Ge_xOI$ and GeOI substrates on CMOS compatible Tunnel FET performance," IEDM 2008, pp.163

[3.4] T. Krishnamohan, K. Donghyun, S. Raghunathan, K. Saraswat, "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) With record high drive currents and ≪60mV/dec subthreshold slope," Electron Devices Meeting, 2008. IEDM 2008. IEEE International, vol., no., pp.1-3, 15-17 Dec. 2008

[3.5] W. Y. Choi, B. G. Park, J. D. Lee and T. K. Liu, "Tunneling Field-Effect Transistors with Subthreshold Swing Less Than 60mV/dec," IEEE Electron Device Letters, Vol. 28, No. 8, Aug 2007, pp.743-745.

[3.6] Compact Modelling of TFETs, P. Patel, A. Bowonder, and C.Hu (To be published)

[3.7] Operation and Modelling of the MOS Transistor, Y. Tsividis, Oxford University Press 2nd edition, Chapter 3.

Chapter 4

Biaxial Strain Engineering for Sub-0.5V Operation Si/Ge Hetero TFET Design

4.1 INTRODUCTION

A transistor that can outperform a MOSFET at ultra low V_{dd} is highly desirable for reducing IC power consumption. Transistors which use band-to-band tunneling (BTBT) to inject carriers into the channel [4.1- 4.10] instead of injecting carriers over a barrier are not limited by 60 mV/decade turn-off. These tunneling transistors (TFETs) are therefore of great interest for high I_{on}/I_{off} at low voltages. TFET performance is limited by the BTBT generation rate which is exponentially dependant on the effective tunneling bandgap of the semiconductor. The obvious means to enhance TFET performance and scale V_{dd} is by scaling this effective tunneling bandgap. This can be achieved by moving from Si [4.5] to Ge [4.6] to even lower bandgap III-V materials.

Since low I_{off} keeps the standby power consumption low, a TFET with lowest I_{off} and largest I_{on}/I_{off} at a reduced V_{dd} is desired for ultra low voltage operation (V_{dd} < 0.4V). In this chapter biaxially strained silicon (Si) and germanium (Ge) based heterostructures are explored as a means to scale V_{dd} below 0.4V [4.7, 4.8]. When a semiconductor film is epitaxially grown lattice matched to a substrate with a different lattice constant, the semiconductor is biaxially strained. The film remains biaxially strained until it reaches a certain critical thickness after which relaxation occurs. The critical thickness depends on the extent of lattice mismatch between the substrate and strained film. Biaxial strain in silicon-germanium based systems has been intensely explored for mobility enhancement in MOSFETs [4.11, 4.12] but in this work strain engineering is explored from a pure bandgap perspective. Biaxial strain shifts the conduction and valance band energy levels of semiconductors [4.13]. The strain induced band offsets of biaxially strained $Si_{1-x}Ge_x$ heterostructures is exploited to engineer ultra low effective tunneling bandgaps for enhanced performance without the enhanced junction leakage and I_{off} of ultra low bandgap semiconductors. In addition Si and Ge based heterostructure TFETs atleast in the immediate future can benefit from existing technology developments used to drive MOSFET scaling.

4.2 EFFECTIVE TUNNELING BANDGAP

Band to band tunneling can be thought of as an electron penetrating the forbidden gap along the imaginary k axis and making a smooth transition from one band into the other band

[4.14]. If the potential varies slowly with distance the WKB (Wentzel-Kramers-Brillouin) approximation can be used to find the solution of the electron wave function and further the tunneling probability through the tunneling barrier as shown in (1).

$$T = \exp(-2i \int_{x1}^{x2} K dx) \tag{4.1}$$

K is the imaginary wave vector (also called dispersion relation) and x1 and x2 are the starting and ending points of the tunneling path. The simplest dispersion relation is a parabolic one band relation [4.14] obtained by conservation of energy of the carrier tunneling from the valence band to the conduction band. This dispersion relation however is most appropriate only when the carrier is close to the conduction band edge or the valence band edge. For a more accurate dispersion relation that is applicable to two band semiconductors and can take into account the unequal effective masses near the conduction and valence band edges, the Franz Two Band dispersion relation (2) can be used.

$$K = \frac{K_c K_v}{\sqrt{K_c^2 + K_v^2}} \tag{4.2}$$

Where

$$K_c(x,\varepsilon) = \frac{\sqrt{2.m_c(x).|E_c(x) - \varepsilon|}}{h} \tag{4.3}$$

$$K_v(x,\varepsilon) = \frac{\sqrt{2.m_v(x).|\varepsilon - E_v(x)|}}{h} \tag{4.4}$$

The concept of an effective bandgap of a heterostructure when considering carriers tunneling across a heterostructure has been explored previously in the context of heterostructure MOSFET GIDL current [4.15]. Here this concept is revisited in the context of heterostructure TFETs and their performance.

Assume a TFET with a thin biaxially strained $Si_{1-x}Gex$ layer on pure Si and the dielectric forms an interface with the strained $Si_{1-x}Ge_x$ layer. The conduction band and valence band energies of biaxially compressively strained $Si_{1-x}Ge_x$ on Si are higher than that of the Si substrate leading to a band alignment as seen in Figure 4.1. The degree of conduction and valence band offsets (ΔE_c and ΔE_v) with respect to the Si substrate is determined by the amount of strain (% Ge in the top $Si_{1-x}Ge_x$ layer). The effective bandgap (**$E_g = E_{g,SiGe} - \Delta Ec$)** at the heterostructure interface is also indicated in Figure 4.1. Clearly a larger ΔE_c leads to a smaller effective tunneling-bandgap.

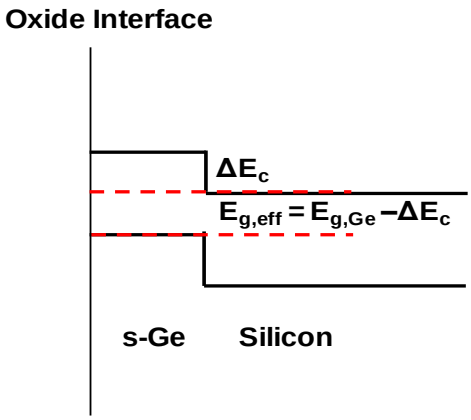


Figure 4.1 Band alignment for strained $Si_{1-x}Ge_x$ cap on Si substrate. Biaxial compressive strain lifts the conduction and valence bands of $Si_{1-x}Ge_x$ above Si. The conduction and valence band offsets are determined by the amount of strain (% Ge in the $Si_{1-x}Ge_x$ layer)

Now assume that the ΔE_c = 0.2eV for a pure Ge top layer on Si substrate. Figure 4.2 is a plot of the imaginary wave vector as it varies along the tunneling path for three different materials. The plot compares the imaginary wave vectors for a tunneling path entirely in Ge (E_g = 0.67eV), a path involving tunneling across a heterostructure of strained Ge on Si (tunneling from the valence band of $Si_{1-x}Ge_x$ to the conduction band of Si) and finally a path involving tunneling entirely in a material with $E_g = E_{g,Ge} - \Delta E_c$ = 0.47eV.

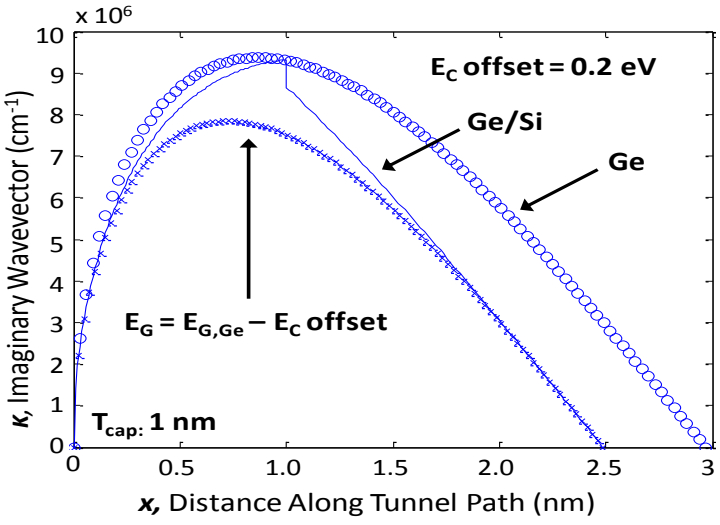


Figure 4.2 Imaginary wave vector as a function of the tunneling path for (a) tunneling path entirely in Ge, E_g = 0.67eV (b) path involving tunneling across a heterostructure of strained Ge on Si (c) path involving tunneling entirely in a material with $E_g = E_{g,Ge} - \Delta E_c$ = 0.47eV

From Figure 4.2 the dispersion relation for tunneling across the heterostructure approaches the dispersion relation for tunneling entirely in a material with $\mathbf{E_g = E_{g,Ge} - \Delta E_c}$. This further implies that the tunneling probability across the heterostructure approaches that of a material $\mathbf{E_g = E_{g,Ge} - \Delta E_c}$ and confirms the concept of an effective tunneling bandgap when considering tunneling across a heterostructure. The larger the $\mathbf{\Delta E_c}$ induced by strain, the smaller the effective tunneling bandgap and the larger the tunneling probability. The concept of using band offsets induced by biaxial strain to engineer ultra low effective bandgap heterostructures is exploited to design TFETs with higher I_{on}/I_{off} at 0.4V V_{dd} than would be possible with ultra low bandgap semiconductors.

4.3 BIAXIALLY STRAINED HETERO TFET STRUCTURES

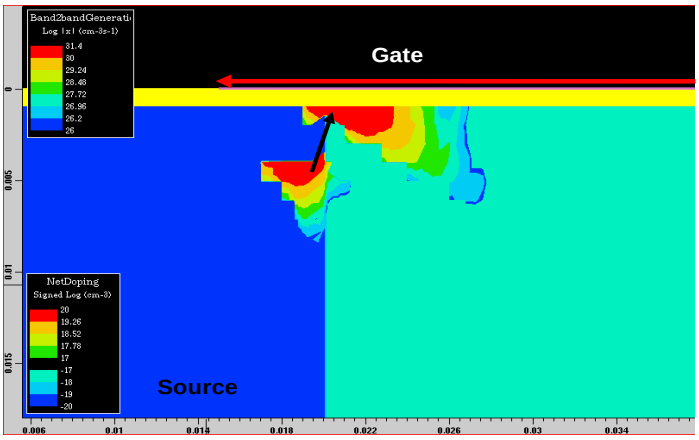


Figure 4.3 The TCAD output with contours of BTBT generation rate of electrons and holes in a PIN TFET. The red regions indicate the peak generation rate and blue indicates the lowest generation rate. The tunneling originates within the heavily P doped source and occurs at an angle, along the direction of the resultant of the vertical and lateral E fields.

The simplest, most widely researched planar single gate TFET design is that of a gated diode (PIN TFET). Figure 4.3 shows the TCAD output with contours of BTBT generation rate of electrons and holes in such a TFET. The tunneling originates within the heavily doped source and occurs at an angle, along the direction of the resultant of the vertical and lateral E fields. Previous heterostructure based TFET designs [4.16] have explored heterostructures at the source channel interface perpendicular to the dielectric interface. In such a design tunneling originates in the source, occurs at an angle across the source channel heterostructure and is limited to a few nanometers below the dielectric interface.

In this work we explore the design and optimization of planar single gate TFETs with an ultra low effective bandgap heterostructure parallel to the dielectric interface (Figure 4.4a). A thin biaxially strained film is inserted between the substrate and the dielectric, parallel to the dielectric interface and BTBT is induced across the heterostructure in the source. In this design the tunneling area and I_{on} can be tuned by varying the length of the source overlapped by the gate. Also this design is easier to integrate into a standard CMOS process as it only requires the growth of a thin biaxially strained capping layer on the substrate before depositing the dielectric and gate electrode. Previously explored techniques such as a thin dopant pocket [4.8] or a sheet of fixed charge at the dielectric interface an also be easily integrated into this design to further enhance the vertical E field and tunneling current at the onset of BTBT.

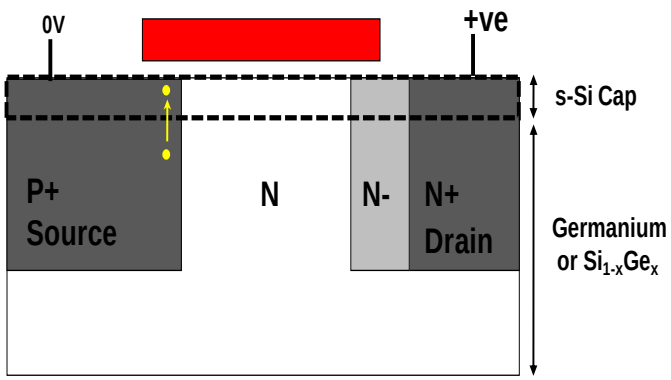


Figure 4.4a N-type hetero TFETs with a P+ source and a biaxial strained capping layer between the substrate end gate dielectric. A positive gate bias pulls the bands down in the source and electrons tunnel from the valence band of the substrate to the conduction band of the biaxially strained capping layer.

In N-type hetero TFETs (Figure 4.4a) with a P+ source, a positive gate bias pulls the bands down in the source and electrons tunnel from the valence band of the substrate to the conduction band of the biaxially strained capping layer. To engineer an ultra low effective tunneling bandgap, the $E_{c,cap}$ needs to be below the $E_{c,subs}$ (Figure 4.4b) The holes generated in the substrate flow to the source terminal and the electrons generated in the cap are collected by the reverse biased N doped drain. In P-type hetero TFETs with a N+ source, a negative gate bias pulls the bands up in the source and electrons tunnel from the valence band of the biaxially strained capping layer to the conduction band of the substrate. To engineer an ultra low effective tunneling bandgap for P-type hetero TFETs, the $E_{c,subs}$ needs to be below the $E_{c,cap}$ (Figure 4.4b). The electrons generated in the substrate flow to the source terminal and the holes generated in the cap are collected by the reverse biased P doped drain.

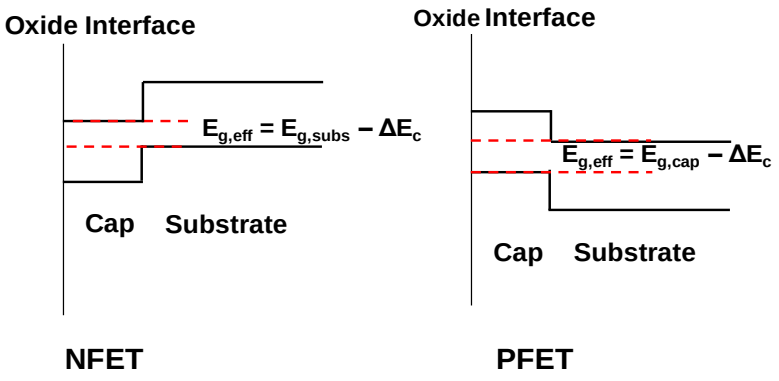


Figure 4.4b The heterostructure band alignment of the strained cap and substrate in the source of NFETs and PFETs. In NFETs a positive gate bias causes the electrons to tunnel from the valence band of the substrate to the conduction band of the strained cap. In PFETs a negative gate bias causes the electrons to tunneling from the valence band of the cap to the conduction band of the substrate.

4.4 Engineering Symmetric Effective Tunneling Bandgaps For Complimentary TFETs

The lattice constant of Si is smaller than that of Ge and the mismatch is 4.2%. When matching the lattice constant (parallel to the interface plane) to that of the unstrained substrate material, the lattice constant perpendicular to the interface also changes (as a consequence of Poisson's ratio) to compensate for the lateral strain in the active material. For example a Si layer grown pseudomorphically on a $Si_{1-x}Ge_x$ substrate is biaxially strained. In pseudomorphically strained layers (1) the hydrostatic strain shifts the energetic position of a band and (2) the uniaxial strain component splits degenerate bands [4.13].

The strain in the pseudomorphic active layer includes a hydrostatic component which shifts the average band energy level and a uniaxial component which splits the degenerate bands [4.13]. The hydrostatic tensile strain raises the average energy levels of the conduction band and valence band, and the uniaxial component splits degenerate bands. Uniaxial stress causes lowering of the two-fold degenerate [001] ellipsoids (Δ_2) and raises the four-fold degenerate [100] [010] ellipsoids (Δ_4) from the six-fold degenerate system. The higher the Ge percentage of the substrate the larger the tensile strain in the Si layer. It is the splitting of the degenerate bands that reduces the inter-valley scattering and enhances carrier mobility.

In the valence band, strain causes the splitting of degenerate valence bands at the Γ point. The strain also shifts the LH and HH up for compressive strain and moves the LH band up and the HH band down for tensile strain (Figure 4.5a) [4.17]. For tensile strain when the Ge content in the substrate is greater than 20% the hole mass of the LH band becomes larger than the hole mass of the HH band (Figure 4.5b) [4.17]. At low Ge content below 20%, both LH and HH bands contribute to hole transport but beyond 20% Ge in the substrate the splitting between LH and HH band increases and reduces the occupation of the HH band.

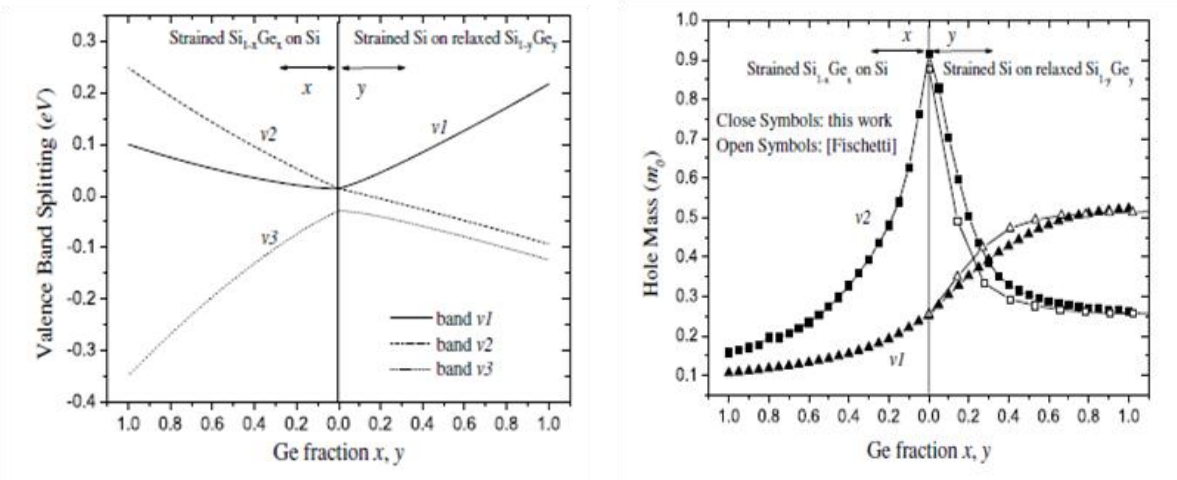


Figure 4.5a Valence band splitting for biaxial tensile and compressive strain [ ] 4.5b Calculated hole effective mass in the bands v1 and v2 from k.p compared with data [4.17].

References	ΔE_C	ΔE_V	$E_{g,sSi}$	$E_{g,eff}$
Van de Walle	0.55	0.31	0.42	0.12
[1] Rieger, et al.	0.56	0.25	0.35	0.11
Yang, et al.	0.58	0.21	0.37	0.09

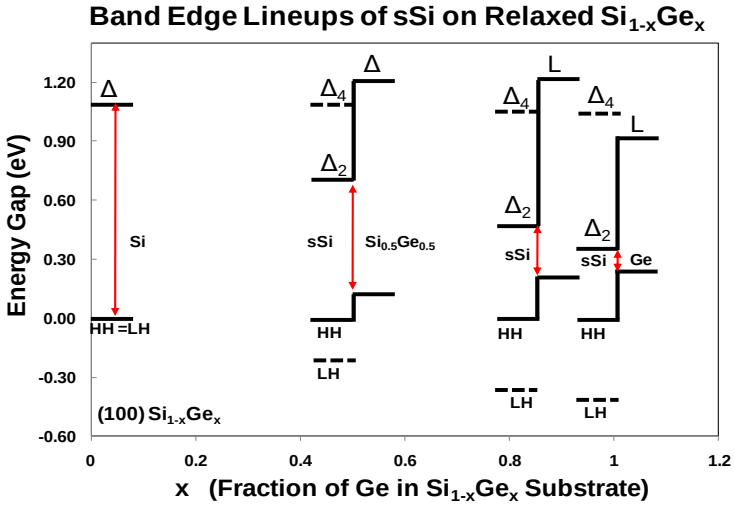


Figure 4.6 The band lineups for biaxial tensile strained Si on a $Si_{1-x}Ge_x$ substrate as a function of the Ge percentage of the substrate [4.18]. Strained Si on pure Ge (100) forms a heterostructure with ultra low effective bandgap (0.12 eV) and $E_{c,s\text{-}Si}$ is below $E_{c,Ge}$ thus providing an easy to integrate solution for N-type TFETs with a s-Si cap layer on a Ge substrate.

Figure 4.6 displays the band lineups for biaxial tensile strained Si on a $Si_{1-x}Ge_x$ substrate as a function of the Ge percentage of the substrate [4.18]. For $Si_{1-x}Ge_x$ with greater than 80% Ge in the substrate the E_c is formed by the L minima and not Δ minima as shown in Figure 4.6. From figure 4.6 we clearly see that strained Si on pure Ge (100) forms a heterostructure with ultra low effective bandgap (0.12 eV) and $E_{c,s\text{-}Si}$ is below $E_{c,Ge}$ thus providing an easy to integrate solution for N-type TFETs with a s-Si cap layer on a Ge substrate.

References	ΔE_C	ΔE_V	$E_{g,sGe}$	$E_{g,eff}$
Van de Walle	0.28	0.84	0.56	0.28
[1] Rieger, et al.	0.28	0.74	0.6	0.32
Yang, et al.	0.30	0.71	0.71	0.41

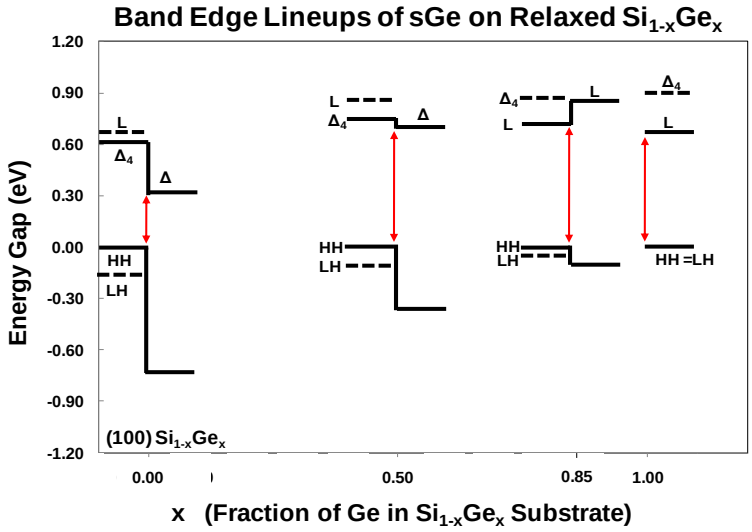


Figure 4.7 The band lineups for biaxial compressive strained Ge on a relaxed $Si_{1-x}Ge_x$ substrate as a function of the Ge percentage of the substrate [4.18]. Lower the Ge % of the substrate greater the compressive strain of the Ge layer. Strained Ge on pure Se (100) forms a heterostructure with ultra low effective bandgap (0.3 eV) and $E_{c,s\text{-}Ge}$ is above $E_{c,Si}$.

For $Si_{1-x}Ge_x$ under compressive strain, the mass in HH band is larger than that in LH band. Since for all Ge % the HH band is raised above the LH band by splitting, in compressively strained $Si_{1-x}Ge_x$, the HH band forms the valence band edge [4.18]. A layer of Ge (100%) grown on a Si substrate is biaxial compressively strained because of the smaller lattice constant of the Si substrate. Figure 4.7 displays the band lineups for biaxial compressive strained Ge on a relaxed

$Si_{1-x}Ge_x$ substrate as a function of the Ge percentage of the substrate [4.18]. Lower the Ge % of the substrate greater the compressive strain of the Ge layer. From Fig we clearly see that strained Ge on pure Se (100) forms a heterostructure with ultra low effective bandgap (0.3 eV) and $E_{c,s\text{-}Ge}$ is above $E_{c,Si}$. This makes a s-Ge cap on Si substrate an easy to integrate solution for P-type TFETs. The larger effective bandgap across the heterostructure of the P-type TFET makes the PFET weaker than the NFET.

Using a s-Ge cap on Si substrate, the PFET performance can be enhanced to match the NFET performance with the use of an oppositely doped pocket region in the s-Ge cap layer. The other way to achieve complimentary hetero TFETs with similar performance would be to find heterostructures with symmetric effective bandgaps for N and P type TFETs. Turns out symmetric effective bandgaps can be achieved using biaxial strain engineering as shown in Figure 4.8. The following 3 are examples of symmetric effective bandgaps and the first example provides the smallest symmetric effective bandgap:

(a) Start with a relaxed $Si_{0.6}Ge_{0.4}$ substrate, a s-Si cap layer on relaxed Ge for NFET and a s-Ge cap layer on s-Si for PFET form symmetric effective bandgaps (0.12eV) for complimentary TFETs.

(b) Start with a relaxed $Si_{0.5}Ge_{0.5}$ substrate, a s-Si cap layer on s- Ge for NFET and a s-Ge cap layer on s-Si for PFET form symmetric effective bandgaps (0.2eV) for complimentary TFETs.

(c) Start with a relaxed $Si_{0.15}Ge_{0.85}$ substrate, a s-Si cap layer on s- Ge for NFET and a s-Ge cap layer on s-Si for PFET form symmetric effective bandgaps (0.15eV) for complimentary TFETs.

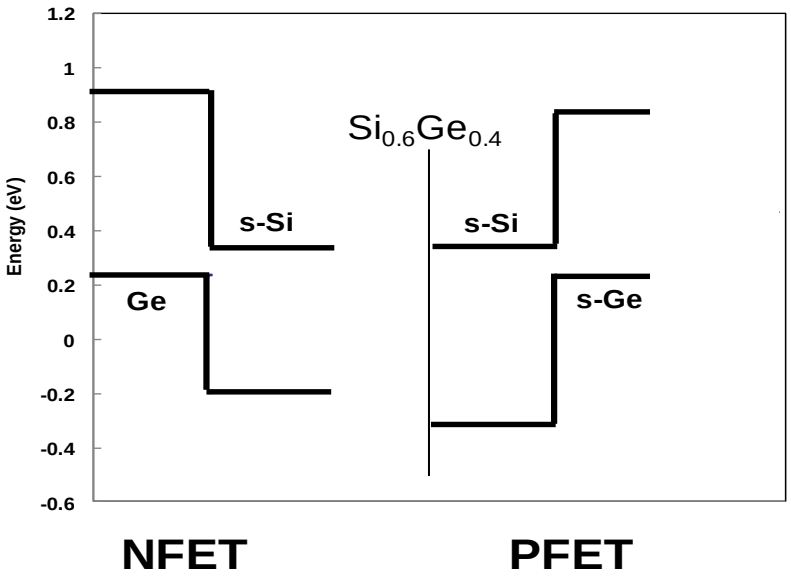


Figure 4.8a Starting with a relaxed $Si_{0.6}Ge_{0.4}$ substrate, a strained-Si cap layer on relaxed Ge for NFET and a strained-Ge cap layer on strained-Si for PFET form symmetric effective bandgaps (0.12ev) for complimentary TFETs.

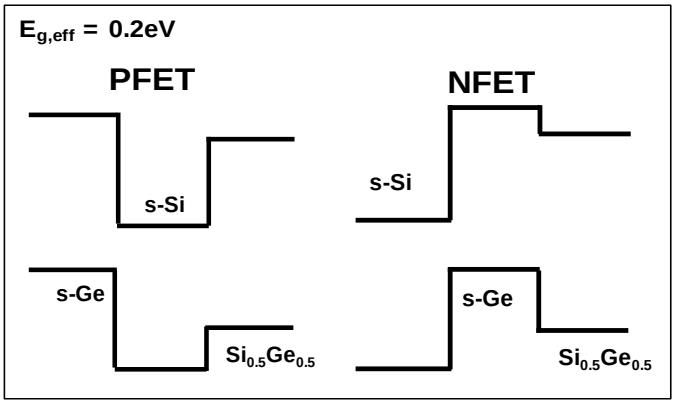


Figure 4.8b Starting with a relaxed $Si_{0.5}Ge_{0.5}$ substrate, a strained-Si cap layer on strained-Ge for NFET and a strained-Ge cap layer on strained-Si for PFET form symmetric effective bandgaps (0.2eV) for complimentary TFETs.

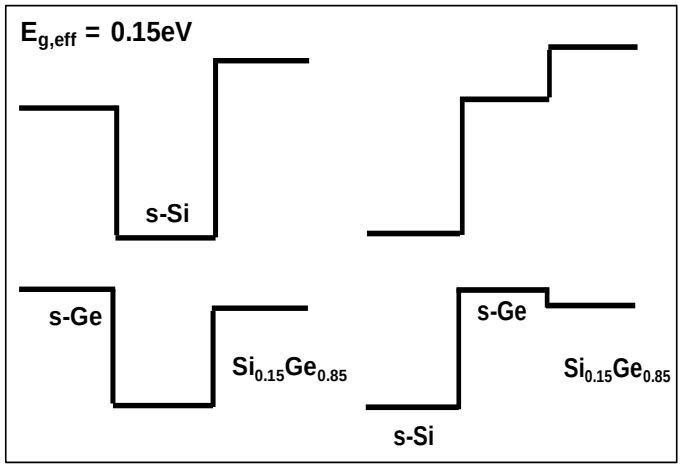
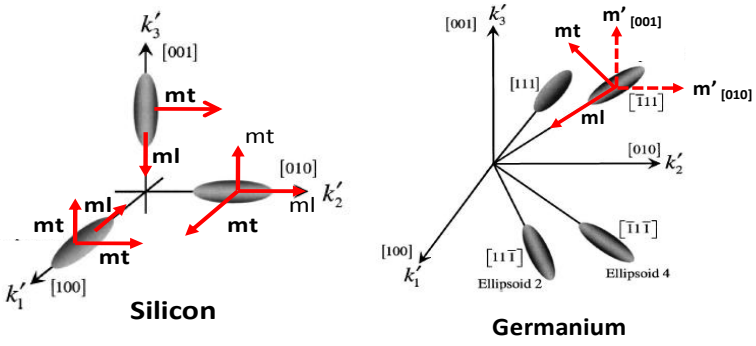


Figure 4.8c Starting with a relaxed $Si_{0.15}Ge_{0.85}$ substrate, a strained-Si cap layer on strained-Ge for NFET and a strained-Ge cap layer on strained-Si for PFET form symmetric effective bandgaps (0.15eV) for complimentary TFETs.

While theoretical studies predicted the following valence band offsets of pure Ge on Si, 0.84 eV [4.19], 0.93 eV [4.20] and 0.74 eV [4.21] and the following valence band offsets of pure Si on Ge, 0.31 eV [4.19], 0.36 eV [4.20] and 0.21 eV [4.21] for pure Si on Ge. These values may be compared to the experimental values for pure Ge on Si, 0.74 ± 0.13 eV [4.22], 0.83 ± 0.11 eV [4.23] and the values for pure Si on Ge 0.17 ± 0.13 eV [4.22] and 0.22 ± 0.13 eV [4.23].

4.5 HETEROSTRUCTURE TFET OPTIMIZATION

The SENTAURUS TCAD simulator is well equipped to do non-local dynamic BTBT calculations across heterostructures. In SENTAURUS the tunneling parameters A and B of indirect band semiconductor materials are defined as a function of the electron and hole effective masses and the bandgap of the semiconductor. The most accurate way to extract A and B for a material would be by fitting to BTBT data from two terminal tunnel diodes, thus eliminating any impact of dielectric interface states that could impact BTBT data from MOSFET GIDL current. While A and B values obtained by fitting to Si and Ge experimental data are available in literature, values for strained Si and Ge and other $Si_{1-x}Ge_x$ alloys are not readily available in literature. Accurate values of bandgap and the electron and hole effective masses of strained and unstrained $Si_{1-x}Ge_x$ layers of various composition are however readily available in literature. In order to make an accurate comparison across the entire hetero TFET design space even without experimentally calibrated A and B values, in this work, based on the direction of tunneling, the A and B values for Si, Ge, strained Si, strained Ge and other $Si_{1-x}Ge_x$ layers were calculated using the relevant effective masses and bandgap from literature. Since the computed A and B values for unstrained Si and unstrained Ge were smaller than those obtained by experimental fitting a scaling factor was determined for both materials to scale the computed values to match the experimentally fitted values. The computed values for strained Si and strained Ge were then scaled by the respective Si and Ge scaling factors. While the exact value of I_{on} from these simulations can only be verified experimentally in the future, this method allows for a fair study of the design and optimization of hetero TFETs with biaxially strained capping layers. With this method it is also possible to compare hetero TFETs with biaxial strained cap layers to homojunction Ge and SiGe TFETs giving us a good indication of the relative benefit of this hetero TFET design.



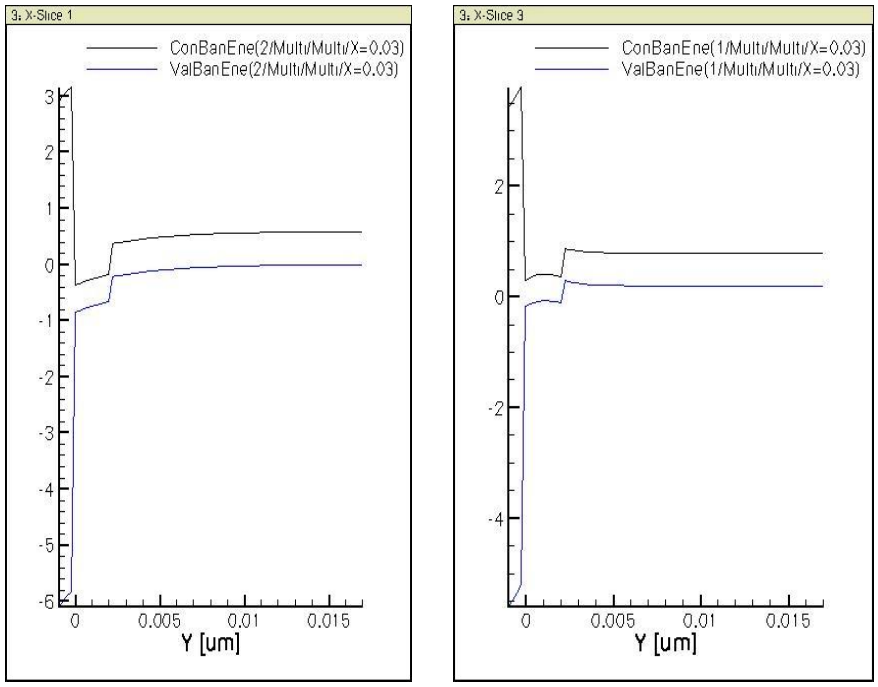
Tunneling Direction	Germanium				Silicon			
	mc	mv	Eg	X	mc	mv	Eg	X
Vertical [001]	$m'_{[001]}$	m_{lh}	0.67	3.95	m_l	m_{hh}	0.53	4.45
Lateral [010]	$m'_{[010]}$	m_{lh}	0.67	3.95	m_t	m_{hh}	0.53	4.45

Fig. 4.9 For a N-type TFET with s-Si cap on Ge (100) substrate with channel along [010], the relevant material parameters used to calculate A and B for strained Si and unstrained Ge for vertical tunneling [001] direction and lateral tunneling [010] direction are shown in the table.

Figure 4.9 uses the example of a N-type hetero TFET with a strained Si cap on a Ge (100) substrate with channel along [010]. The relevant material parameters used to calculate A and B for strained Si and unstrained Ge for vertical tunneling in the [001] direction are shown in the table. In tensile strained Si the Δ_2 conduction band minima move to lower energy and form the conduction band edge while the Δ_4 minima are 0.4eV above E_c (Figure 4.6).

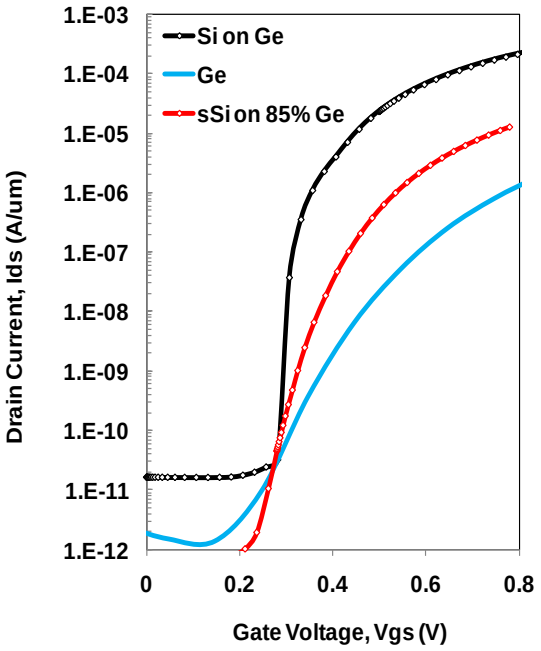
The dominant tunneling in this case corresponds to the 0.53eV (Δ_2) bandgap and longitudinal strained Si electron mass. Unstrained Ge has conduction minima along (111) and so the relevant electron effective mass is the component of the Ge electron effective mass along [010]. The table also shows material parameters that would be relevant to tunneling along [010] incase the heterostructure was perpendicular to the dielectric interface.

When the P^+ source (cap and substrate) is doped to 1e20 cm^{-3} the band lineup in the source as shown in Figure 4.10 does not benefit from the low effective bandgap of the s-Si on Ge heterostructure. Tunneling occurs across the heterostructure at an angle from the P^+ source into the lightly doped s-Si channel region. When the source is doped to 1e19 cm^{-3} however the band lineup in the source is optimal to benefit from the reduced effective bandgap of the heterostructure and allows for vertical tunneling across the heterostructure at higher gate voltages. For all simulations from here on the source doping is maintained at 1e19 cm^{-3}.



1e19 cm^{-3} P+ Source 1e20 cm^{-3} P+ Source

Figure 4.10 Band lineup in the source when the P^+ source (cap and substrate) is doped to 1e20 cm^{-3} does not make use of the low effective bandgap of the s-Si on Ge heterostructure. Band lineup in the source when the P+ source is doped to 1e19 cm^{-3} is optimal to benefit from the reduced effective bandgap of the heterostructure and allows for vertical tunneling across the heterostructure.



Parameter	Optimal
Nsource	1e19 cm^{-3}
Source Underlap	6nm
Cap Thickness	2nm
Oxide Thickness	1nm
Gate Length	30nm
Body Doping	1e16 cm^{-3}
Drain Doping	1e20 cm^{-3}

Figure 4.11 Comparison of a pure Ge N-type TFET, a TFET with a s-Si cap layer on pure Ge (100) and a TFET with a s-Si layer on $Si_{15}Ge_{85}$ (100). The s-Si on pure Ge TFET gives 100x more current than the pure Ge TFET and 10x more than the s-Si on $Si_{0.15}Ge_{0.85}$ TFET. The various device parameters used in all the three simulations are presented in the table.

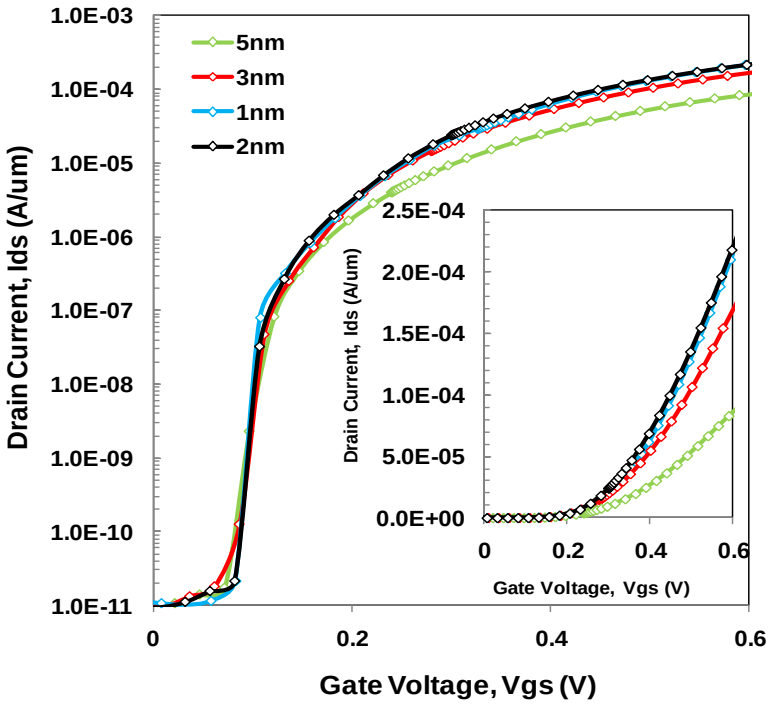


Figure 4.12a Impact of cap thickness on hetero TFET performance assuming WF engineering can achieve identical turn on voltage in all cases. 2nm Cap gives highest I_{on} for the same gate overdrive.

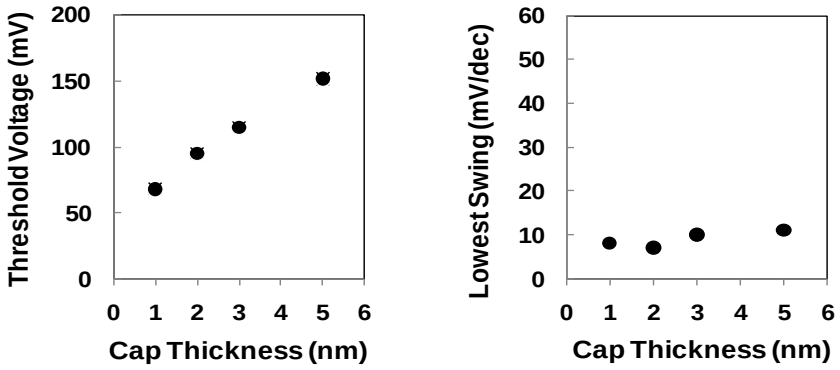


Figure 4.12b Assuming identical gate WF for all cap thicknesses plot of turn on voltage vs. cap thickness and steepest swing vs. cap thickness for N TFET with s-Si on Ge (100) substrate

Figure 4.11 compares a pure Ge N-type TFET, a TFET with a strained Si cap layer on pure Ge (100) and a TFET with a strained Si layer on $Si_{15}Ge_{85}$ (100). The strained Si on pure Ge TFET gives 100x more current than the pure Ge TFET and 10x more than the s-Si on $Si_{0.15}Ge_{0.85}$ TFET confirming the concept of enhanced BTBT across a low effective bandgap heterostructure. The various device parameters used in all the three simulations are presented in the table.

Figure 4.12a explores the impact of cap thickness on hetero TFET performance assuming WF engineering can be used to achieve identical turn on voltage in all cases. Using an identical gate WF, the turn on voltage increases with cap thickness as seen in Figure 4.12b. From Figure 4.12a and figure 4.12b we see that the 2nm cap thickness is ideal in terms of the both the steepest swing as well as highest I_{on}. From the TCAD BTBT generation rate output a 2nm cap is sufficient to capture the peak tunneling volume. With thicker caps, the gate coupling to the Ge source decreases degrading I_{on}.

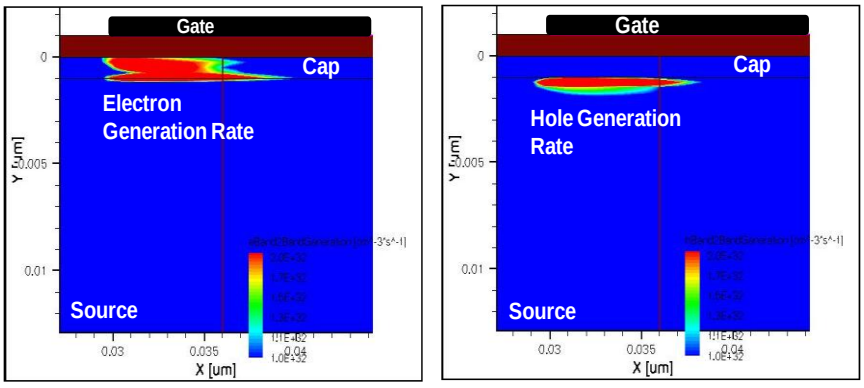


Figure 4.13 The BTBT generation rate of electrons and hole for a 1nm s-Si cap on pure Ge (100) substrate. Electrons tunnel vertically across the heterostructure.

For a given cap thickness, the length of the source overlapped by the gate can be tuned to control the tunneling area. Figure 4.13 shows the TCAD output of BTBT generated electrons and holes for a 1nm s-Si cap on Ge hetero TFET with 6nm of the source overlapped by the gate. Since vertical tunneling across the heterostructure occurs increasing the length of the source overlapped by the gate increases the Ion. The maximum overlap with negligible impact on I_{off} needs to be determined since increasing overlap decreases the distance between the source and drain. Figure 4.14 shows the impact of increasing the source overlap from 0nm to 18nm for a 30nm gate length device. The heavily doped drain is spaced 20nm away from the gate edge to maintain a low I_{off}.

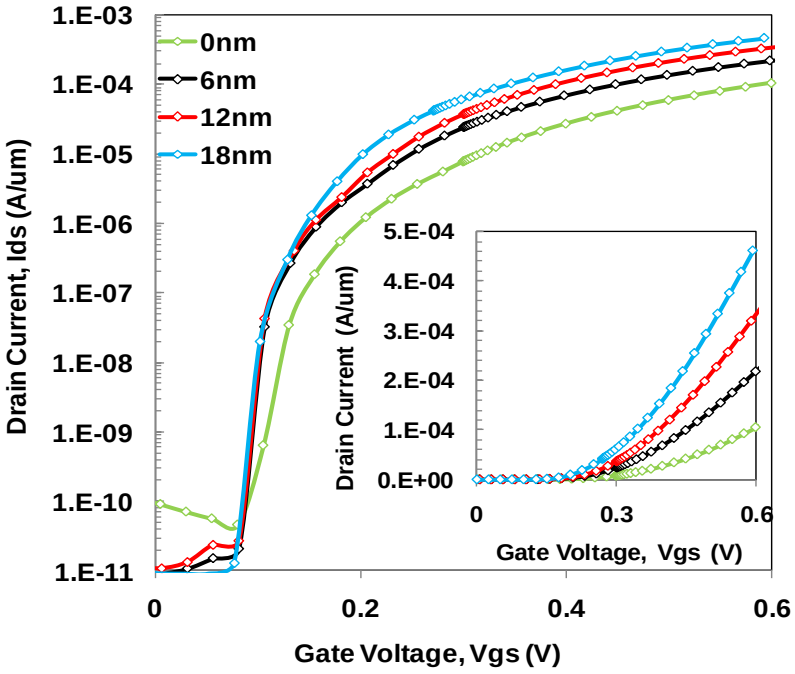


Figure 4.14 Impact of length of the source overlapped by the gate. Larger overlap enhances tunneling area and provides larger I_{on}.

Figure 4.15 shows the I_d-V_g for the most optimized N-type hetero TFET and the dotted line indicates the ideal MOSFET with 60mV/decade turn on and the same I_{off} as the hetero TFET. Below a gate overdrive of 0.4V, for the same I_{off} the hetero TFET performance exceeds that of the most ideal MOSFET. The hetero TFET at 0.4V overdrive provides $I_{on}/I_{off} = 4x10^7$ making it possible to scale V_{dd} below 0.4V and still achieve good performance.

Figure 4.16 shows the I_d-V_d for the most optimized N-type hetero TFET. It shows slight non linearity at low V_{ds} and saturates at high V_{ds}.

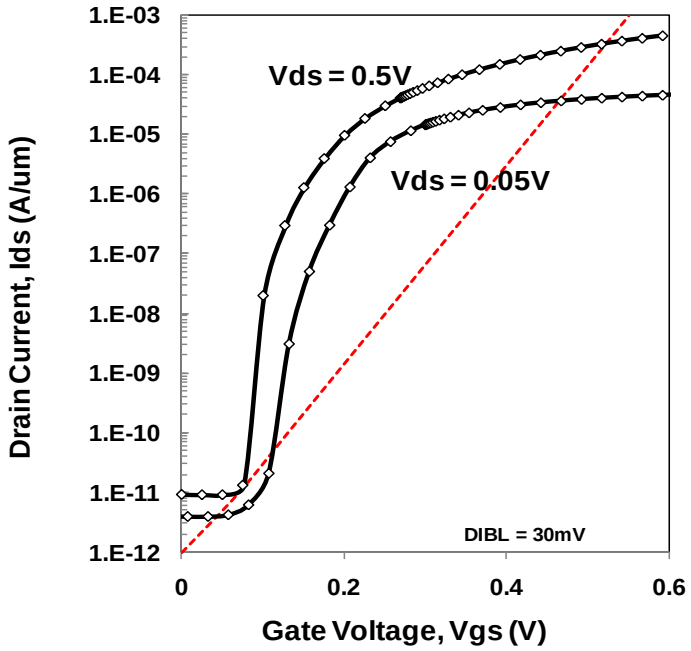


Figure 4.15 I_d-V_g characteristics for optimized 30nm gate length N hetero TFET with s-Si on Ge (100). The dotted line indicates the ideal MOSFET with 60mV/decade turn on. Below a gate overdrive of 0.4V, for the same I_{off} the hetero TFET performance exceeds that of the ideal MOSFET. The hetero TFET at 0.4V overdrive provides I_{on}/I_{off} = 4x10^7 making it possible to scale V_{dd} below 0.4V and still achieve good performance.

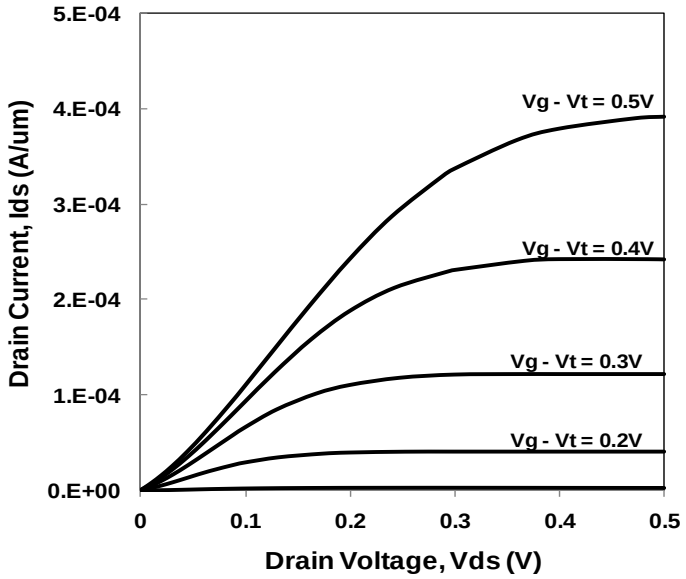


Figure 4.16 I_d-V_d characteristics for a 30nm gate length N hetero TFET with s-Si on Ge (100) exhibit non linearity at low V_{ds} and saturation at high V_{ds}.

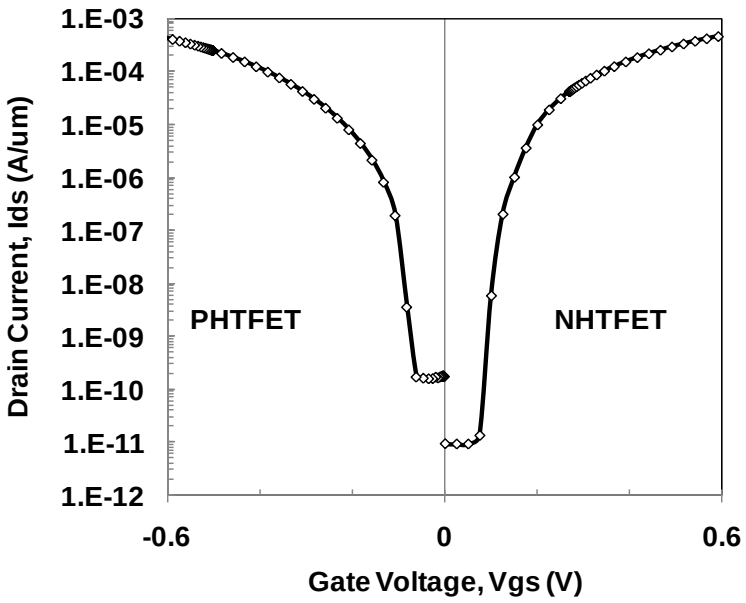


Figure 4.17 The I_d-V_g of the most optimized N-type and P-type hetero TFETs. The ratio of I_{on} for NFET to PFET is 1.2

Figure 4.17 compares the I_d-V_g of the most optimized N-type and P-type hetero TFETs. The NFET uses a strained-Silicon cap on unstrained Ge while the PFET uses a strained-Germanium cap on s-Si on $Si_{0.6}Ge_{0.4}$. The ratio of I_{on} for NFET to PFET is 1.2

4.6 CONCLUSIONS

Prototype LPTFETs are fabricated in silicon and shown to have enhanced current and swing compared to PIN TFETs. Enhancement in swing and on current due to lateral pocket opens up a path to lower TFET operating voltage which can be combined with bandgap scaling to achieve high I_{on} and high I_{on}/I_{off} at ultra low V_{dd}. Further performance enhancement than demonstrated here can be achieved with better control of the source and lateral pocket profiles and thinner gate dielectric. If implantation is used, flash annealing or other diffusion-less annealing methods can help realize more abrupt source-pocket junctions than possible by RTA. Selective epitaxial growth of source and pocket regions would be ideal to provide maximum enhancement in LPTFET performance and enable V_{dd} scaling.

4.7 REFERENCES

[4.1] W. M. Reddick, G. A. Amaratunga, "Silicon surface tunnel transistor," Applied Physics Letters, vol.67, no.4, pp.494-496, Jul 1995

[4.2] C. Aydin, A. Zaslavsky, S. Luryi, S. Cristoloveanu, D. Mariolle, D. Fraboulet, S. Deleonibus, "Lateral interband tunneling transistor in silicon-on-insulator," Applied Physics Letters , vol.84, no.10, pp.1780-1782, Mar 2004

[4.3] Krishna K. Bhuwalka, Mathias Born, Markus Schindler, Matthias Schmidt, Torsten Sulima and Ignaz Eisele, "P-Channel Tunnel Field-Effect Transistors down to Sub-50 nm Channel Lengths", Japanese Journal of Applied Physics,45, pp.3106-3109, 2006

[4.4] W. Y. Choi, B. G. Park, J. D. Lee and T. K. Liu, "Tunneling Field-Effect Transistors with Subthreshold Swing Less Than 60mV/dec," IEEE Electron Device Letters, Vol. 28, No. 8, Aug 2007, pp.743-745.

[4.5] F. Mayer, C. L. Royer, J. F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali and S. Deleonibus, "Impact of SOI, $Si_{1-x}Ge_xOI$ and GeOI substrates on CMOS compatible Tunnel FET performance," IEDM 2008, pp.163

[4.6] S. H. Kim, H. Kam, C. Hu and T. J. K. Liu, "Germanium-Source Tunnel Field Effect Transistors with Record High ION/IOFF", Symposium on VLSI Technology, 2009, pp.178.

[4.7] C. Hu, D. Chou, P. Patel, A. Bowonder, "Green Transistor – A V_{DD} Scaling Path for Future Low Power ICs", VLSI Technology, Systems and Applications, 2008. VLSI-TSA 2008. International Symposium on. April 2008, pp.14-15.

[4.8] P. Patel, K. Jeon, A. Bowonder, C. Hu, "A Low Voltage Steep Turn-off Tunnel Transistor Design," SISPAD Sept 2009.

[4.9] A. Bowonder, P. Patel, K. Jeon, J. Oh, P. Majhi, H. H. Tseng, C. Hu, "Low Voltage Green Transistor Using Hetero-Tunneling", International Workshop on Junction Technology, May, 2008, pp.93-96

[4.10] V. Nagavarapu, R. Jhaveri and J. C. S. Woo, "The Tunnel Source (PNPN) n-MOSFET: A Novel High Performance Transistor," IEEE Transaction on Electron Devices, Vol. 55, No.4, April 2008, pp.1013-1019.

[4.11] M. L. Lee, E. A. Fitzgerald, M. T. Bulsara, M. T. Currie, and A. Lochtefeld, "Strained Si, SiGe, and Ge channels for high-mobility metal-oxide-semiconductor field-effect transistors", J. Appl. Phys. **97**, 011101 (2005)

[4.12] J. L. Hoyt, H. M. Nayfeh, S. Eguchi, I. Aberg, G. Xia, T. Drake, E. A. Fitzgerald, A. D. Antoniadis, "Strained silicon MOSFET technology," Electron Devices Meeting, 2002. IEDM '02. Digest. International, vol., no., pp. 23- 26, 2002

[4.13] Y. Sun, S. E. Thompson and T. Nishida, "Physics of strain effects in semiconductors and metal-oxide-semiconductor field-effect transistors", J. Appl. Phys. 101, 104503 (2007)

[4.14] S. Wang, Fundamentals of Semiconductor Theory and Device Physics, 1989

[4.15] T. Krishnamohan, Z. Krivokapic, K. Uchida, Y. Nishi, and K. C. Saraswat, "Low defect ultrathin fully strained Ge MOSFET on relaxed Si with high mobility and low band-to-band-tunneling (BTBT)," in VLSI Symp. Tech. Dig., 2005, pp. 82–83.

[4.16] O. M. Nayfeh, C. N. Chleirigh, J. Hennessy, L. Gomez, J. L. Hoyt, D. A. Antoniadis, "Design of Tunneling Field-Effect Transistors Using Strained-Silicon/Strained-Germanium Type-II Staggered Heterojunctions," Electron Device Letters, IEEE , vol.29, no.9, pp.1074-1077, Sept. 2008
[4.17] Lianfeng Yang, Jeremy R. Watling, Richard C. W. Wilkins, Mirela Boric, John R. Barker, Asen Asenov and Scott Roy, "Si/SiGe heterostructure parameters for device simulations", Semiconductor Science and Technology, 19**,** 1174–1182, 2004.
[4.18] M. Rieger and P. Vogl,"Electronic-band parameters in strained $Si_{1-x}Ge_x$ alloys on $Si_{1-y}Ge_y$ substrates", Physical Review B, Vol. 48, 19, 14276–87 [10], 1993
[4.19] C. G. Van de Walle and R. M. Martin, "Theoretical calculations of heterojunction discontinuities in the Si/Ge system", Physical Review. B, 34**,** 5621–34, 1986
[4.20] Walle C G Van de and Martin R M, J. Vac. Sci. Technology.B, 4 **,** 1055–9, 1986
[4.21] Colombo L, Resta R and Baroni S, "Valence-band offsets at strained Si/Ge interfaces", Physical Review. B, 44, 5572–9, 1991
[4.22] G. P. Schwartz, M. S. Hybertsen, J. Bevk, R. G. Nuzzo, J. P. Mannaerts, G. J. Gualtieri," Core-level photoemission measurements of valence-band offsets in highly strained heterojunctions: Si-Ge system", Physical Review. B, 39**,** 1235–41, 1989
[4.23] E. T. Yu, E. T. Croke, D. H. Chow, D. A. Collins, M. C. Phillips, T. C. McGill and R. H. Miles, "Measurement of the valence band offset in novel heterojunction systems: Si/Ge (100) and AlSb/ZnTe (100)", J. Vac. Sci. Technology. B, 8 908–15, 1990

Chapter 5

Heterostructure MOSFETs and Verification of Hetero Band to Band Tunneling Concept

5.1 INTRODUCTION

The previous chapter explored the use of biaxial strain engineering to design high performance silicon-germanium TFETs for sub 0.5V operation. In this chapter, electrical measurements from biaxial strained heterostructure MOSFETs are explored as preliminary verification of enhanced tunneling across heterostructures with reduced effective bandgap. Biaxial strain has been explored greatly for enhanced mobility in MOSFETs [5.1, 5.2]. A relaxed $Si_{1-x}Ge_x$ buffer on a silicon substrate helps create a larger lattice constant. This epitaxially grown buffer layer transforms the silicon substrate so the lattice constant of the top surface is that of the $Si_{1-x}Ge_x$ layer. The buffer layer can then be used as a template on which epitaxial Si-rich layers are biaxially tensile strained and epitaxial Ge-rich layers are biaxially compressive strained [5.1].

The earliest attempts of biaxial strain involved a thick uniform $Si_{1-x}Ge_x$ buffer layer, but these were not fully relaxed and contained threading dislocation densities ~ $10^8 cm^{-2}$ [5.1]. Complete relaxation was achieved with relaxed $Si_{1-x}Ge_x$ graded buffers. These graded buffers produced a 100x reduction in defect density. The growth of low defect-density, fully relaxed $Si_{1-x}Ge_x$ buffers is vital to the fabrication of enhanced performance hetero-TFETs. The low defect density ensures low junction leakage and I_{off}. The growth of these graded buffers relies on minimizing the dislocation nucleation rates while maximizing the dislocation glide velocity. The low dislocation nucleation is achieved by slowly increasing the Ge content over several microns of thickness of film while high dislocation glide velocity is achieved by epitaxial growth at high temperatures [5.1]. The glide of dislocation threads is impeded by dislocation pileups which form as a result of crosshatches that arise from mismatched Epitaxy. Since threads get trapped in pileups, new dislocations need to nucleate to relax graded layers, increasing the final defect density of the epitaxial film. Typically $Si_{1-x}Ge_x$ growth is interrupted at x=0.5 and CMP is used to remove crosshatches [5.1]. The work presented in this chapter however does not attempt to use graded buffer layers. Instead thin completely relaxed germanium layers (30-80nm) were used as a relaxed template on which to grow strained silicon layers.

5.2 STRAINED $SI_{1-X}GE_X$ ON SILICON HETEROSTRUCTURE MOSFET GIDL

High mobility channel materials like strained germanium and strained $Si_{1-x}Ge_x$ have received much interest for enhanced P-MOSFET performance. The strained $Si_{1-x}Ge_x$ channel MOSFETs fabricated at SEMATECH involve very thin, high germanium percentage, layers

grown on a silicon substrate and capped with a thin relaxed silicon layer to form a good interface with the high-K dielectric. A relaxed silicon cap (passivation layer) on the strained channel material was found to drastically improve the quality of the interface with the dielectric. This chapter deals with gate induced BTBT current or GIDL measurements from such MOSFETs. Valuable processing concerns and design concerns are addressed using the GIDL data from these strained $Si_{1-x}Ge_x$ channel MOSFETs.

The s-$Si_{1-x}Ge_x$ layer and relaxed silicon cap are deposited using UHVCVD at 500 $^{\circ}$C. Once the high K, metal gate stack is formed, the source and drain are ion implanted and annealing is performed at 500 $^{\circ}$C for 1min. The low temperature anneal is to prevent any strain relaxation of the $Si_{1-x}Ge_x$ channel. A 900 $^{\circ}$C anneal temperature split is also fabricated. The Source and drain are reverse biased with respect to the body and the gate is biased so the source and drain are in deep depletion, allowing for BTBT to occur. The BTBT generated carriers are collected by the reverse biased body.

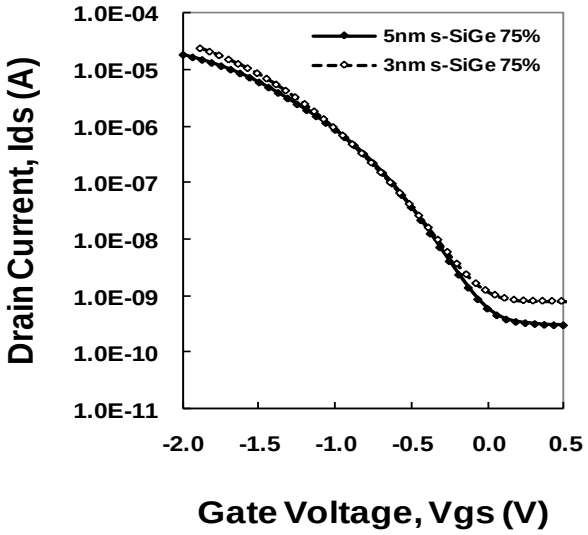


Figure 5.1 GIDL current from two heterostructure N-MOSFETs with 3nm HfO2, one with a 5nm thick strained $Si_{0.25}Ge_{0.75}$ layer on silicon substrate and one with a 3nm thick strained $Si_{0.25}Ge_{0.75}$ layer on silicon substrate. Both transistors were activated at 500 $^{\circ}$C.

Figure 5.1 shows the GIDL current measured from two heterostructure N-MOSFETs, one with a 5nm thick strained $Si_{0.25}Ge_{0.75}$ layer on silicon substrate and one with a 3nm thick strained $Si_{0.25}Ge_{0.75}$ layer on silicon substrate. Both transistors have 3nm of HfO_2 as gate dielectric and were activated at 500 $^{\circ}$C. The negligible difference seen between them at low gate biases could possibly be explained by a smeared out, non-abrupt heterostructure interface. At higher voltages, the thinner 3nm s-$Si_{0.25}Ge_{0.75}$ cap layer seems to outperform the thicker s-$Si_{0.25}Ge_{0.75}$ layer as expected and seen in the simulations in the previous chapter.

Figure 5.2 shows the temperature dependence of GIDL current of the heterostructure N-MOSFET with a 3nm thick strained $Si_{0.25}Ge_{0.75}$ layer on silicon substrate. Because of the large decrease in I_{off} with temperature the swing at low currents is lower with decreasing temperature. The I_{on} seems fairly independent of temperature. Figure 5.3 shows the GIDL current measured as

the drain and source bias are swept, keeping the bulk at ground. The curve generated at different gate biases is similar to the Id-Vd sweep of a TFET. The GIDL current increases non-linearly with low body bias and saturates at large body biases just like in a MOSFET.

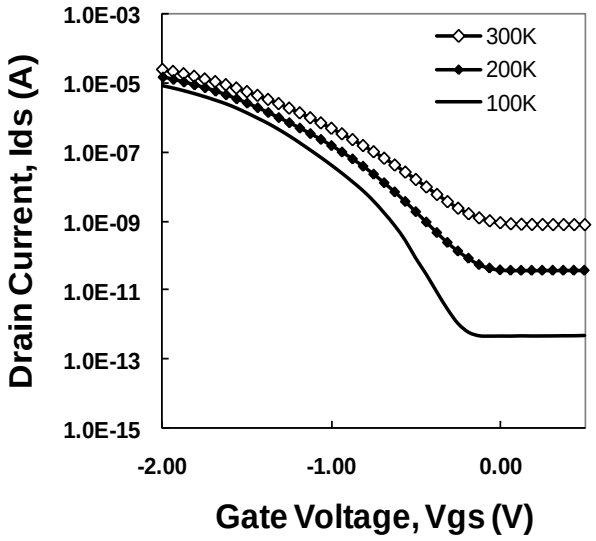


Figure 5.2 Temperature dependence of GIDL current of the heterostructure N-MOSFET with a 3nm thick strained $Si_{0.25}Ge_{0.75}$ layer on silicon substrate. I_{off} changes pretty significantly with temperature.

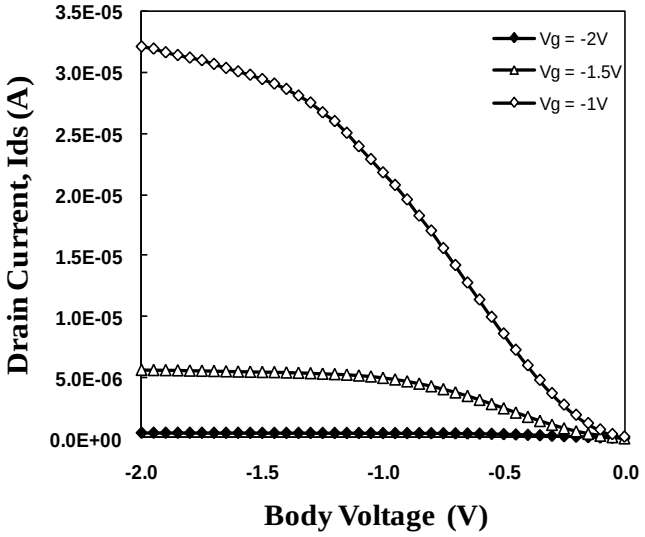


Figure 5.3 GIDL current vs. body voltage characteristics as the drain and source bias are swept, keeping the bulk at ground. This is equivalent to the Id-Vd sweep of a TFET. The GIDL current increases non linearly at low body bias and then saturates at higher bias.

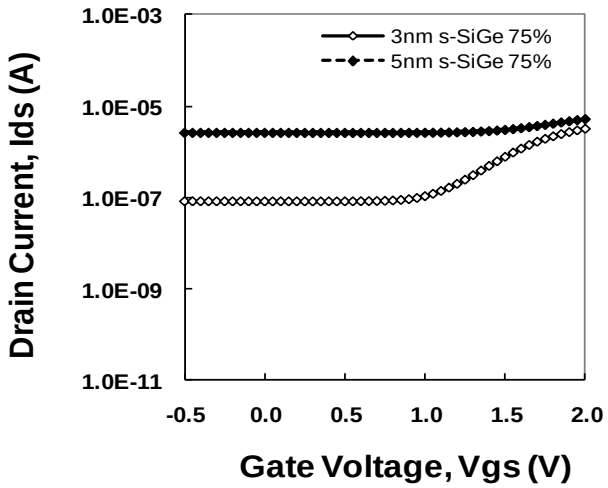


Figure 5.4 The GIDL current measured from two heterostructure P-MOSFETs, one with a 5nm thick strained $Si_{0.25}Ge_{0.75}$ layer on Si substrate and one with a 3nm thick strained $Si_{0.25}Ge_{0.75}$ layer on Si substrate. Both were activated at 500 °C and show higher off-state leakage than the corresponding N-MOSFETs.

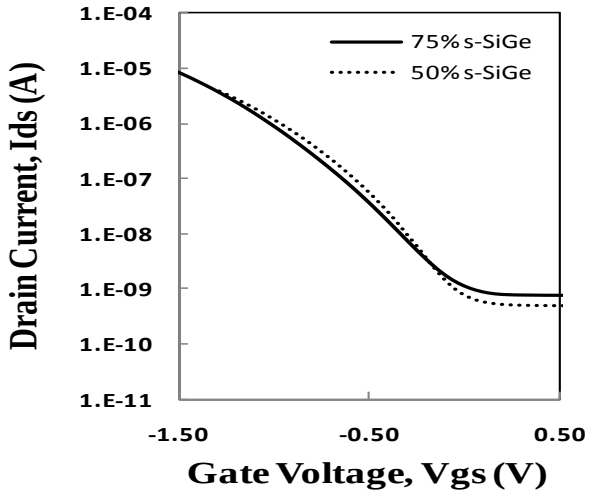


Figure 5.5 The GIDL current measured from two heterostructure N-MOSFETs, with a 3nm thick strained $Si_{1-x}Ge_x$ (Split1 - x = 50%, Split2 -x= 75%) layer on silicon substrate. Both transistors have 3nm of HfO_2 as gate dielectric and were activated at 500 °C.

Figure 5.4 shows the GIDL current measured from two heterostructure P-MOSFETs, with 5nm and 3nm thick strained $Si_{0.25}Ge_{0.75}$ layers on silicon substrate. Again both were activated at 500 °C. Both show higher junction leakage than the corresponding N-MOSFETs. This indicates that with a low temperature process (500 °C) the damage annealing for N-type dopants is more effective than for P-type dopants.

Figure 5.5 shows the GIDL current measured from two heterostructure N-MOSFETs, with a 3nm thick strained $Si_{0.25}Ge_{0.75}$ layer on silicon substrate and a 3nm thick strained $Si_{0.5}Ge_{0.5}$ layer on silicon substrate. Both transistors have 3nm of HfO_2 as gate dielectric and were activated at 500 °C. Theoretically the effective bandgap for hetero BTBT is smaller for the transistor with a strained $Si_{0.25}Ge_{0.75}$ layer on silicon, but the negligible difference seen for varying germanium percentage is again possibly because of a smeared out heterostructure interface. Even though the thermal budget was maintained fairly low, the temperature of initial growth (550 °C) could have also contributed to some intermixing and a less abrupt heterostructure interface.

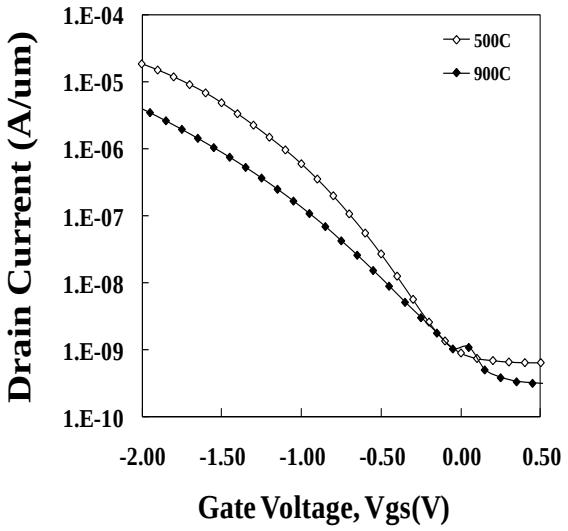


Figure 5.6 The GIDL current measured from two heterostructure N-MOSFETs, with 3nm thick strained $Si_{0.25}Ge_{0.75}$ layer on silicon substrate and 3nm HfO_2 as gate dielectric, one activated at 500 °C and the other at 900 °C. A significant degradation is noticed for the 900 °C annealed MOSFET confirming the importance of a low temperature process.

Figure 5.6 shows the GIDL current measured from two heterostructure N-MOSFETs, with a 3nm thick strained $Si_{0.25}Ge_{0.75}$ layer on silicon substrate and 3nm HfO2 as gate dielectric, one activated at 500 °C and the other at 900 °C. A significant degradation is noticed for the 900 °C annealed MOSFET. Higher temperature processing poses the risk of increased strain relaxation and increased inter-diffusion across the heterostructure. Not only can the heterostructure interface get less abrupt, but the greater amount of germanium atoms diffusing to the interface can also lead to worse dielectric interface quality and therefore worse modulation of the surface potential by the gate. The data in figure 5.6 confirms the importance of using a low temperature process that can anneal implant damage and achieve dopant activation without strain relaxation and smearing out of the $Si_{1-x}Ge_x$/Si hetero interface, both unwanted effects for heterostructure TFETs.

From GIDL measurements of strained $Si_{1-x}Ge_x$ channel MOSFETs, several valuable observations are made to help aid a more optimal heterostructure TFET design and process.

5.3 Strained Silicon on Relaxed Germanium Heterostructure MOSFET GIDL and Verification of Hetero Band to Band Tunneling Concept

The cross section of a heterostructure MOSFET with silicon cap on a relaxed germanium layer on silicon substrate is illustrated in Figure 5.7. Since the critical thickness of an epitaxial germanium layer on a silicon substrate is close to 1nm, a fully relaxed germanium layer can be achieved by depositing tens of nanometers. Since the germanium relaxes by forming defects, the thickness of the relaxed germanium layer is found to affect the junction leakage of the MOSFETs as shown in Figure 5.8.

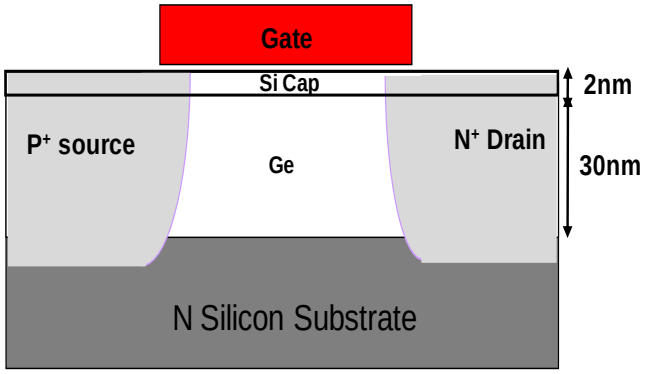


Figure 5.7 Cross section of a heterostructure MOSFET with s-silicon cap on relaxed germanium on a silicon substrate.

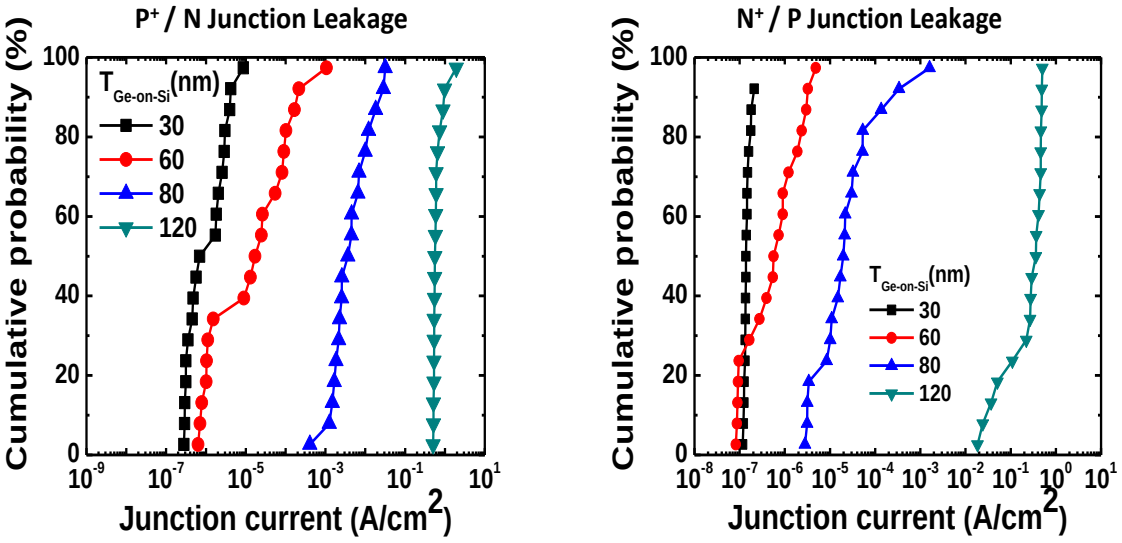


Figure 5.8 Effect of relaxed germanium thickness on the P^+/N and N^+/P junction leakage. The 30nm film gives the lowest junction leakage in both cases. (Data Courtesy Dr Jungwoo. Oh, SEMATECH)

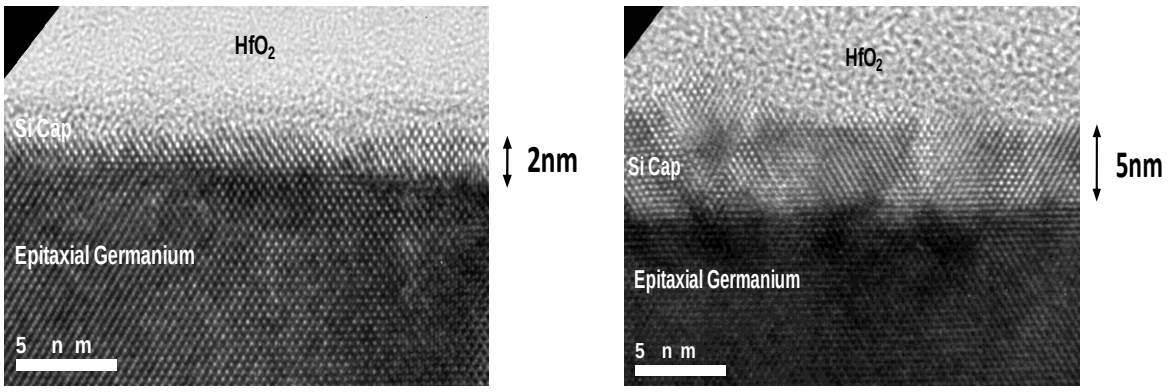


Figure 5.10 TEM images of a 2nm silicon cap on relaxed germanium and 5nm silicon cap on relaxed germanium. The interface between germanium and silicon does not look perfectly abrupt in either case. The 5nm cap is relaxed, with the formation of defects.

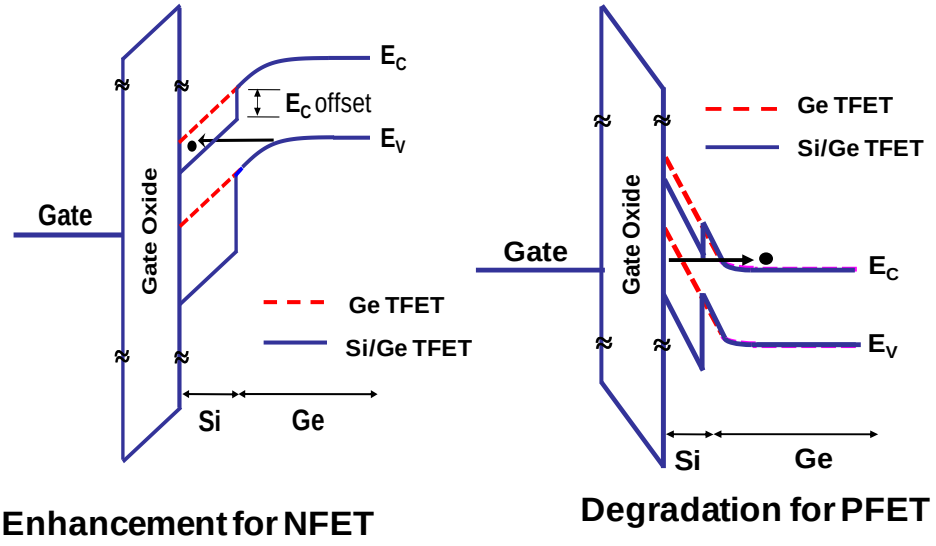


Figure 5.11 Band alignment indicating benefit of reduced effective bandgap across the heterostructure for PMOSFET GIDL measurement (bands bend down). For NMOSFET GIDL measurement (bands bend up), the reduced bandgap is not beneficial because of the band alignment.

If the silicon cap grown epitaxially on the germanium layer is below the critical thickness, it is strained to have the lattice constant of the germanium and the band alignment is that seen in Figure 5.11. Figure 5.11 shows the band diagram in the source and drain of N-type and P-type MOSFETs when the gate is biased to induce deep depletion. In the N-type MOSFETs when the bands bend down, the strained silicon cap on relaxed germanium heterostructure has a reduced effective bandgap ($E_{g,Ge} - \Delta E_c$) for tunneling allowing for enhanced BTBT. In the P-type MOSFETs when the bands bend up, the strained silicon cap on relaxed germanium heterostructure increases the effective bandgap for tunneling degrading the BTBT current.

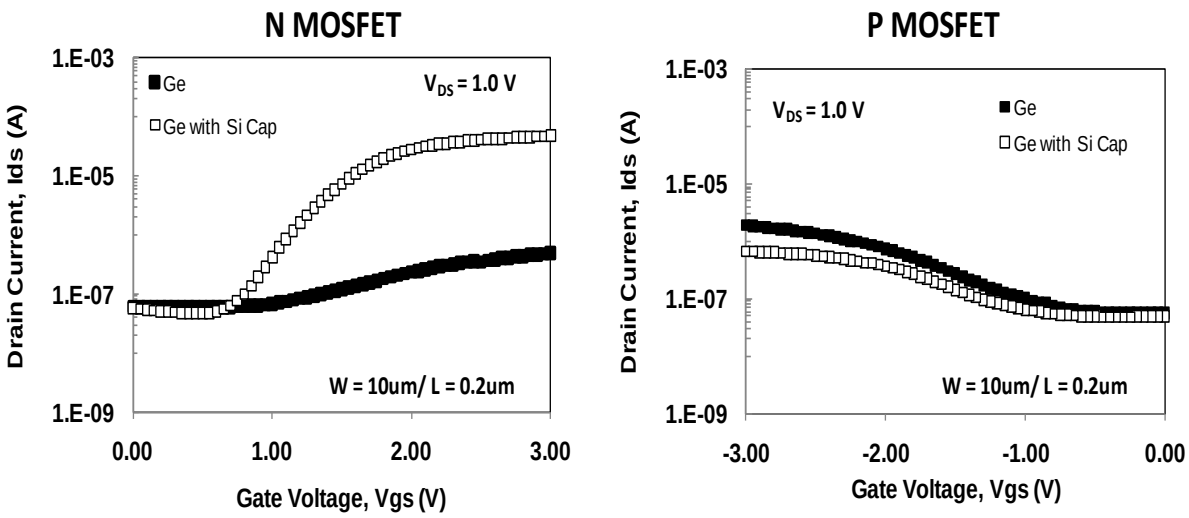


Figure 5.12 Comparison of GIDL current from hetero NMOSFET and PMOSFET with pure relaxed germanium and 2nm s-silicon cap on relaxed germanium. The Ge with Si cap NMOSFET outperforms the pure Ge as expected from the band alignment in Figure 5.11. The Ge with Si cap PMOSFET underperforms the pure Ge despite the better dielectric interface as indicated by the band alignment in Figure 5.11.

As shown in Figure 5.12, the GIDL current from the heterostructure N-type MOSFET with 2nm silicon cap is more than 100x larger than the current from the relaxed germanium N-type MOSFET. Since silicon forms a higher quality interface with high K dielectric than germanium, the presence of the silicon cap in the heterostructure MOSFET could explain the enhancement in GIDL current. If this however was the only cause for the enhancement in current, the GIDL current from heterostructure P-type MOSFETs would also be enhanced compared to germanium P-type MOSFETs. This however is not the case as seen in Figure 5.12. The degraded current from the heterostructure P-type MOSFET despite the better interface quality and gate control of the surface potential can only be a result of the enhanced effective bandgap across the silicon/ relaxed germanium heterostructure when the bands bend up. The relative enhancement in heterostructure GIDL current over germanium GIDL current when the bands bend down and degradation when the bands bend up acts as confirmation of the concept of heterostructure BTBT.

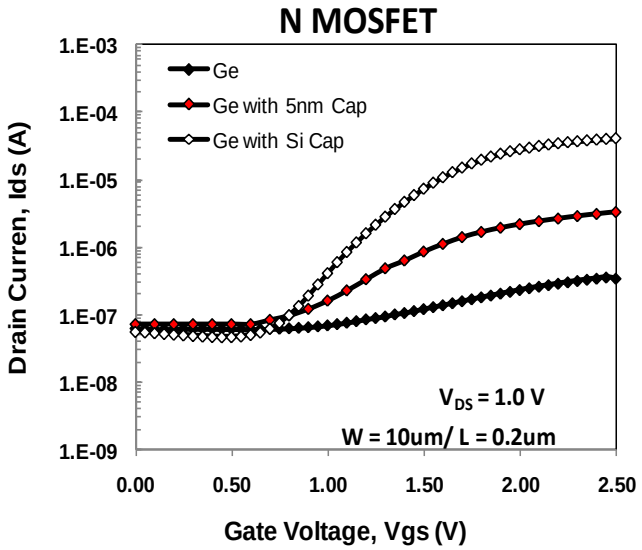


Figure 5.13 NMOSFET GIDL current for pure Ge, Ge with 2nm Si cap and Ge with 5nm Si Cap MOSFETs. Despite both having the advantage of better interface quality, the Ge with 2nm Si cap outperforms the Ge with 5nm Si cap again verifying the enhancement in BTBT from a reduced effective bandgap across the heterostructure.

Figure 5.13 compares the GIDL current from MOSFETs with (i) a relaxed germanium layer and no silicon cap (ii) with a strained silicon cap (2nm) on the relaxed germanium layer and (iii) with a relaxed silicon cap (5nm) on the relaxed germanium layer. If the enhancement of the heterostructure with the 2nm silicon cap is only because of an improved dielectric interface, the heterostructure with the 5nm silicon cap would be expected to show similar enhancement. If however the enhancement of the heterostructure with the 2nm silicon cap is due to the reduced effective bandgap ($E_{g,Ge} - \Delta E_c$) across the heterostructure then it would outperform the heterostructure with the 5nm silicon cap since the ΔE_c is larger for a strained silicon cap on relaxed germanium. From Fig. 13 the heterostructure with the 2nm cap does outperform the heterostructure with the 5nm cap again confirming the concept of enhanced heterostructure tunneling performance due to a reduced effective bandgap.

5.4 Impact of Mechanical Strain on Heterostructure MOSFET GIDL

Uniaxial strain is known to reduce crystal symmetry, lift degeneracy and warp both conduction and valence bands. Applying stress along a low symmetry direction causes more destruction of crystal symmetry and band warping than stress along other directions [5.3]. In both silicon and germanium optical phonon scattering dominates. In silicon uniaxial tensile strain causes splitting of the Δ_2 and Δ_4 bands in silicon, while in germanium it causes splitting of the L valleys [5.4]. This splitting leads to mobility enhancement due to reduced intervalley scattering.

Uniaxial compressive strain causes both band warping and splitting of the LH and HH bands. At low stress the splitting is enough to ensure that most of the holes reside in the top band atleast when the hole density is low. At large strains the valence band becomes asymmetric by warping and the splitting between LH and HH is large enough to prevent all optical phonon scattering between top and bottom valence bands [5.3].

Since GIDL is due to BTBT in the drain of the MOSFET and since BTBT generation rate is exponentially related to the carrier mass and bandgap, changes in bandgap and carrier mass due to mechanical strain should be reflected by the GIDL current. Mechanical strain is applied by bending the wafer using a strain jig [5.5]. The NMOSFETs and PMOSFETs are oriented with channel along the <110>. Figure 5.14 compares the variation in GIDL current with mechanical strain for strain applied along two directions: (i) the channel <110> direction and (ii) along the <100> direction perpendicular to the channel direction. The figure shows increase in GIDL current with tensile strain and decrease in GIDL current with compressive strain for bending along both <110> and <100> directions. The change in GIDL current with strain is slightly larger for wafer bending along the <100>.

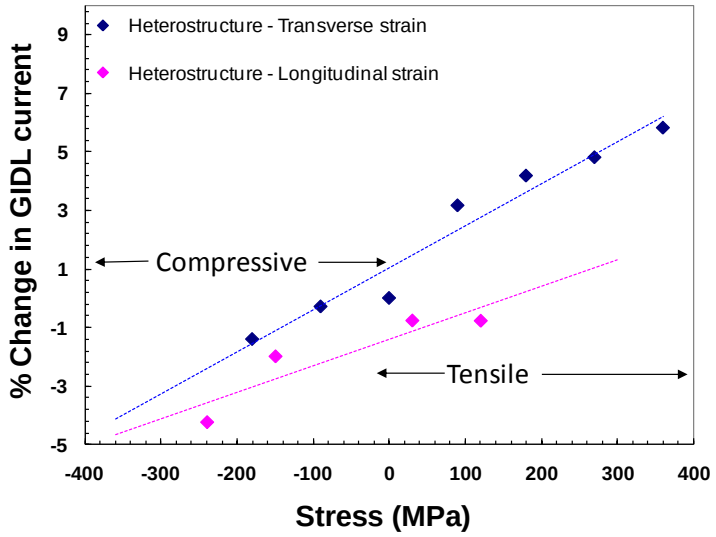


Figure 5.14 Variation in GIDL current with mechanical strain for strain applied along two directions: (i) the channel <110> direction (longitudinal) and (ii) along the <100> direction perpendicular to the channel (transverse) direction. Increase in GIDL current seen with tensile strain and decrease in GIDL current seen with compressive strain for bending along both <110> and <100> directions. The change in GIDL current with strain is slightly larger for wafer bending along the <100>.

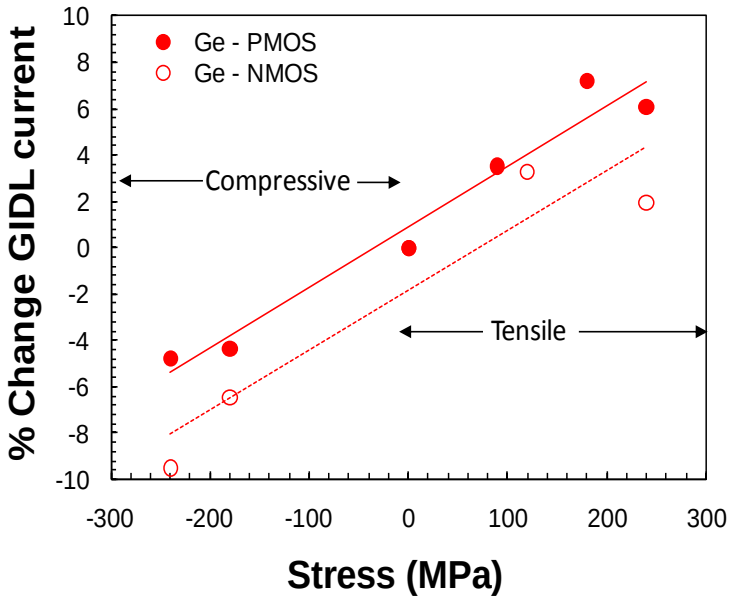


Figure 5.15 Variation of GIDL current with strain along the <110> direction for germanium N and P MOSFETs. GIDL current increases with tensile strain and decreases with compressive strain in both cases.

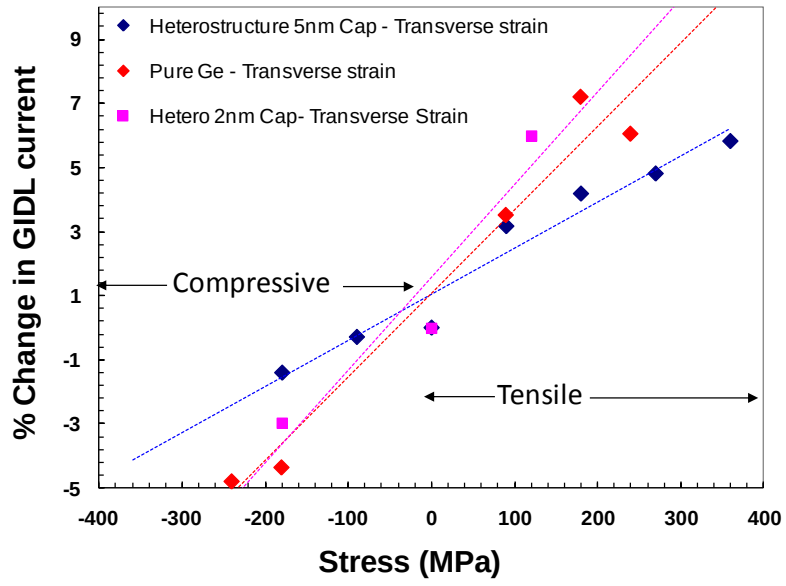


Figure 5.16 Variation of GIDL current with strain along the <100> direction (transverse to the channel) for pure Ge, Ge with 2nm Si cap and Ge with 5nm Si cap MOSFETs. Again GIDL current increases with tensile strain and decreases with compressive strain.

Figure 5.15 compares the change in GIDL current with strain for strain along <110> for germanium NMOSFETs and PMOSFETs. Both NMOSFETs and PMOSFETs show increase in GIDL current with tensile strain and decrease in GIDL current with compressive strain. Figure 5.16 shows the same trend seen in Figure 5.14 and figure 5.15 for both germanium MOSFETs as well as heterostructure MOSFETs with varying silicon cap thickness.

Wafer breakage set the limit on the maximum mechanical strain applied to the samples. This measured change in GIDL due to uniaxial strain can further be exploited in TFETs by including process induced stressing layers such as nitride capping layers.

5.5 REFERENCES

[5.1] M. L. Lee, E. A. Fitzgerald, M. T. Bulsara, M. T. Currie, and A. Lochtefeld, "Strained Si, SiGe, and Ge channels for high-mobility metal-oxide-semiconductor field-effect transistors", J. Appl. Phys. **97**, 011101 (2005)

[5.2] J. L. Hoyt, H. M. Nayfeh, S. Eguchi, I. Aberg, G. Xia, T. Drake, E. A. Fitzgerald, A. D. Antoniadis, "Strained silicon MOSFET technology," Electron Devices Meeting, 2002. IEDM '02. Digest. International, vol., no., pp. 23- 26, 2002

[5.3] Y. Sun, S. E. Thompson and T. Nishida, "Physics of strain effects in semiconductors and metal-oxide-semiconductor field-effect transistors", J. Appl. Phys. 101, 104503 (2007)

[5.4] C. G. Van de Walle and R. M. Martin, "Theoretical calculations of heterojunction discontinuities in the Si/Ge system", Physical Review. B, 34**,** 5621–34, 1986

[5.5] S. Suthram, J. C. Ziegert, T. Nishida, and S. E. Thompson, Fellow, IEEE, "Piezoresistance Coefficients of (100) Silicon nMOSFETs Measured at Low and High (~1.5 GPa) Channel Stress", IEEE Electron Device Letters, Vol. 28, No. 1, January 2007

Chapter 6

Bulk Germanium Homojunction and Heterostructure TFET Processing

6.1 INTRODUCTION

There has been tremendous interest in using germanium as a channel material for highly scaled MOSFETs because germanium can provide higher carrier mobility than silicon and therefore provide increased drive currents [6.1, 6.2]. The biggest challenge with integrating germanium channels however has been the formation of a high quality gate stack on germanium crystal. Deposited high-K dielectrics on germanium are being heavily researched and are the most viable gate dielectric option [6.3, 6.4]. Dopant activation [6.5, 6.6], damage annealing [6.7, 6.2] and forming good ohmic contacts [6.8, 6.9] are other aspects of integrating germanium channels that are being heavily researched. [6.2] give experimental evidence that it is possible to achieve low junction leakage using bulk germanium substrates using an optimized dopant activation anneal. This chapter discusses why fabricating N-type hetero TFETs on bulk germanium is even considered. It also discusses in detail several important modules required to fabricate these transistors.

6.2 HETERO TFET PROCESSING CONCERNS

N-type Hetero TFETs with a strained silicon cap on relaxed germanium and P-type Hetero TFETs with a strained germanium on strained silicon on relaxed $Si_{0.5}Ge_{0.5}$ can be fabricated either with a gate first process or with a gate last process. Each process has its own merits and demerits. The key factors that need to be taken into account when considering either process are as follows:

(a) Relaxed $Si_{1-x}Ge_x$ layer

Starting with a silicon substrate, a relaxed germanium layer is needed for N-type TFETs and a relaxed $Si_{0.5}Ge_{0.5}$ layer is needed for P-type TFETs. The thickness and defect density (quality) of this layer will determine the junction leakage and TFET I_{off}. For N-type TFETs starting with a bulk germanium substrate eliminates the need for the relaxed germanium layer and could ideally allow for lower junction leakage if implant induced damage is annealed sufficiently.

(b) Strained Silicon/Strained Germanium capping layer

For N-type TFETs the quality and thickness of the strained silicon capping layer atop the relaxed germanium layer will determine the TFET I_{on} and for P-type TFETs the quality and thickness of both the strained germanium layer and the strained silicon layer atop the relaxed $Si_{0.5}Ge_{0.5}$ will determine the TFET I_{on}. In order for the TFET to benefit from the reduced effective bandgap

across s-Si/relaxed Ge heterostructure or the s-Ge/s-Si heterostructure, any chosen process needs to ensure no strain relaxation through the entire fabrication process.

The ideal process would also ensure reduced diffusion across the heterostructure and retain an abrupt heterostructure interface. The growth temperature of the strained silicon on Ge (N-type TFET) or on $Si_{0.5}Ge_{0.5}$ (P-type TFET) needs to be relatively low for minimum germanium inter-diffusion. This is a challenge because silicon source gases such as SiH_4, Si_2H_6, and SiH_2Cl_2 decompose at an impractically slow rate at 350–400 °C. Hence a two-step growth process is therefore required to grow the s-Si cap. The first step is to initiate SiH_4 flow while simultaneously raising the furnace temperature to 450 °C [6.10]. Although the Si growth rate is virtually negligible at these low temperatures, the ability of hydrides to pyrolize more rapidly on pure germanium allows a very small amount of silicon to be deposited, helping to stabilize the morphology of the underlying Ge layer. With SiH_4 still flowing, the second step is to raise the temperature to 550 °C, where the Si cap growth can be completed in several minutes.

The ideal P-type Hetero TFETs require a compressive strained pure germanium layer on a tensile strained pure silicon layer on a relaxed $Si_{0.5}Ge_{0.5}$ substrate. Germanium rich layers have a tendency to form coherent islands or surface ripples via stress-driven surface diffusion [6.10]. This can be avoided by lowering the epitaxial growth temperature to around 350 $^{\circ}$C and therefore preventing islanding and growing the strained germanium layer on top of the strained silicon layer is easily achievable by lowering the temperature to 350 $^{\circ}$C. Keeping the growth temperature of strained Ge between 300-400 $^{\circ}$C automatically keeps inter-diffusion from and strain relaxation of any layers beneath to a minimum.

Lastly it is vital that the topmost strained layers are not etched away during processes such as gate etch or spacer etch since the BTBT generated carriers in the source flow into the drain through the strained cap layer.

(c) Dielectric interface

The quality of the interface between the gate dielectric and the top most strained layer also severely impacts both the TFET I_{on} and steep turn on characteristics. A low interface trap density is highly desirable for steep swing. For N-type TFETs the up-diffusion of germanium into the strained cap is least desirable since germanium forms a more unstable interface than silicon. For P-type TFETs a thin Si passivating layer needs to be deposited on top of the topmost s-Ge layer to form a higher quality dielectric interface. In the presence of interface traps, gate voltage which would have otherwise caused more band bending in the semiconductor, is wasted in filling and un-filling these traps.

(d) Source and Drain regions

The creation of the source and drain regions impacts the junction leakage and TFET I_{off}. In-situ doping during epitaxial growth is one method after which no anneal is required as the dopants are almost 99% activated and there is no damage which needs to be annealed. For electrons generated by BTBT in the source to be collected by the drain, a hetero TFET needs a continuous s-Si or s-Ge layer which is doped oppositely in the source and in the drain. This requirement makes it challenging to use a source/drain etch and refill process for hetero TFETs.

The simplest method of creating the source and drain is using ion implantation. This is easy to integrate but yields non abrupt junctions and requires a sufficient high temperature anneal after the implant to activate the dopants as well as remove implant induced damage. The high temperature anneal is advantageous to anneal implant induced damage but would increase possible strain relaxation of strained layers, increase inter-diffusion across the layers making the heterostructure interface less abrupt, and degrade the dielectric interface by allowing up

diffusion of more germanium atoms to the interface. The anneal needs to be carefully optimized taking all these factors into account.

6.3 GATE LAST HETERO-TFET PROCESSING

The GIDL measurements from gate first heterostructure MOSFETs with relaxed germanium were a clear indication that in a gate first heterostructure TFET process high junction leakage would obscure the true nature of hetero-BTBT close to turn on. Increasing the annealing time could be a means to lower junction leakage but raises the risk of relaxing the strained layer atop the germanium and increases the risk of inter-diffusion and therefore a less abrupt junction at the heterostructure interface. To try and decouple the implant damage annealing from the heterostructure interface quality, a gate last hetero-TFET process was attempted.

The gate last hetero-TFET process flow involved first the growth of relaxed epitaxial germanium on the silicon substrate followed by ion implantation of the source and drain dopants using half masks. This is followed by a long anneal to completely remove all implant induced damage before growing the strained silicon cap on the relaxed germanium. The gate stack is then deposited and patterned before finally contacting the source, drain and gate with metal.

Module development for this process flow involved development of epitaxial growth of the strained silicon cap on relaxed germanium after ion implantation and anneal. Two splits involving 10keV, 3e15 Bf_2 implants into the source side (drain side covered) were created. They were then annealed at 500 $^{\circ}$C, for 3 minutes and 6 minutes. The wafers were then cleaned and loaded into the RTCVD chamber for epitaxial growth of the silicon cap using SiH_4. Inspection of the wafers with an SEM after the growth revealed islanding of the silicon cap on the implanted and annealed source side. Figure 6.1 shows the top down SEM image with a clear difference between the implanted and the non implanted side.

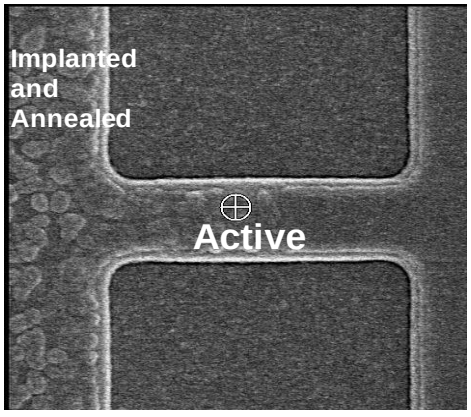
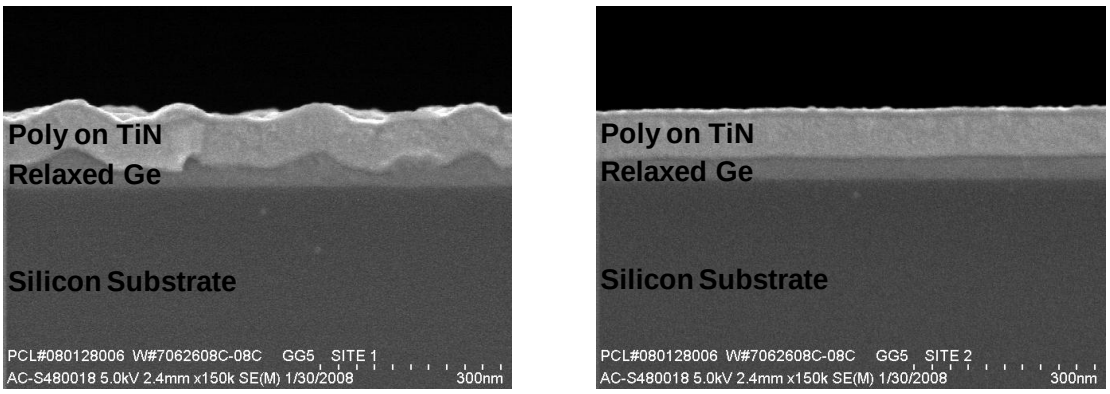


Figure 6.1 Top down SEM image of the active area after source side implant, anneal, clean and strained-Silicon epitaxial growth. Islanding seen instead of a uniform film on the implanted and annealed side of the active area.



Implanted Side **Non-Implanted Side**

Figure 6.2 Cross sectional SEM images of the implanted and non-implanted sides of the active area. The implanted side shows undulations and roughness while the non implanted side is smooth and regular.

Figure 6.2 shows cross section SEMs of the implanted and non implanted side, indicating that the relaxed germanium on the implanted side was no longer planar after implant and the long anneal. The surface had severe undulations and roughness making it far from ideal to be the base on layer for a thin strained silicon cap layer. One possible solution to this could be using solid source diffusion to try and introduce dopants into the source and drain although it is not clear what the effect of the long thermal budget would be on the relaxed germanium film. The other possible solution could be to etch out the source and drain and use in-situ doped expitaxial growth to create the source and drain regions.

6.4 GATE FIRST HETERO-TFET PROCESSING ON BULK GERMANIUM SUBSTRATE

To try and get away from relaxed $Si_{1-x}Ge_x$ related junction leakage issues, bulk germanium processing is considered. This is not an ultimate solution since it makes it difficult to integrate with other circuits fabricated on silicon substrate but is a first step to try and understand the true nature of hetero BTBT turn on characteristics without facing enhanced junction leakage due to relaxed $Si_{1-x}Ge_x$ layers. Gate first hetero-TFET processing on bulk germanium substrates required the development of several modules. The three main modules which are discussed in detailed here are A) developing a process to grow strained-Si caps on bulk Ge substrates, B) developing a working high-k based gate stack on bulk germanium substrates and testing Capacitors and C) developing a working process flow for homojunction TFETs on bulk germanium substrates.

A) S-SILICON CAP GROWTH ON BULK GERMANIUM FOR N-TYPE HETERO TFETS

Bulk germanium substrates were cleaned in 25:1 HF, DI water rinsed and dried before loading into LPCVD chamber. To remove the GeO_x not removed by the HF clean the substrates were baked to 800 $^{\circ}$C in H_2 and then cooled to 400 $^{\circ}$C before growing an epitaxial germanium layer.

The strained silicon cap was to be grown after growing a thin lattice matched epitaxial germanium layer. Slip lines due to thermal stress were visible (figure 6.3) on unloading the wafer. Only the center of the wafer had no slip lines.

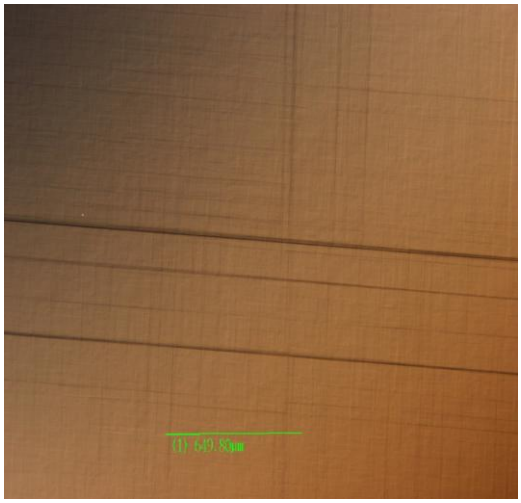


Figure 6.3 Slip lines seen visually across the entire wafer except at the center of the wafer.

To reduce the thermal stress the process was optimized. The wafers were loaded at 250 $^{\circ}$C, H_2 baked at 650 $^{\circ}$C and then cooled down to 400 $^{\circ}$C with a very slow thermal ramp rate. A lattice matched epitaxial germanium layer was then grown at 400 $^{\circ}$C. After the germanium growth, at 400 $^{\circ}$C, SiH_4 flow is initiated at 400 $^{\circ}$C and the temperature is ramped up to 550 $^{\circ}$C to allow the growth of a thin silicon cap. The growth rate of silicon by decomposition of SiH_4 at these low temperatures (below 550 $^{\circ}$C) is very low allowing for good control of a thin strained silicon cap layer. Because of the large lattice mismatch between silicon and germanium, the silicon is expected to relax at a fairly early stage and influence the growth mode.

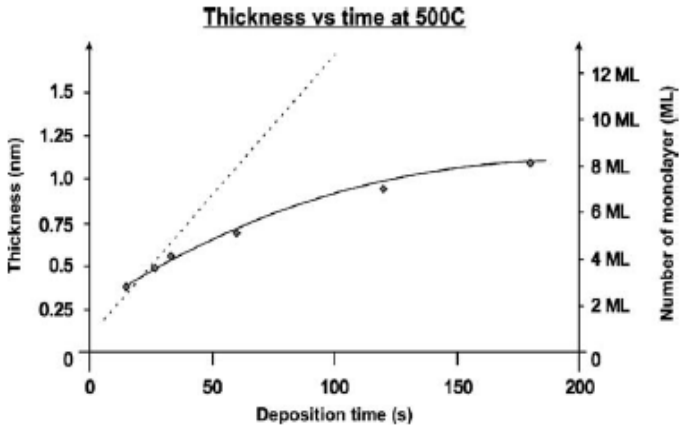


Figure 6.4 Growth curve of s-silicon cap layer on bulk germanium wafers by IMEC [6.11].

Similar studies of silicon cap on bulk germanium wafers by IMEC [6.11] showed that at 500 °C, the thickness of silicon grown on germanium was self limiting to around 1nm after 200s. Growth temperatures of 550 °C or 575 °C would allow for silicon thickness greater than 1nm but these films would be relaxed through the formation of regular arrays of misfit dislocations. In the study by IMEC, for 1-1.2nm silicon films grown at 575 °C, no dislocations were observed in the TEM images, while for 1.8-2.2nm silicon films grown at 575 °C, relaxation through dislocations with a periodicity of ~10nm was observed in the TEM images. In our experiment, TEMs were not available during the characterization of the silicon cap growth at Lawrence Semiconductor Research Lab. A growth curve was established with 3 different growth times at 550 °C and then a range of growth times at 550 °C were chosen to account for both an initial incubation period and still try and capture within the various splits silicon films both <1nm and >1nm in thickness.

B) GERMANIUM CAPACITORS

In order to determine a working gate stack for germanium substrates, a cleaning and degreasing procedure for germanium wafers had to be established along with a pre-clean procedure before depositing the gate dielectric on the substrate. A soak in a 120 °C piranha bath followed by a DI water rinse is the standard procedure to clean silicon wafers but piranha attacks germanium and therefore the cleaning procedure was changed to an acetone soak followed by DI water rinse. A PRS 3000 soak would also work well.

A thermally grown SiO_2 gate dielectric with in-situ doped LPCVD poly silicon deposited at 615 °C is the standard gate stack used for silicon substrates. Germanium's thermal oxide is known to have very poor stability and is water soluble making deposited dielectrics the most viable option. Nitridation of the substrate followed by deposition of LTO has been studied to create a working gate dielectric on germanium substrates. But this is greatly surpassed in quality by ALD deposited high-K dielectrics such as Al_2O_3 and HfO_2 which not only allow for thicker physical films and lower EOT but are also less leaky.

Al_2O_3 was available to be deposited with the ALD tool in the UC Berkeley fabrication facility, making this the preferred gate dielectric. The standard pre-clean procedure for silicon substrates is a 30sec soak in 25:1 HF followed by a DI water rinse and spin dry. Germanium's native oxide is a combination of GeO_2 and GeO_X ($x<1$). While GeO_2 is easily removed in water, GeO_x is not so easily removed. The pre-clean found to be most effective for germanium substrates involves a cyclic HF + DI water rinse, ending with the HF to leave the surface hydrogen terminated. The wafer is then loaded into the ALD chamber for 50 cycles of Al_2O_3 deposition. This results in a 5nm thick film after which the wafers are loaded into the LPCVD furnace at 300 °C for poly $Si_{0.5}Ge_{0.5}$ deposition. The in-situ doped film is 1000Å thick and deposited at 425 °C using Si_2H_6, GeH_4 and Ph_3. The Si_2H_6 is needed to allow deposition to occur below 500 °C. For pure germanium TFETs, poly silicon deposited at 615 °C works well as a gate stack, while for hetero TFETs, the lower thermal budget is preferred to keep up-diffusion of germanium into the silicon cap to a minimum and to prevent strain relaxation of the silicon cap. Ideally a metal gate such as TiN would be the preferred gate electrode because of the minimal thermal budget involved in their deposition. TiN can also withstand further high temperature processing and can be etched easily in a Cl_2 ambient. This however was not a viable option in our UCB fabrication facility.

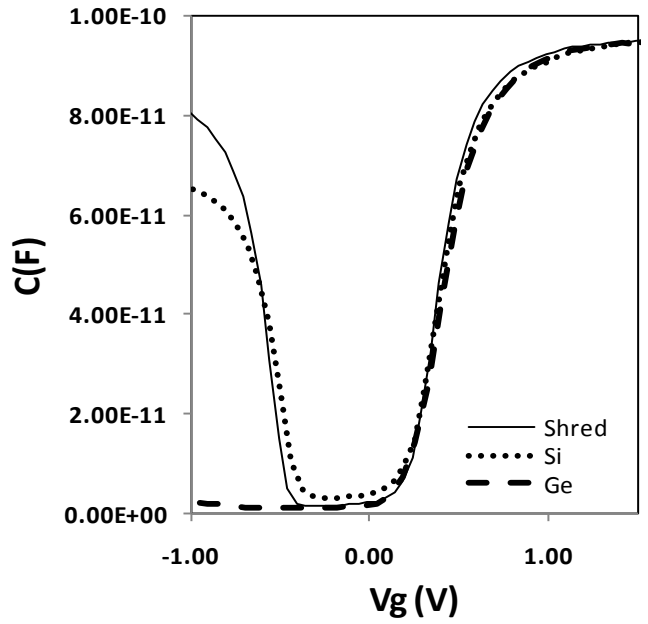


Figure 6.5 Capacitor measurements comparing a poly-Si/Al_2O_3/n-silicon capacitor and a poly-SiGe/Al_2O_3/ n-germanium capacitor with a Shred simulation of Al_2O_3 on germanium. From the ellipsometer, 5.6nm of Al_2O_3 was present on the silicon and germanium substrates. The germanium substrate was pre-cleaned with cyclic HF and DI water with HF last.

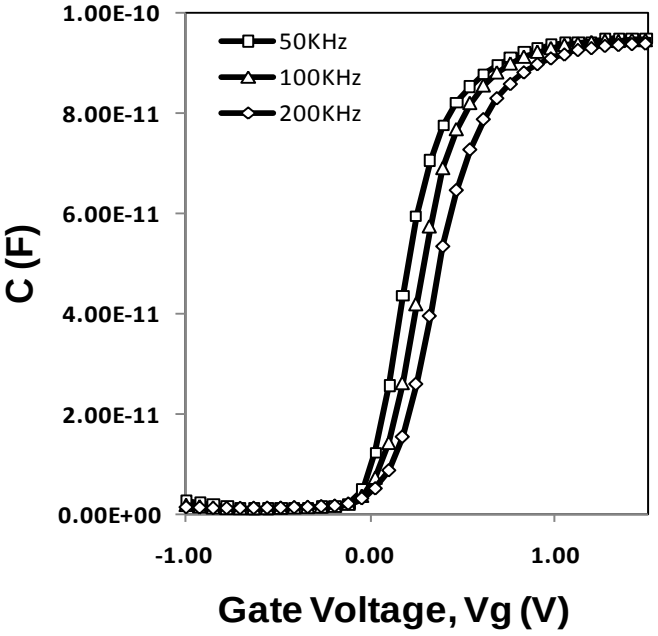


Figure 6.6 Frequency dependence of CV measurements of poly-$Si_{0.5}Ge_{0.5}$/Al_2O_3/ n-germanium capacitor

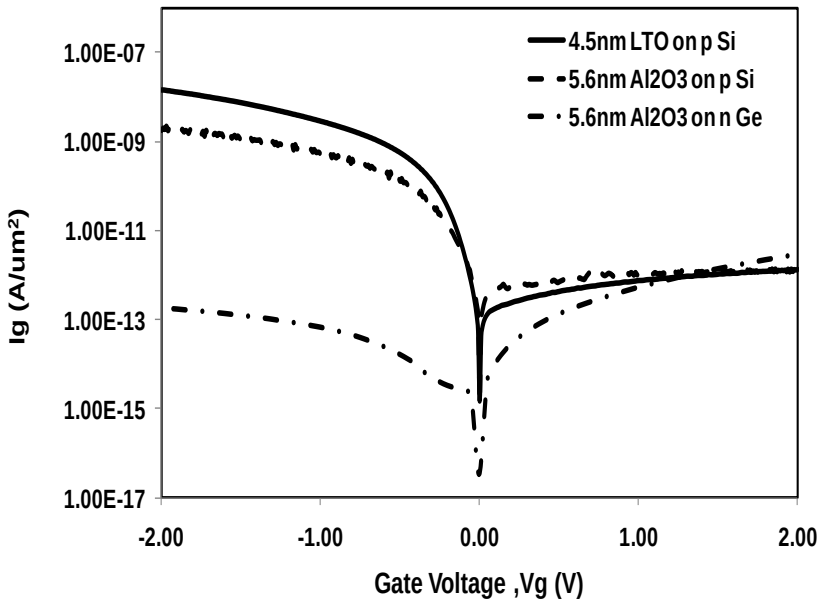


Figure 6.7 Gate current from capacitors of poly-Si/Al_2O_3/n-silicon capacitor, a poly-$Si_{0.5}Ge_{0.5}$/Al_2O_3/ n-germanium capacitor and poly-Si/SiO_2/n-silicon capacitor. The LTO oxide exhibits highest leakage as expected. The Al_2O_3 on germanium exhibits lower leakage than Al_2O_3 on silicon possibly because of a thicker interfacial layer.

Figure 6.5 shows a comparison between measured CV curve from a poly-$Si_{0.5}Ge_{0.5}$/Al_2O_3/ n-germanium capacitor and a Shred simulation. The dielectric constant of the in house Al_2O_3 was determined to be 8.5. In using this and fitting the Shred simulation to the measured CV curve, it was required to use an Al_2O_3 thickness of 6.3nm instead of 5.6nm indicating the formation of an interface layer between the germanium and Al_2O_3. Figure 6.5 also include CV measurements from a poly-Si/Al_2O_3/n-silicon capacitor. The increase in capacitance in the inversion regime is seen only for capacitors on n-silicon substrates and not for capacitors on p-silicon substrates. This increase is believed to be due to minority carrier generation and seen to occur in small bandgap semiconductors and if there exists trap sites for generation either in the bulk or at the interface.

Comparison of the gate current from a poly-Si/Al_2O_3/n-silicon capacitor, a poly-SiGe/Al_2O_3/ n-germanium capacitor and poly-Si/SiO_2/n-silicon capacitor is presented in Figure 6.7. An identical gate etch is used to etch all three capacitor gate stacks removing that as a variable. LTO is known to be porous and leaky as a gate material and is typically densified at higher than deposition temperature to reduce bulk leakage. This densification automatically occurs when poly silicon gate is deposited at 615 $^{\circ}$C in a furnace. The Al_2O_3 on p-silicon is found to be marginally less leaky than the LTO on p-silicon for negative gate biases (Accumulation). The Al_2O_3 on n-germanium is less leaky at low gate biases (<1V) for both negative and positive gate biases making it the preferred dielectric for germanium in our facility.

C) BULK GERMANIUM TFETS

Fabrication of TFETs on bulk germanium substrates was done using a ring type layout for the transistor gate. This layout allows for automatic isolation of one transistor from other transistors without the need for a LOCOS or field isolation process. The RingFET layout is shown in Figure 6.8. The gate dielectric was 5nm thick Al_2O_3 deposited using ALD at 300 °C and the gate electrode was a poly $Si_{0.5}Ge_{0.5}$ layer deposited using LPCVD at 425 °C. The wafers are very lightly n-type doped and the area of the source is limited to 3x3 um^2 for reduced junction leakage. The square gate pad attached to the gate ring is also limited to 3x3 um^2 to reduce gate leakage. Since the source and gate pad are very small in area, metal pads are used to contact them through 1x1 um^2 contact holes. Two wafer splits were fabricated one annealed at 500 °C for 1minute and one annealed at 600 °C for 1minute. The wafers after contact lithography and etch were loaded into a cluster sputter tool. In the sputter tool, the first step was to sputter etch the native oxide away, then sputter a 300Å TiN diffusion barrier layer and then finally sputter a 1000Å Aluminum (2% silicon) layer. Since the target available was Aluminum with 2% silicon and not 2% germanium, the TiN layer was needed to prevent spiking. The first split unfortunately broke in the sputter tool.

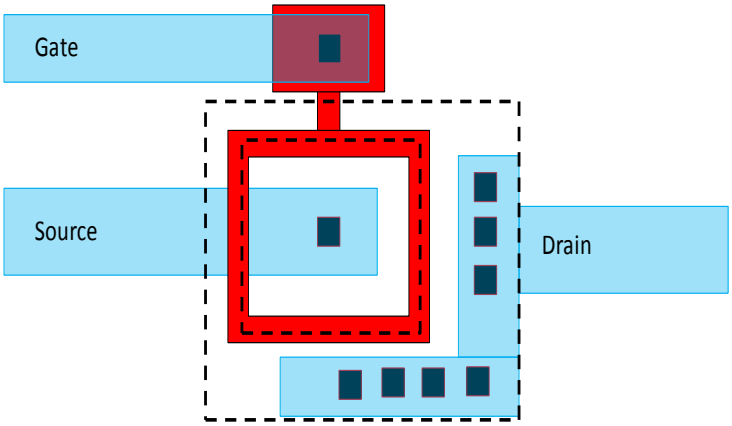


Figure 6.8 RingFET layout for fabrication of transistors on bulk germanium wafers without concerns about isolation. The gate pad area and the area enclosed by the gate are minimized for lower gate leakage and junction leakage respectively. Contacts are made through metal layer.

The remaining wafer had two on wafer splits with regards to the drain implant. Several rows of dies on the wafer received a drain implant aligned to the gate edge, while the other rows of dies received a drain implant aligned to a 50nm drain spacer. Sketches representing the two different splits are presented in Figure 6.9

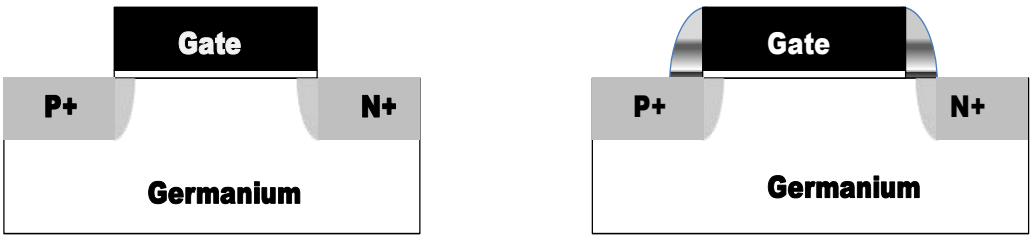
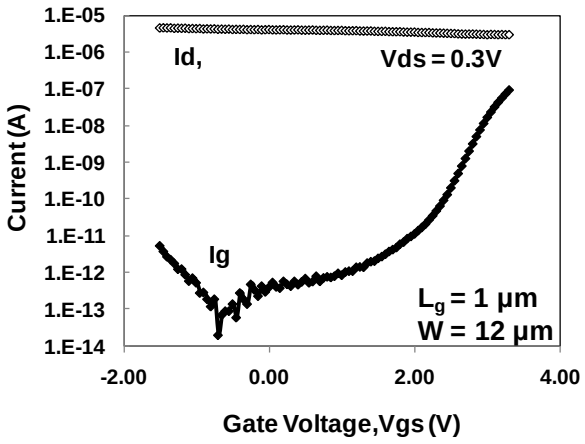


Figure 6.9 On wafer splits on the bulk germanium wafer included TFETs with drain implanted aligned to the gate edge and TFETs with drain implanted aligned to an offset drain spacer.

Figure 6.10a and 6.10b shows the measured I_d-V_g characteristics from the two on wafer splits. The drain current is independent of the gate voltage for all transistors with drain aligned to the gate edge and the drain current shows high I_{off} (at Vg = 0V) but some modulation with gate voltage for the transistors with the drain aligned to a 50nm spacer. From the figure below clearly offsetting the drain from the gate edge is a better design.



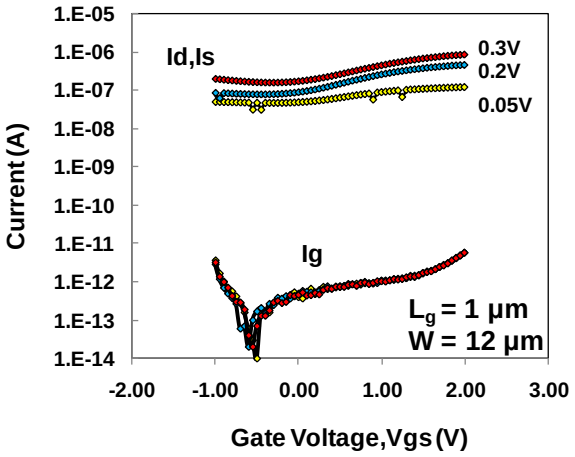


Figure 6.10a I_d-V_g characteristics of TFETs with drain aligned to the gate 6.10b I_d-V_g characteristics of TFETs with drain aligned to the spacer.

Since devices with gate length of 1um, 5um and even 10um without a drain spacer seem to show drain current with no gate voltage dependence, it seems unlikely that this can be attributed to pushing the drain farther from the source and averting the formation of an N+ - P+ junction.

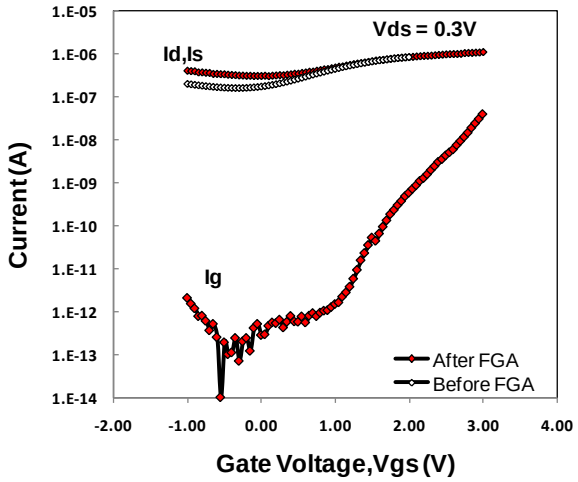


Figure 6.11 I_d-V_g characteristics of a TFET with offset spacer before and after 350 ^{0}C FGA anneal.

The wafer was subjected to FGA at 350 $^\circ$C for 30 minutes to try and improve the dielectric interface quality. The drain current with increasing positive gate voltage remained unchanged after FGA (Figure 6.11) while the drain current with increasing negative gate voltage was larger after FGA than before. Also the drain current seemed to increase slightly with increasing negative gate voltage indicating P-type TFET turn on characteristics are likely superimposed on the N-type TFET characteristics. Since TFETs are ambipolar, the onset of trap assisted or band to band tunneling in the N+ region overlapped by the drain would cause the P-type TFET turn on, causing the drain current to increase with increasing negative voltage.

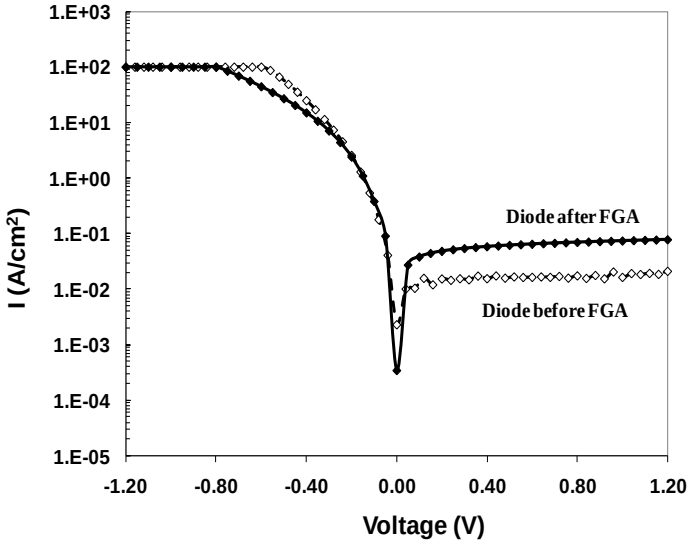


Figure 6.12 P^+/N diode characteristics before and after 350 $^\circ$C FGA. The reverse diode leakage is seen to increase after FGA.

Figure 6.12 compares P^+-N diode characteristics from the wafer before and after FGA. The reverse bias diode leakage is found to increase after FGA. Also superimposed on the diode leakage curves are two I_d-V_d curves from a 1um gate length TFET with an offset drain spacer.

Figure 6.13 shows the I_d-V_g characteristics for multiple drain voltages and I_d-V_d characteristics for multiple gate voltages for a TFET with gate length of 1um. In literature it is well known there exists a large density of interface states close to the valence band of Ge which makes modulation of the surface potential of N-doped Ge by application of a negative gate voltage (bands bend up) better than that of the P-doped Ge by application of a positive gate voltage (bands bend down)[6.12]. This can possibly explain why the P-type TFET turn on is more obvious than the N-type TFET turn on Fig.

At V_g = 0V, with no onset of tunneling the I_d-V_d characteristic for all positive drain voltages should have been identical to the diode leakage and therefore fairly constant for increasing drain voltages. The drain current is however seen to increase tremendously with increasing drain voltage. This indicates the presence of some generation mechanism (mechanism 1) other than just the normal reverse bias diode leakage. The I_d-V_d curve with positive TFET gate voltages is seen to be larger than that at V_{gs} = 0V at very low drain bias, indicating the onset of some gate induced tunneling/other generation mechanism which is overshadowed very quickly (by V_{ds} = 0.5V) by mechanism 1.

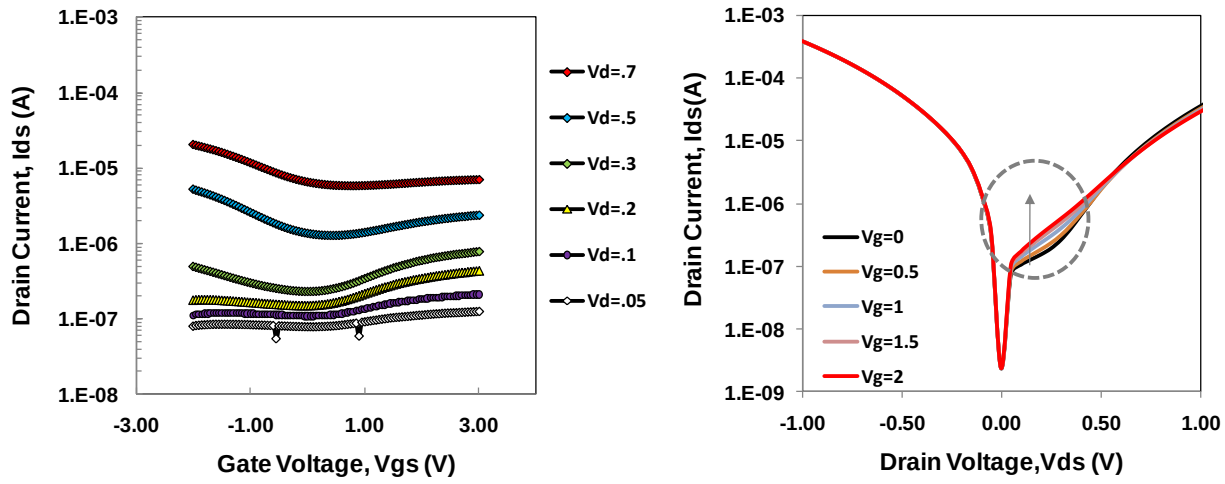


Figure 6.13 I_d-V_g for multiple drain voltages and I_d-V_d for multiple gate voltages for TFET with drain spacer.

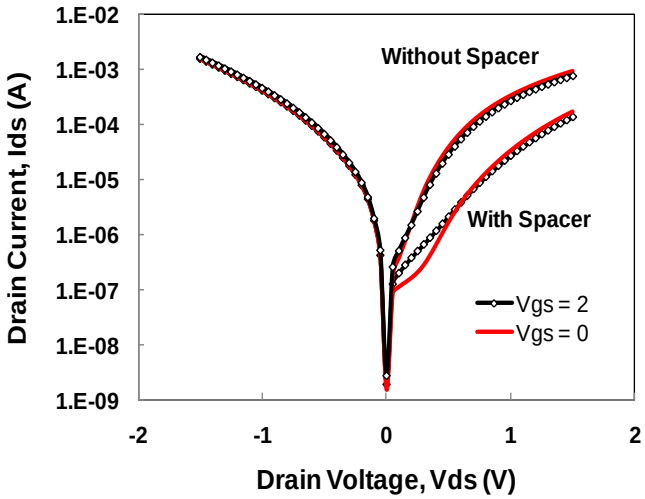


Figure 6.14 TFET I_d-V_d characteristics for two different gate voltages for TFETs with and without offset drain spacer. For positive drain voltages, the ideal TFET drain current would be reverse bias diode leakage at V_{gs} = 0V. The increase in drain current with positive drain voltage is indication of an unwanted carrier generation mechanism. The spacer pushes out the drain implant and seems to reduce this generation mechanism indicating that the generation is possibly in the N+ drain region under the gate.

Figure 6.14 compares the I_d-V_d characteristics at V_{gs} = 0V and 2V, for 1um gate length TFET with offset drain spacer and without offset drain spacer. For the TFETs with drain aligned to the gate edge the drain current at positive drain voltages is significantly larger and completely independent of gate voltage even at small drain voltages.

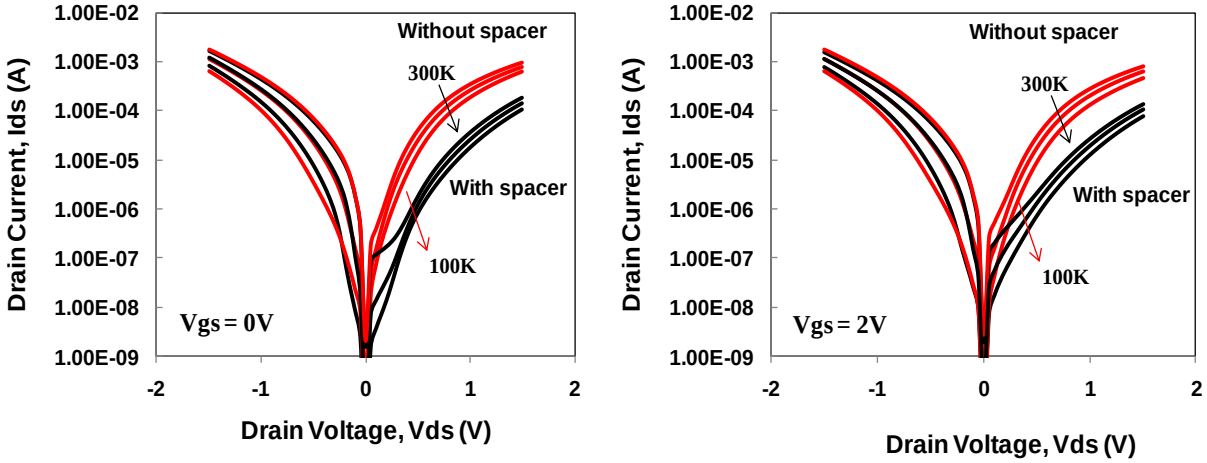


Figure 6.15a Impact of temperature on I_d-V_d characteristics for TFET with and without spacer at V_{gs} = 0V; 6.15b Impact of temperature on I_d-V_d characteristics for TFET with and without spacer at V_{gs} = 2V. Drain current from TFETs without spacer fairly insensitive to temperature. The drain current from TFETs with spacer is fairly sensitive to temperature at low drain voltages.

Figure 6.15a and 6.15b shows the impact of temperature on (L_g = 1um) TFET I_d-V_d characteristics. For the TFETs with drain aligned to the gate edge, drain current is fairly insensitive to temperature. For the TFETs with drain aligned to the spacer edge, the drain current at V_{gs} = 0V is very sensitive to temperature at low drain voltages and then fairly insensitive to temperature. The same trend is noticed for V_{gs} = 2V except that the sensitivity to temperature at low drain voltages is lesser than that seen for V_{gs} = 0V. The different sensitivity to temperature in the different regions is an indication of different carrier generation mechanisms at low and high drain voltages.

To try and understand the mechanism at low drain voltages a little better the activation energy for the drain current at various drain voltages is extracted at both V_{gs} = 0V and V_{gs} = 2V.

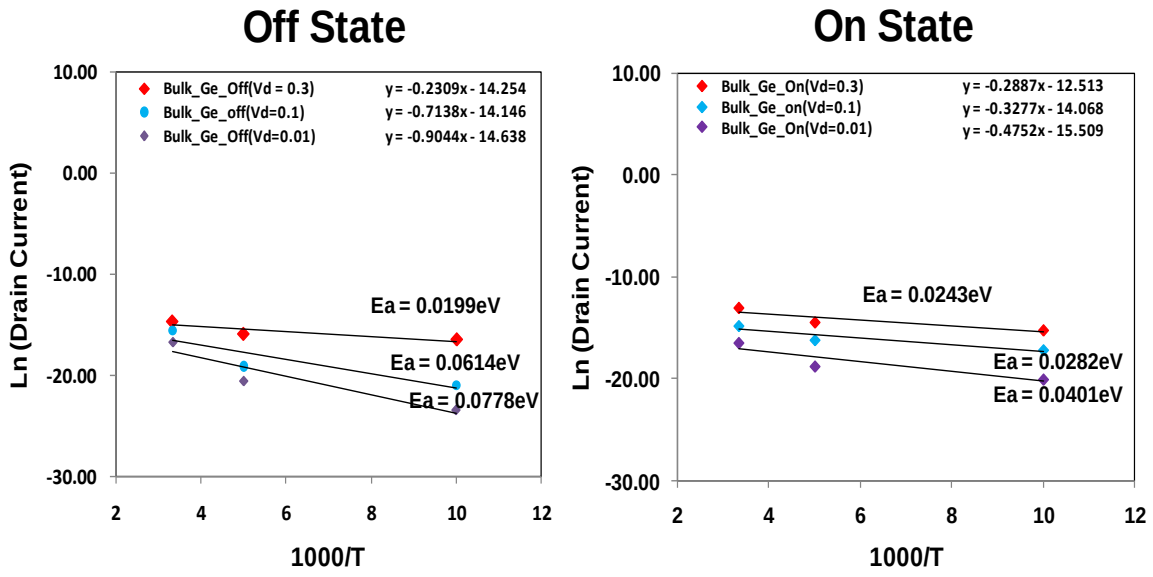


Figure 6.16 Activation energy of drain current from TFETs with spacer in the off state (V_{gs} = 0V) and on state (V_{gs} = 2V) for low drain voltages.

Figure 6.16 shows natural log of the TFET drain current vs. 1000/ T at drain voltages of 0.01V, 0.1V and 0.3V. The activation energy for very low drain voltages of 0.01V and 0.1V at V_{gs} = 0.V, is found to be ~ 0.1eV, while at 0.3V it is found to be 0.02eV. At V_{gs} = 2V, the activation energy for all drain biases is ~0.02eV. At V_{gs} = 0V, if the drain current was due to reverse bias diode leakage, the activation energy would be ~ E_g. This indicates that the TFET drain current at V_{gs} = 0 is not due to diode leakage and is possibly due to the P-type TFET turn on.

To truly see the N-type or P-type TFET turn on at low voltages and ensure that the drain current at V_{gs} = 0V is entirely due to diode leakage, the drain needs to be implanted aligned to a much wider spacer (~0.5µm) to ensure that even after annealing, the N^+ dopant diffusion underneath the gate is minimal.

Several crucial aspects of fabricating hetero TFETs with bulk germanium substrates have been experimentally studied. Since this was the first study using bulk germanium wafers at the

UC Berkeley fabrication facility, several fabrication aspects were experimented with and new procedures that work with germanium substrates were successfully established. The knowledge from the experimental modules can be used to successfully fabricate hetero TFETs on bulk germanium substrates to try and understand the true nature and benefits of hetero BTBT turn on characteristics.

6.5 REFERENCES

[6.1] H. Shang, H. Okorn-Schmidt, K. K. Chan, M. Copel, J. A. Ott, P. M. Kozlowski, S. E. Steen, S. A. Cordes, H. S. P. Wong, E. C. Jones, W. E. Haensch, "High mobility p-channel germanium MOSFETs with a thin Ge oxynitride gate dielectric," Electron Devices Meeting, 2002. IEDM '02. Digest. International, vol., no., pp. 441- 444, 2002

[6.2] Huiling Shang, *Member, IEEE,* Kam-Leung Lee, P. Kozlowski, C. D'Emic, I. Babich, E. Sikorski, Meikei Ieong, H.-S. P. Wong, *Fellow, IEEE*, Kathryn Guarini, and W. Haensch, "Self-Aligned n-Channel Germanium MOSFETs with a Thin Ge Oxynitride Gate Dielectric and Tungsten Gate", Electron Device Letters, IEEE, vol.25, no.3, pp. 135- 137, March 2004

[6.3] Chi On Chui, H. Kim, D. Chi, P. C. McIntyre, K. C. Saraswat, "Nanoscale germanium MOS Dielectrics-part II: high K gate dielectrics," Electron Devices, IEEE Transactions on, vol.53, no.7, pp.1509-1516, July 2006

[6.4] Nan Wu, Qingchun Zhang, D. S. H. Chan, N. Balasubramanian, Z. Chunxiang, "Gate-first Germanium nMOSFET with CVD HfO_2 gate dielectric and silicon surface passivation," Electron Device Letters, IEEE , vol.27, no.6, pp. 479- 481, June 2006

[6.5] A. Satta, E. Simoen, R. Duffy, T. Janssens, T. Clarysse, A. Benedetti, M. Meuris, and W. Vandervorst, "Diffusion, activation, and regrowth behavior of high dose P implants in Ge", Applied Physics Letters, 88, 162118, 2006.

[6.6] A. Satta, E. Simoen, T. Clarysse, T. Janssens, A. Benedetti, B. De Jaeger, M. Meuris, W. Vandervorst, "Diffusion, activation, and recrystallization of boron implanted in preamorphized and crystalline germanium," Applied Physics Letters, vol.87, no.17, pp.172109-172109-3, Oct 2005

[6.7] H. Shang, M. M. Frank, E. P. Gusev, J. O. Chu, S. W. Bedell, K. W. Guarini, M. Ieong, "Germanium channel MOSFETs: Opportunities and challenges," IBM Journal of Research and Development, vol.50, no.4.5, pp.377-386, July 2006

[6.8] L. Rui, S. J. Lee, H. B. Yao, D. Z. Chi, M. B. Yu, D. L. Kwong, "Pt-Germanide Schottky source/drain Germanium p-MOSFET with HfO_2 gate dielectric and TaN gate electrode," Electron Device Letters, IEEE, vol.27, no.6, pp. 476- 478, June 2006

[6.9] R. Li, H. B. Yao, S.J. Lee, D. Z. Chi, M.B. Yu, G.Q. Lo and D. L. Kwong, "Metal-germanide Schottky Source/Drain transistor on Germanium substrate for future CMOS technology", Thin Solid Films, Vol. 504, Issues 1-2, 10 May 2006, Pages 28-31

[6.10] M. L. Lee, E. A. Fitzgerald, M. T. Bulsara, M. T. Currie, and A. Lochtefeld, "Strained Si, SiGe, and Ge channels for high-mobility metal-oxide-semiconductor field-effect transistors", J. Appl. Phys. 97, 011101 (2005)

[6.11] F.E. Ley,R. Bonzom, R. Loo, O. Richard, B. De Jaeger, J. Van Steenbergen, K. Dessein, T. Conard, J. Rip, H. Bender, W. Vandervorst, M. Meuris, M. Caymax, "Epitaxy solutions for Ge MOS technology", Thin Solid Films, Volume 508, Issues 1-2, June 2006, pp 292 – 296

[6.12] A. Dimoulas, A. Tsipas, A. Sotiropoulos, E. K. Evangelou, "Fermi-level pinning and charge neutrality level in germanium," Applied Physics Letters, vol.89, no.25, pp.252110-252110-3, Dec 2006

Chapter 7

Conclusion

7.1 SUMMARY

To combat with the increasing power density that comes along with scaling and increased speed and circuit density, a low power alternative to MOSFETs is needed. Gate induced Band to Band tunneling based transistors (TFETs) are a heavily researched alternative because of their potential to achieve much steeper than 60mV/decade turn on. In this work TFETs are researched to understand their potential to enable low voltage operation. Starting with the simplest gate diode TFET design, experimental silicon TFETs are implemented to confirm characteristics seen in TCAD simulations. Beyond the simplest design two novel design concepts are explored in great detail to determine the ultimate potential of TFETs as ultra low voltage transistors:

1) Dopant pocket engineering (lateral pocket adjacent to the source)
2) Biaxial strain engineering to achieve ultra low effective tunneling bandgap

The lateral pocket TFET design, operation and optimization are performed using TCAD simulations and finally an experimental implementation is presented. The fabricated LPTFET was far from the ideal LPTFET design because of the use of ion implantation and anneal to create the pocket and therefore the lack of precise control of the pocket thickness and doping. Regardless experimental verification of the enhanced behavior of the LPTFET over a control TFET with no pocket was achieved.

Biaxial strain engineering for ultra low effective bandgap hetero TFETs was explored in great detail. Based on literature, a systematic search was performed to determine the most optimal strain system for lowest effective bandgap for both N and P-type TFETs. It was found that achieving an ultra low effective bandgap for N-type hetero TFETs is far easier than for P-type hetero TFETs. With careful use of a vertical dopant pocket, or a more complex double strained layer system, P-type TFETs which are complementary to N-type TFETs can be achieved. In this work using TCAD simulations which integrate strain and dopant engineering it seems possible to achieve N-type hetero TFETs with 10pA of leakage and $I_{on}/I_{off} = 4.6x10^{7}$ in 0.4V V_{dd} and complementary P-type TFETs with $I_{on,NFET}/I_{on,PFET} = 1.2$. Complimentary hetero TFETs with such high performance in 0.4V V_{dd} seems extremely attractive from a power saving point of view.

Using existent biaxially strained heterostructure MOSFETs, GIDL current is measured carefully to understand and experimentally observe the nature of hetero BTBT. Experimental confirmation of the concept of reduced effective tunneling bandgap across a strained silicon cap on relaxed germanium heterostructure is presented. Given the large junction leakage from the

gate first processed hetero MOSFETs, a gate last hetero TFET process was implemented and found to be unsuccessful. In order to understand the true nature of hetero TFET turn on characteristics at low current ranges without the turn on being overshadowed by enhanced junction leakage, bulk germanium substrates are considered as a viable alternative to implement N-type hetero TFETs. Since there has been no prior processing experiment with bulk germanium substrates at UC Berkeley, process development for several modules was required. A working gate stack with ALD Al_2O_3 and in-situ doped LPCVD poly $Si_{0.5}Ge_{0.5}$ was developed.

A process to grown a thin strained silicon cap layer on germanium substrates was established in collaboration with Lawrence Semiconductor Research Lab. And finally homojunction gate first processed germanium TFETs were fabricated and analyzed. While the true N-type tunneling turn on characteristics at low currents were masked by the P-type turning on, the experiment was extremely valuable to help debug all the various process modules required to fabricate N-type hetero TFETs. The contact to germanium using a thin TiN barrier before sputtering Aluminum was found to work well. Also the use of a thin ALD spacer after gate etch was found to protect the gate edges and keep the gate leakage current low. Lastly the need for a very wide offset drain spacer to misalign the drain side from the gate edge and ensure no P-type turn on was the most valuable lesson learned from the experiment. Unlike in silicon, the P-type turn on is made worse in germanium by the fact that the surface modulation of N-doped germanium is better than P-doped germanium because of the large density of interface traps close to the valence band edge of germanium. Therefore it is essential to ensure that no N-type dopants diffuse underneath the gate to ensure that the off state is determined by reverse bias diode leakage.

7.2 Suggestions for Future Work

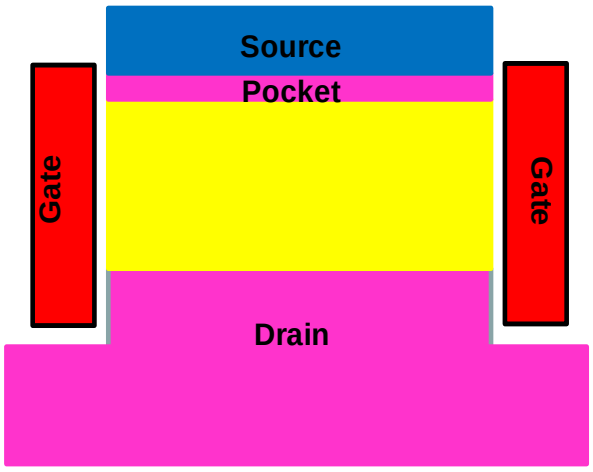


Figure 7.1 Vertical TFET with a lateral pocket. The drain, channel, pocket and source are all grown with in-situ epitaxy for precise control of the thickness and doping concentration of the pocket.

From simulations lateral pockets in TFET can be engineered to achieve very steep TFET turn on characteristics. The limitation of creating these pockets by ion implantation and anneal is the lack of abruptness in the doping profile and the lack of precise control over the pocket thickness. This can be achieved very easily with a vertical TFET configuration as shown in Figure 7.1. In this design the source, pocket, channel and drain are all grown with in-situ doped epitaxial layers. This design also allows for easy integration of a heterostructure at the source-pocket interface which would provide a reduced effective tunneling bandgap for enhanced performance at ultra low voltages.

With any chosen process flow to make pocket TFETs, leakage current through the gate terminal needs to be minimized to be able to study the true characteristics of tunneling current close to the turn on voltage. For this, any process should involve a gate etch with high selectivity to the dielectric beneath the gate electrode. Low selectivity leads to punching through the dielectric and etching into the source and drain regions which is undesirable. Further after the gate etch, the edges need to be healed and if ion implantation is used to introduce dopants the edges need to be protected before the implantation. If epitaxial growth of TFETs and ion implantation are not preferred solid source diffusion can be used to introduce dopants into the source and drain without damaging the silicon substrate or the gate edges. Care needs to be taken to ensure that before the deposition of the solid source such as PSG or BSG, the surface of the silicon substrate is cleared of any residual oxide which might not allow the dopants to penetrate into the substrate. While from a damage annealing point of view solid source diffusion is probably a good option for N-type hetero TFETs, it is unclear what the effects will be on strain relaxation of the strained silicon capping layer.

While chapter 6 dealt with the module development for fabricating N-type hetero TFETs, the same is yet to be done for P-type hetero TFETs. As mentioned in chapter 6 in order to create symmetric P and N-type hetero TFETs, P-type TFETs can either involve:

1) Strained germanium cap on silicon substrate along with a pocket or
2) Strained germanium layer on strained silicon layer on $Si_{0.5}Ge_{0.5}$

Growth of a thin strained germanium cap on silicon needs to be developed. The critical thickness for germanium on pure silicon is ~1nm and further since the growth temperature for strained germanium on silicon substrate is around 350 $^{\circ}$C, all processing after the growth of germanium needs to stay below this temperature to ensure no strain relaxation. A couple of angstroms of silicon need to be deposited on the germanium layer to form a better dielectric interface. This process again needs careful development to ensure that no strain relaxation occurs while passivating with silicon. Studies have shown that the passivating layer can even be amorphous and still form better dielectric interface with high-K then pure germanium. Since effective dopant activation in silicon occurs at very high temperatures, a gate last process which allows for dopant implant and anneal before the germanium layer growth maybe better for P-type hetero TFETs.

Strained Silicon (silicon strained to relaxed $Si_{0.5}Ge_{0.5}$) on insulator wafers are currently available for research purposes. Starting with these substrates would allow for very easy fabrication of P-type hetero TFETs by the growth of a strained germanium layer on the s-SOI wafer. Since the germanium is strained to the relaxed $Si_{0.5}Ge_{0.5}$ lattice constant, the critical thickness of the germanium layer is ~5nm making growth of this s-germanium layer far easier than the growth of s-germanium on a silicon substrate, which involves a much larger lattice

constant mismatch. Also since the wafer consists of just the s-silicon which was lattice matched to the $Si_{0.5}Ge_{0.5}$ and not the $Si_{0.5}Ge_{0.5}$ itself, enhanced junction leakage issues when using relaxed $Si_{0.5}Ge_{0.5}$ substrates are eliminated, making this ideal to study the nature of hetero TFET turn on.

A lot of the simulations in this work have explored the potential of TFETs without regard for technological issues since TFETs are relatively new and their physics and optimization is of great importance. In pushing the boundaries of TFET design, it seems possible to achieve complementary TFETs with 10pA of leakage and $I_{on}/I_{off} \sim 10^7$ in 0.4V V_{dd}. While TFETs have been heavily researched in the last couple of years, lack of easy availability of some of the advanced technologies required to demonstrate ultra steep turn on TFETs has been a huge limitation. Future technological advances will hopefully be leveraged in the coming years to make ultra low voltage operation TFETs a reality.

Appendix A

Process Integration of Silicon TFETs

Step	Process Name	Tool and Recipe	Process Specification
1	Wafers		100nm Si on 200nm Box (6" SOI wafers)
2	Alignment Mark Lithography	SVGOAT6, SVGDEV6, ASML	9000Å resist, E = 24, rotation = -0.5
3	Alignment Mark Etch	LAM5	Oxide Breakthrough to etch through silicon and box
4	Resist Strip	Matrix	2min 30s
5	Alignment Mark Measurement	ASIQ	Ensure step height of 1200Å or more
6	Wafer cleaning	Sink 6	Piranha + DI Water
7	Body Thinning Oxidation	Tystar 2	2Dryoxa, 800 C
8	Wet etch	Sink6	Piranha + DI water + 25:1 HF Dip (check for dewetting)
9	Etch inspection	Nanoduv	Measure to ensure desired body thickness before gate oxidation. Repeat steps 2-4 is desired thickness not reached.
10	Gate Pre Clean	Sink6	RCA 1 clean HF bath + RCA 2 Clean HF bath + 25:1 HF dip + DI Water
11	Gate Oxidation	Tystar1	TCA Clean (8hours) + 1Gateoxa (800 °C, 3min + 900° C N_2 Anneal, 20min)
12	Oxide Measurement	SOPRA	
13	Gate Electrode Deposition	Tystar10	10sdplya, 1hour (in-situ N+ doped poly silicon)
14	Gate Lithography	SVGOAT6, SVGDEV6, ASML	9000Å resist, E = 19, rotation = -0.5
15	Gate etch	Lam 5	3s OB + 10s ME + OE (20%) {Ringfet process and so no OE needed to clear spacers on the sides of the MESA.}
16	Etch inspection	Nanoduv	Measure body thickness to ensure no substrate etch during gate etch
17	Resist Strip	Matrix	2min 30s
18	Polymer removal	Sink7	100:1 HF, 12s
19	Wafer Clean	Sink6	Piranha + DI water

20	Source Implant Lithography	SVGOAT6, SVGDEV6, ASML	9000Å resist, E = 19, rotation = -0.5
21	Inspection	LEO	Check alignment to gate
22	Source Implant	Core Systems	Bf2, 10keV, 7e15
23	Resist Strip	Matrix	2min 30s
24	Wafer Clean	Sink6	Piranha + DI water
25	Drain Implant Lithography	SVGOAT6, SVGDEV6, ASML	9000Å resist, E = 19, rotation = -0.5
26	Inspection	LEO	Check alignment to gate
27	Drain Implant	Core Systems	As, 10keV, 4e15
28	Resist Strip	Matrix	2min 30s
29	Wafer Clean	Sink6	Piranha + DI water
30	ILD deposition	Tystar11	11SULTOA, 8 min, 1000Å
31	Dopant Anneal	Heatpulse 4	500C 30s intermediate step; ramp to 1020 °C in 3s
32	Contact lithography	SVGOAT6, SVGDEV6, ASML	9000Å resist, E = 24, rotation = -0.5
33	Contact etch	Centura-MXP	MXP_OX_ET_EP (20% OE)
34	Oxide removal	Sink7	100:1 HF dip, 5s
35	Metal deposition	Novellus	Sputter etch + 1000 Å Al-2%Si deposition
36	Metal lithography	SVGOAT6, SVGDEV6, ASML	9000Å resist, E = 16, rotation = -0.5
37	Metal etch	Centura-MET	Aluminum ME (until no shiny film seen on wafer)
38	Resist Strip	Matrix	2min 30s
39	Wafer clean	Sink8	Water rinse
40	FGA anneal	Tystar 18	H2SINT400, 30min

Appendix B

Process Integration of Bulk Germanium TFETs

Step	Process Name	Tool and Recipe	Process Specification
1	Wafers		6” Bulk Ge wafers
2	Gate Pre Clean	Sink6	Cyclic HF clean (end with HF and make sure wafer dewets)
3	Gate Dielectric Deposition	Picosun	50 cycles of std Al_2O_3 recipe at 300°C (gives 5nm)
4	Dielectric Measurement	SOPRA	
5	Gate Electrode Deposition	Tystar19	SGDEPF.019, 425°C, 1hour
6	Alignment Mark Lithography	SVGOAT6, SVGDEV6, ASML	9000Å resist, E = 24, rotation = -0.5
7	Alignment Mark etch	LAM 7	3s OB + ME
8	Resist Strip	Matrix	2min 30s
9	Alignment Mark Measurement	ASIQ	Ensure mark is 1200Å or deeper
10	Wafer cleaning	Sink9	Acetone+ DI Water
11	Gate Lithography	SVGOAT6, SVGDEV6, ASML	9000Å resist, E = 19, rotation = -0.5
12	Gate etch	Lam 7	3s OB + ME (extra 20% etch time is added to endpoint)
13	Resist Strip	Matrix	2min 30s
14	Wafer cleaning	Sink9	Acetone+ DI Water
15	Polymer removal	Sink7	100:1 HF, 12s
16	Gate protection liner	Picosun	30 cycles of std Al_2O_3 recipe at 300°C (gives 3nm)
17	Source Implant Lithography	SVGOAT6, SVGDEV6, ASML	9000Å resist, E = 19, rotation = -0.5
18	Inspection	LEO	Check alignment to gate
19	Source Implant	Core Systems	Bf2 implant
20	Resist Strip	Matrix	2min 30s
21	Wafer Clean	Sink9	Acetone + DI water

22	Drain offset spacer deposition	P5000	PE_THIN_USG (25s for RF to stabilize + 45s deposition with flat facing up; flip wafer and flat faces down, then 25 for RF to stabilize + 45s deposition)
22	Drain spacer etch	Centura-MXP	MXP_OT_ET_EP, etch to leave 200Å of oxide behind (to ensure silicon cap not etched away for HETERO TFETs)
23	Drain Implant Lithography	SVGOAT6, SVGDEV6, ASML	9000Å resist, E = 19, rotation = -0.5
24	Inspection	LEO	Check alignment to gate
25	Drain Implant	Core Systems	As implant
26	Resist Strip	Matrix	2min 30s
27	Wafer Clean	Sink9	Acetone + DI water
28	ILD deposition	P5000	PE_USG_0.5 (25s for RF to stabilize + 75s deposition with flat facing up; flip wafer and flat faces down, then 25 for RF to stabilize + 75s deposition)
29	Dopant Anneal	Heatpulse 4	Ramp to 500°C in 10s, steady at 500°C for required time. Use Germanium TC
30	Contact lithography	SVGOAT6, SVGDEV6, ASML	9000Å resist, E = 24, rotation = -0.5
31	Contact etch	Centura-MXP	MXP_OX_ET_EP (20% OE)
32	Oxide removal	Sink7	100:1 HF dip, 5s
33	Metal deposition	Novellus	Sputter etch +250Å TiN + 1000 Å Al deposition (TiN is required as a diffusion barrier since only Al- 2% Si is available)
34	Metal lithography	SVGOAT6, SVGDEV6, ASML	9000Å resist, E = 16, rotation = -0.5
35	Metal etch	Centura-MET	Aluminum ME (until no shiny film seen on wafer) Etch rate of Al = 2x etch rate of TiN
36	Resist Strip	Matrix	2min 30s
37	Wafer clean	Sink8	Water rinse