Maitra, Kingsuk, *Electron Transport in Bulk-Si NMOSFETs in presence of High-κ Gate Insulator – Charge Trapping and Mobility*

Advisor: Dr. Veena Misra
Industrial Collaborator: Dr. Eduard Cartier, IBM Research

Abstract

Recent advancements in gate stack engineering has led to the development of aggressively scaled, high mobility, high-κ dielectric based NMOSFETs with metal gates. Most of the current literature on the subject also stressed on the need for a high temperature process step to attain the high mobility under minimal change of effective oxide thickness. However, the physical origin of high mobility is not well understood. In this work, fundamental insight into the necessity of the high temperature process step is provided. Novel experimental strategies are developed to understand the impact of interface states and bulk traps separately and exclusively on channel mobility. It is conjectured that the interface states at the SiO_2/(100) bulk-Si interface are identical in nature (as far as coupling with the channel electrons is concerned) to those at the high-κ/SiO_2/(100) bulk-Si interface. Thus, the response of interface states on channel electrons in high-κ insulator based NMOSFETs is properly calibrated by a novel thermal desorption of hydrogen experiment on SiO_2/(100) bulk-Si NMOSFETs to yield a highly accurate parameterized equation. The value of

interface state response parameter determined by the aforementioned experiment is compared with theoretical predictions, and independently determined projections from electrical stress measurements. The impact of transient charging on transport in the channel is investigated. It is conclusively shown that remote charge has minimal impact on mobility in the channel. The role of nitrogen induced fixed oxide charge is studied on a set of Hf-silicate samples. Role of soft optical phonon scattering and the beneficial impact of metal gates on soft optical phonon limited mobility are thoroughly investigated both theoretically and experimentally. Conclusions are drawn on the fundamental limit of mobility attainable in high-κ dielectric based NMOSFETs.

Electron Transport in Bulk-Si NMOSFETs in presence of High-κ Gate Insulator – Charge Trapping and Mobility

By

Kingsuk Maitra

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

ELECTRICAL ENGINEERING
Raleigh, North Carolina
October, 2005

APPROVED BY:

_____

CARL OSBURN

_____

DOUGLAS BARLAGE

_____

MARK JOHNSON

_____

VEENA MISRA
Chair of advisory committee

To

My Parents

Biography

Education

PhD – December 2005, Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC – 27695

Master of Science (Engineering) in Microelectronics – July 2002, Department of Electrical Communication Engineering, Indian Institute of Science, Bangalore, Karnataka, India

Master's Thesis – "Transistor Design Methodology for sub-100 nm CMOS Technology in presence of Gate Oxide Leakage"

Master of Science in Physics – August 2000, Department of Physics, University of Kalyani, Kalyani, West Bengal, India

Bachelor of Science (Honors) in Physics – July 1998, University of Calcutta, Calcutta, India

Professional Experience

Three Industrial Summer Internships (2003, 2004, 2005) at IBM T.J Watson Center, IBM Corporation, Yorktown Heights, NY-10598

Currently at Portland Technology Division, Intel Corporation, Hillsboro, Oregon

Acknowledgements

I would like to thank my thesis advisor, Dr Veena Misra for her guidance and support throughout the course of this work. I would always cherish the liberty and independence she gave me to pave my own path. Wherever, she gave me an inch, I tried to go a full mile. I sincerely hope I lived up to her expectations.

I would like to thank my committee members Dr Carl Osburn, Dr Doug Barlage, and Dr Mark Johnson for taking time out of their busy schedules to critique this work.

None of this would be possible without the active collaboration and participation of the high-κ team at the IBM T. J Watson Research Center at Yorktown Heights. Dr Eduard Cartier (Eddie), my mentor, friend and philosopher stood beside me like a rock these three long years and constantly guided me throughout my formative years as a graduate student. I would like to thank him for believing in me, and believing in this project. Words would always fall short while describing Eddie's invaluable contribution to this thesis work.

I would like to thank Raj (Dr Raj Jammy), my manager, for giving me the break when I needed it.

I would like to thank Dr Max Fischetti for helping me with the theoretical component of this thesis. Dr Vijay Narayanan deserves a special mention for the samples used in this work. Thank you, Vijay for helping me whenever I needed it. I would also cherish the interactions with Dr Martin M Frank and having the opportunity to work with him. I would like to thank Dr Evgeni Gousev for supporting me throughout my stay at Yorktown and beyond.

Dr Barry P. Linder, Gill Singcko, Dr Vamsi Paruchuri were always there whenever I needed them. I would also like to thank Dr Wilfried Haensch, Dr Paul Solomon, Dr Arvind Kumar, and Dr Zhibin Ren for the invaluable comments and suggestions which enriched the content of this thesis.

I would also like to acknowledge my friends and fellow interns especially Saibal Mukherjee, Dr Kaushik Das, Dr Soumyadip Ghosh, Dr Syed Alam, and Anindita Basu who made my stay at Yorktown a very memorable one.

I would like to thank Bei Chen, Tuhin Maity, Dr Debdip Bhattacharyya and Jaehoon Lee for making my stay at NC-state University a very rewarding and pleasurable one.

I would like to thank my parents for being an integral part of my quest and endeavor right from the word go. I am what I am today because of their constant encouragement and unflinching support throughout my early years.

I would like to thank National Science Foundation (NSF) and IBM Corporation for funding this work.

And last but not the least, a special word of thanks to my lovely wife Sujata, fellow graduate student at NC-state, who stood beside me through thick and thin and never let me lose sight of my goal. I definitely overachieved when I married her.

November 2005

Raleigh, NC-27695

TABLE OF CONTENTS

LIST OF TABLES

List of Figures

“The woods are lovely dark and deep

But I have promises to keep

And miles to go before I sleep”

-Robert Frost (1874 – 1963)

Chapter 1

Introduction

1.1 Introduction

Continuous downscaling of MOS transistor gate length/channel length's requires extremely thin gate oxides for (i) improvement in vertical oxide electric field leading to increase in inversion charge and subsequent decrease in channel resistance [1] which leads to increase in on-current and hence improvement in performance (ii) better short channel control, since short channel control is determined by the relative thickness of gate dielectric to Si depletion width [2]. However, extremely thin gate oxides lead to runaway direct tunneling leakage through the gate oxide [3]. This in turn leads to unacceptably high static power dissipation which seriously affects the functionality and reliability of a chip. As a specific example, the International Technology Roadmap for Semiconductors (ITRS) predicts sub-1 nm electrical oxide thickness (EOT) gate oxides for the 32 nm technology node and below [4], gate oxide thicknesses at which direct tunneling leakage will be greater than 10, 000 A cm^{-2} [5]. The only way to contain this high gate leakage is to use a thicker gate dielectric layer without changing the value of the EOT. This necessitates the replacement of SiO_2 with a high-κ insulator. The motivation behind using high-κ insulator in advanced MOS transistors is discussed in more detail in the following section.

1.2 Why High-κ?

Elementary electrostatics predicts that the physical thickness of a high-κ dielectric can be connected to an equivalent electrical oxide thickness (EOT) through the following simple mathematical relationship,

$$\varepsilon_{Ox} / T_{Ox} = \varepsilon_{Hi\text{-}\kappa} / T_{Hi\text{-}\kappa} \quad (1)$$

, terms being usual. Clearly, the simple relationship stems from constancy of gate oxide/dielectric capacitance (C_{Ox}), before and after replacement of SiO_2 by high-κ. Since, $\varepsilon_{Hi\text{-}\kappa}$ is greater than $\varepsilon_{Ox,}$ the above equality holds if and only if $T_{Hi\text{-}\kappa}$ is greater than T_{Ox}. This is pictorially surmised in Figure 1.1.

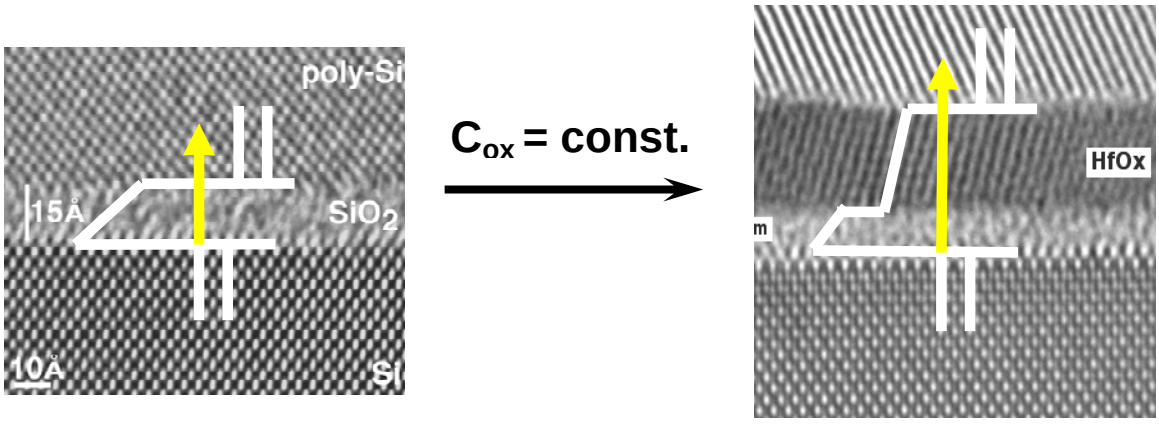


Figure 1.1. Why High - κ? A set of representative TEMs showing the increase in $T_{Hi\text{-}\kappa}$ with replacement of SiO_2 with Hf based high-κ gate stack [6]

This definitely comes about with marked decrease in gate leakage as depicted in Figure 1.2, where a one-to-one comparison is made between gate leakage behavior of a SiO_2 dielectric layer with a similar EOT representative high-κ Al_2O_3 gate stack. A 10^4 X

reduction in leakage current is observed. Evidently, higher the magnitude of the dielectric constant (κ or ε), higher is $T_{Hi-\kappa}$, and larger is the reduction in gate leakage.

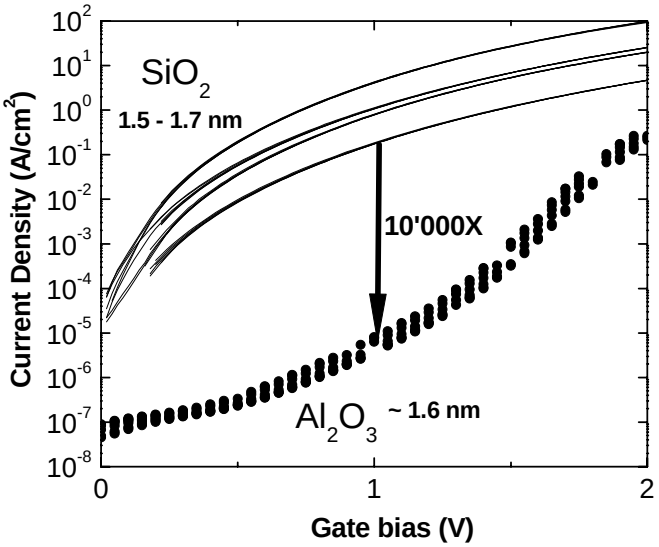


Figure 1.2. Around 10,000 X reduction of leakage current observed with Al_2O_3 gate dielectric over SiO_2 at a gate bias of 1 V [6]

1.3 High-κ and Metal Gates

However, the paradigm shift from SiO_2 to high-κ is not the only major change that needs to be made to enable sub-1 nm oxide scalability. Poly crystalline Si (Poly-Si) gate, the gate electrode material of choice for the chip industry for the last two decades, is incompatible with aggressively scaled high-κ gate stacks [7]. Primarily, because of poly depletion, a phenomenon which occurs in MOS transistors during inversion, a severe EOT penalty has to be paid if high-κ were to be integrated with Poly-Si. So, solely from a device design perspective, the sub-1 nm EOT as required by ITRS [4] for sub-32 nm gate length, may be extremely difficult to achieve. To get around this problem, a metal gate

needs to be used with high-κ. Metal gates have high carrier concentration and thus may be used to eliminate the impact of EOT increase with poly-depletion. This is clearly shown in Figure 1.3, where the same leakage level with the lowest possible EOT is obtained for a metal gate (TiN)/HfO_2 gate stack [7].

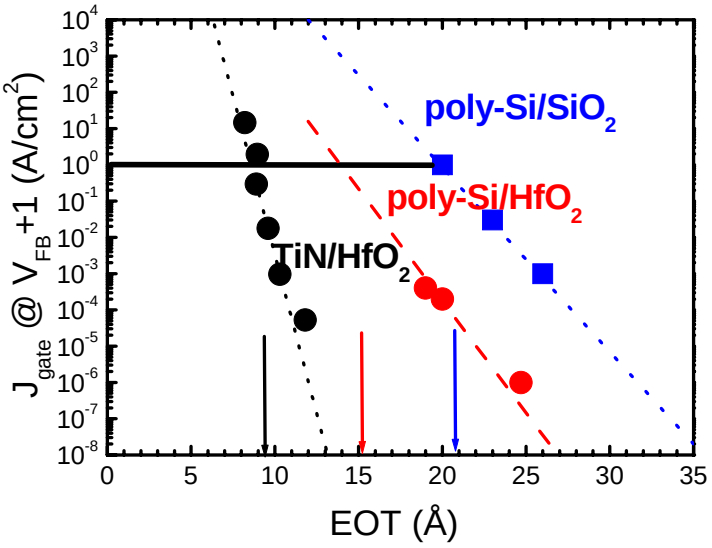


Figure 1.3. At the same gate leakage level (1 A cm^{-2}) @ (V_{FB}+1) V gate voltage, the lowest EOT is achieved for metal gate (TiN)/HfO_2 stack as opposed to a Poly-Si/HfO_2 stack because of absence of poly depletion with the former [7].

However, high-κ and metal gates have their problems when they are integrated with bulk-Si devices. In this work, we focus on one such problem, namely mobility degradation in bulk-Si NMOSFETs in presence of high-κ insulator. Fundamental insight into this problem will be provided, and a discussion on the physical origin of high mobility in aggressively scaled metal gate/high-κ gate stacks, as recently reported [8] will be thoroughly investigated. In the next section, we cite the different physical mechanisms

discussed in literature which cause mobility degradation in the bulk-Si channel in presence of a high-κ. Each of those mechanisms is then discussed briefly based on the current state-of-the-art. Also, the beneficial impact of replacing conventional poly-Si gate electrode with an ideal metal gate (in the context of mobility degradation) is discussed. In each of the following sections of this chapter, we also give an overview of how each of those topics relate to the rest of the thesis.

1.4 Mobility degradation mechanisms

Mobility degradation mechanisms in high-κ insulator based MOSFETs may be classified into two broad categories: (i) Extrinsic like charge trapping (ii) Intrinsic like soft optical phonon scattering. As the name suggests, the former may be controlled or minimized by careful process optimization, while the latter is related to the basic physics of the high-κ insulator system.

1.5 Charge Trapping

The charge trapping related mobility degradation or additional scattering mechanisms may be further sub-divided into the following two categories:

(i) Scattering from the interface states, henceforth referred to as N_{it}.

(ii) Scattering from remote charge or "charge at a distance", which may include but not limited to fixed oxide charge from impurities at the interface, or, compensated charge in the gate stack, which may not be detectible by independent electrical measurements like flatband/threshold voltage shift [9].

Going back to the first topic of N_{it} limited scattering, we take guidance from extensive work done on the topic in SiO_2 based NMOSFETs. It is instructive to note that charge trapping belongs to the larger category of the Coulomb scattering limited mobility degradation mechanisms which also include ionized impurity scattering in the bulk-Si channel from the dopants in the channel. However, here the emphasis is on those charge trapping related mobility degradation mechanisms which are exclusive to the high-κ gate insulator. It has been pointed out in literature quite extensively, that the high-κ/bulk-Si system comes very often with a large number of N_{it} [6, 10]. So a very careful investigation of this scattering mechanism is necessary when dealing with the mobility problem in high-κ gate insulators. Substantial insight into the problem has been provided by W. Zhu et. al [10] from a characterization standpoint. From a fundamental standpoint, the Coulombic term has been thoroughly investigated in works such as [11]. However, these treatments are extremely complex, and do not provide closed form solutions to the Boltzmann transport equation, which forms the basis of any mobility study in MOSFET structures. To gain a phenomenological insight into the problem, we adopt an approach which is very similar to the one used by Villa et. al. [12] . This approach also helps us to separate out the impact of ionized impurity limited scattering and scattering from the N_{it} within the Coulombic mobility term very easily. This will be dealt with in more detail in Chapter 4.

The influence of remote charge in the high-κ insulator (or "charge at a distance" to distinguish it from the N_{it} which are right adjacent to the channel electrons) on the

channel mobility is largely based on the work of Yang et. al [13], Saito et. al [14], and more recently Gamiz et. al [15] . Though they investigated the role of remote charge in the depleted poly-Si gate on the channel mobility, it may be extrapolated to study the impact of remote charge in the high-κ insulator itself on the channel electrons, since the mechanism of coupling of the "charge at a distance" with the channel electrons is through remote Coulomb scattering (RCS) mechanism. An attempt to study this phenomenon experimentally in detail, has been attempted by R.J Carter et. al., L. A Ragnerssen et. al, and E. A Cartier [16, 17, 6 respectively], where the magnitude of remote charge (in this case fixed oxide charge) was estimated from a V_{FB} vs. EOT plot, and the mobility estimated by split-CV/I_d-V_g technique (Figure 1.4). However, no attempt was made to extend the understanding so obtained to the existing theory of RCS. Also it is not immediately clear what the role of N_{it} is, and how and why the two should be treated differently from a very fundamental perspective. For the first time, we will provide objective experimental evidence on the impact of remote charge in the gate dielectric layer itself on transport in the channel. As pointed out in [15], the proximity of the gate (poly-Si or metal) decides the strength of coupling of the remote charge with the channel electrons. This is because the gate provides a ground plane where the charge in the dielectric may be screened. Larger the screening, lower is the strength of coupling with the channel electrons [15]. By using a systematic set of samples, where the distance of the poly-Si gate is pushed farther and farther away from the channel by using an insulator capping layer (viz.: Al_2O_3), we try to provide an empirical understanding of the role of screening by the gate on the RCS. This will be discussed very briefly in Chapter 3.

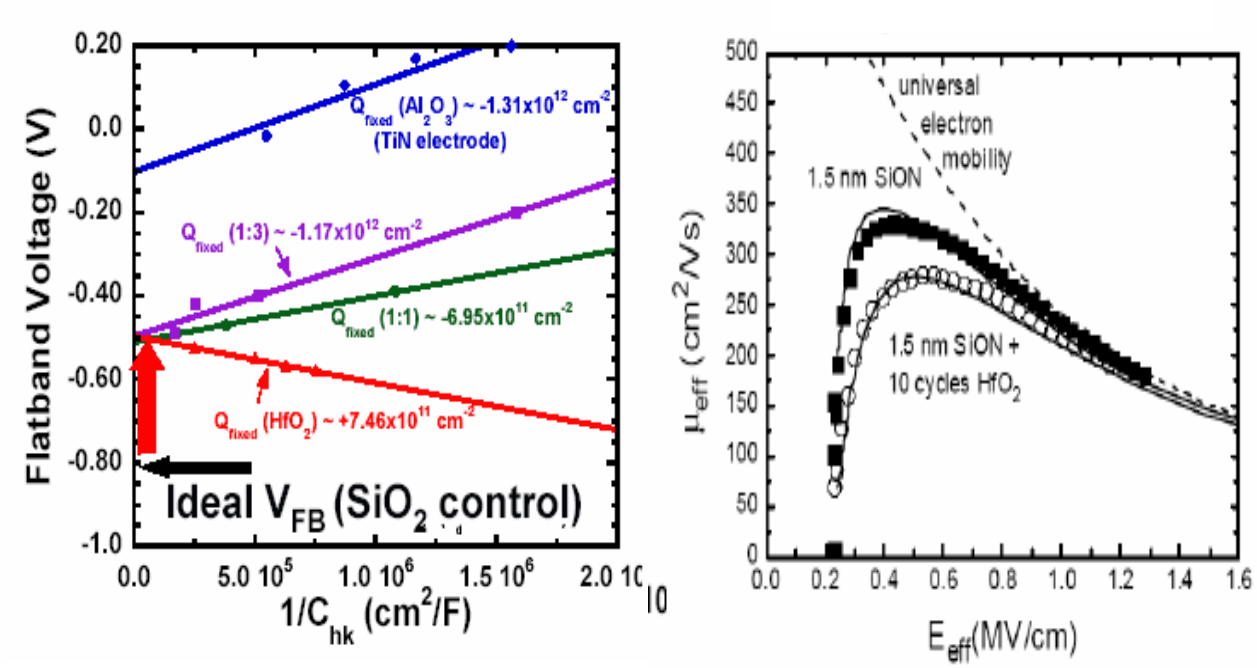


Figure 1.4. Estimation of fixed oxide charge from V_{FB}-1/C_{hk} (EOT) plots, Degradation in mobility in a HfO_2 based gate oxide stack due to the possible presence of fixed charge [6].

Another RCS limited mobility degradation mechanism floated recently by Hiratani et. al. [18] is possible coupling of the channel electrons with compensated charge in the dielectric layer itself. They pointed out that the entire mobility degradation cannot be explained by charge scattering induced by the charge present in stack which may be detectible by voltage shifts obtained by conventional electrical measurements. In order to explain the additional decrease of mobility, Hiratani et. al [18] proposed that there is an additional compensated charge located in the dielectric layer itself, this is schematically described in Figure 1.5. The problem with this scheme is that an independent experimental confirmation would be almost impossible to come up with.

And a back of the envelope estimate of the amount of compensated charge that may be necessary to bring about a mobility degradation cited in [18] is on the order of 10^{13} cm^{-2} , an extremely large quantity of charge solely because the mechanism of coupling between channel electrons and compensated charge is through dipole scattering, a coupling which is substantially weaker than RCS induced by unipolar charge.

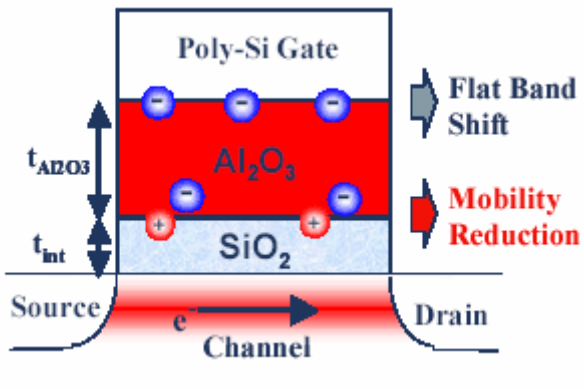


Figure 1.5. Schematic illustrating the compensated charge hypothesis, unipolar charge can cause both mobility reduction and flatband voltage shift but compensated charge can cause mobility degradation without affecting flatband or threshold voltage.

The impact of bulk trapping, or transient charge trapping unique to high-κ insulator systems will also be discussed in Chapter 3. It will be shown fast trapping has negligible impact on mobility.

This brings us to the next high-κ specific scattering mechanism namely that of soft optical phonon scattering limited mobility, an intrinsic scattering mechanism which is guided by the basic physics of the high-κ insulator system.

1.6 The role of Remote Phonon Scattering

In 2001, M.V Fischetti et. al [19] pointed out that there are some fundamental limitations to the maximum mobility attainable in high-κ insulator based NMOSFETs. The dielectric constant of a non-metallic solid stems from two factors: the ionic and the electronic polarization. Pure insulators, by definition have large bandgap and since the electronic polarization is roughly proportional to the square of the direct bandgap of the solid averaged over the first Brillouin zone [19], there is little that can be done to tune this component. Naturally, the high permittivity of a high-κ dielectric can come only from the large ionic polarization, often due to the highly polarizable ('soft') metal-oxygen bonds. Associated with soft bonds are low energy "optical phonons" which cause severe remote phonon scattering. By contrast, the hard Si-O bonds in SiO_2 yield reduced ionic polarization. Associated with "hard" bonds, are stiff optical phonons. Expectedly, this scattering component will be less in SiO_2 based NMOSFETs. In Figure 1.6, theoretical estimates of mobility in the inversion layers of MOS systems are depicted for different high-κ dielectrics. Anisotropic scattering with acoustic phonons and remote scattering with surface optical phonons are accounted for. Surface roughness and ionized impurity scattering have been incorporated using simple Matthiesen's rule as described in Figure 7. As predicted, higher the dielectric constant, worst is the mobility degradation from the

SiO_2 control. It was also observed that in dual layer stacks, higher the interfacial oxide thickness, less severe is the mobility degradation. M.V Fischetti et. al [19] conjectured that this is consistent with the remote phonon scattering theory. Theoretical calculations predict that proximity of the high-κ to the inversion channel worsens scattering due to remote phonons.

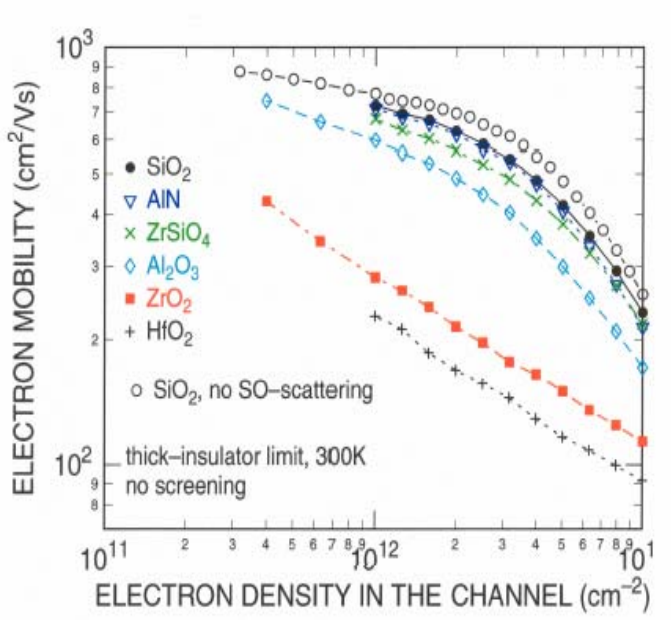


Figure 1.6. Soft optical phonon scattering limited mobility for different high-κ insulator systems. It is instructive to note that higher is the dielectric constant of these stacks, lower is the soft optical phonon scattering limited mobility [19].

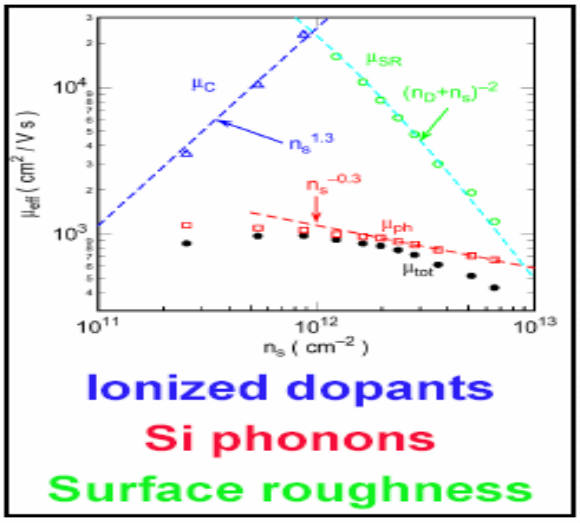


Figure 1.7. Different scattering components merged together by Matthiessen's rule (N_{it} limited scattering factor not included) [19]

Direct experimental evidence on this topic was provided by Zhibin Ren et. al. [20] in IEDM 2003. They noted the following two points:

1. Weak temperature dependence with strong remote phonon scattering was predicted and measured. (Figure 1.8)
2. Hf-silicate which has lower dielectric constant compared to HfO_2 has higher mobility in the intermediate electric field (or inversion charge) regime where phonon scattering is expected to dominate.

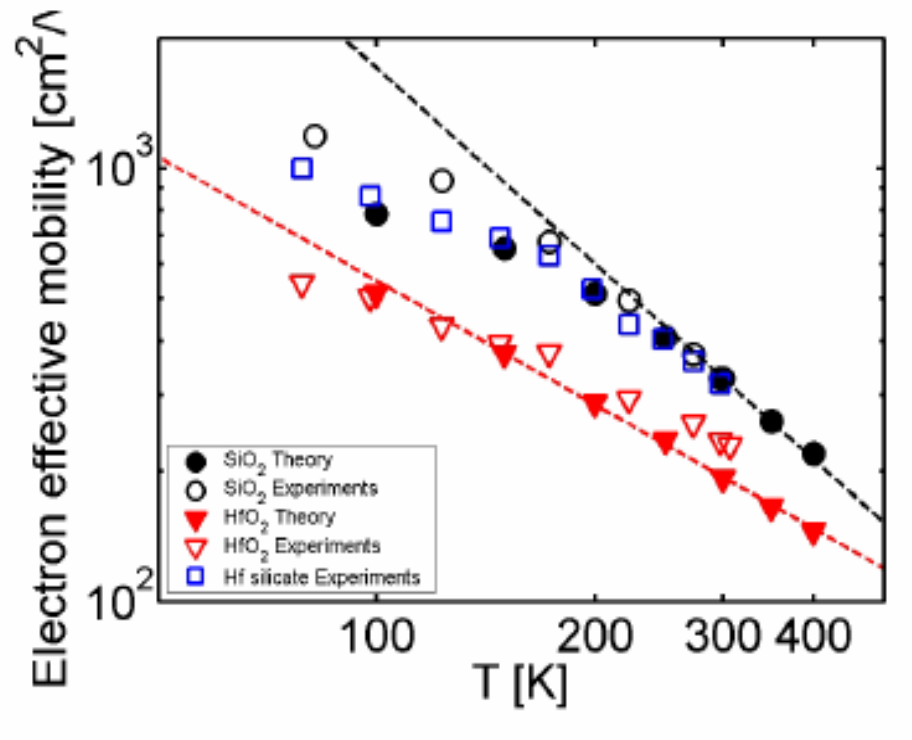


Figure 1.8. Weak temperature exponent of mobility at intermediate inversion charge densities (N_{inv}= 3 X 10^{12} cm^{-2}) for high-κ suggesting dominance of optical phonons which predictably have weak temperature dependence [20].

It is instructive to note that temperature dependence will be used copiously in Chapter 5 while investigating physical origin of high mobility in advanced high-κ gate stacks.

1.6 High Mobility in High-κ/Metal gate stacks – Role of Metal gate Screening

In 2003, R. Chau et. al. demonstrated high mobility in aggressively scaled (T_{inv} of 1.45 nm) HfO_2/TiN stacks [8]. The reason for high mobility in these stacks was attributed to metal gate screening of some of the soft phonon modes in the high-κ insulator by the ideal metal gate (TiN in this case) on top. The underlying physics related to this phenomena was pointed out for the first time by A.R Bhatt et. al in 1993 [21] in polar semiconductors (as opposed to high-κ insulator) sandwiched between a metal layer and a bulk non-polar semiconductor in a classic MSS structure. Using simple electrostatics and the dielectric continuum model, they showed disappearance of the transverse optical phonon modes in a MSS structure. It should be noted at this point that a polar semiconductor can be easily replaced by a high-κ insulator without making any major changes in the formalism proposed. Particularly, from a phonon spectrum standpoint, both can be approximated by a linear chain of dissimilar atoms which leads to emission in the optical region of the spectra. The phonon dispersion relations for an ionic insulator such as HfO_2 are very similar to that of an ionic semiconductor such as GaAs. From a comparison of the dispersion relations of an ideal metal/high-κ/bulk-Si system with the non-ideal metal or poly-Si/high-κ/bulk-Si , it is clear that disappearance of some of the optical phonon modes with the latter has profound implications on transport of electrons in the channel. For instance, in a metal gate/high-κ NMOSFET, the optical phonon-electron coupling in the channel will be significantly reduced and this is what is observed experimentally as shown below (Figure 1.9).

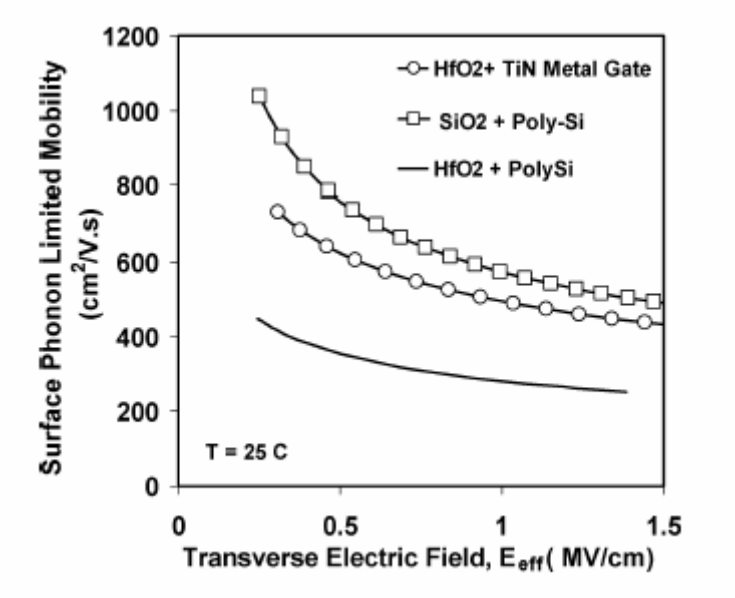


Figure 1.9. Comparison of surface phonon limited mobility of a HfO_2/TiN gate stack with HfO_2/poly-Si gate stack. Clearly, the HfO_2/TiN stack is seen to recover some of the mobility lost by remote optical phonon scattering with the HfO_2/poly-Si stack.

In course of this thesis, we will perform novel ultra low temperature measurements on aggressively scaled, high mobility metal gate HfO_2 stacks to study for the first time the impact of metal gate screening on soft optical phonon limited mobility in high-κ/ metal gate stacks. Next, we discuss the physical origin of high mobility in aggressively scaled metal gate/high-κ stacks.

1.7 Physical origin of high mobility in High-κ gate stacks

It is worth pointing out here that though [8] reported high mobility in high-κ gate stacks, no information was given as to the process steps that led to such high mobility in

advanced high-κ gate stacks. Very recently, however, A. Callegari et. al. [22] gave some insight into the high mobility phenomenon in aggressively scaled high-κ/metal gate (in this case, W/HfO_2) stacks. It was pointed out that a high temperature process step, post metal gate deposition brings about a structural change in the gate stack itself, leading to formation of a low dielectric constant Hf-silicate layer, which screens the soft optical phonon modes in the HfO_2 layer itself. This leads to significant recovery in the mobility with the metal gate/high-κ gate stack itself (Figure 1.10). The problem with this scheme is that under high temperature treatment , an EOT penalty (about 0.2-0.4 nm) is paid which affects overall scalability into the sub-1 nm regime.

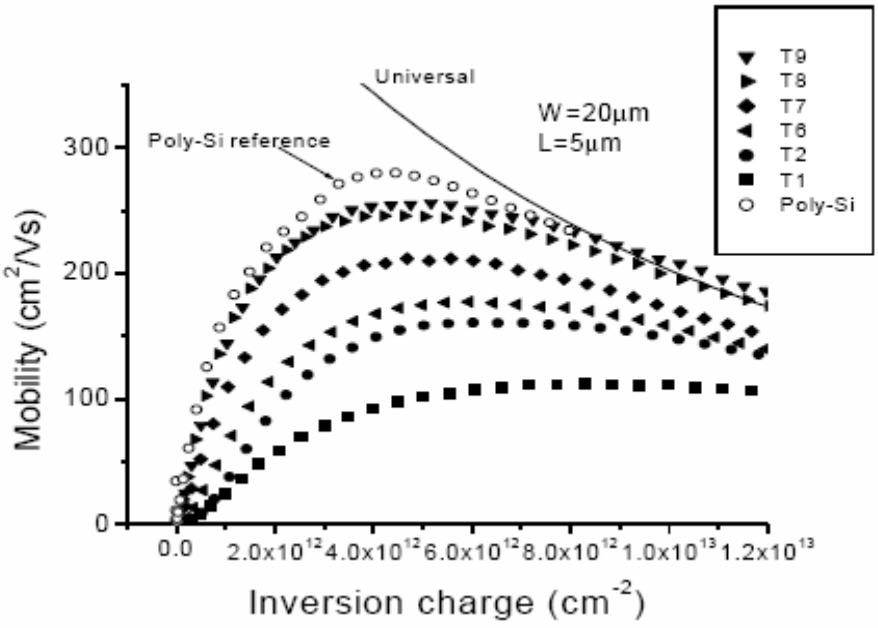


Figure 1.10. High Mobility attained in W/HfO_2 gate stack after high temperature treatment (T9>….>T1, T-temperatures) [22].

Very recently, however, L. A Ragnarsson tried to investigate the role of the high temperature process step in attaining high mobility in metal gate/high-κ gate stacks [23].

Using SPER (solid phase epitaxial regrowth) technique, a low temperature activation anneal was used to form the source/drain (600 ^{0}C) junction, thereby ensuring that the high-κ/metal gate stack is protected from high temperature treatment. However, the high mobility so attained is low compared to those values reported in [8]. So the necessity of a high temperature processing step cannot be excluded. However, there are exceptions. Very recently, in VLSI 2005, Y Akasaka et. al. demonstrated high mobility in $TaSi_X/HfO_2$ stack NMOSFETs [24] under low processing temperature conditions. Though the physical origin of this high mobility was not well understood, but it has been attributed to possible nitrogen profiling in the gate stack itself.

1.8 Scope and Organization of the Thesis

From the foregoing discussion, it is clear that recent developments in gate stack engineering has led to the development of aggressively scaled, high mobility, high-κ dielectric based NMOSFETs with metal gates. Most of the current literature on the subject also stressed on the need for a high temperature process step to attain the high mobility provided there is minimal change of effective oxide thickness. However, the physical origin of high mobility is not well understood. In this thesis, fundamental insight into the necessity of the high temperature process step is provided. Novel experimental strategies are developed to understand the impact of interface states and bulk traps separately and exclusively on channel mobility. Role of soft optical phonon scattering and the beneficial impact of metal gates on soft optical phonon limited mobility are thoroughly investigated both theoretically and experimentally. Conclusions are drawn on the fundamental limit of mobility attainable in high-κ dielectric based NMOSFETs.

In Chapter 2, the amplitude sweep charge pump technique, an useful charge trapping characterization technique as it applies to high-κ gate dielectrics, is discussed. The inversion charge pump based mobility characterization technique is also discussed, and a novel variant to the one and two contact correction scheme is proposed. The split-CV/dc I_d-V_g conventional mobility estimation technique is also compared against the non-standard inversion charge pump technique to identify the best split-CV (frequency) for use in Chapter 4 to estimate mobility in presence of varying interface state densities (N_{it}).

In Chapter 3, a novel experimental strategy is developed on a controlled set of samples to get a semi quantitative estimate of the location and the magnitude of trapped charge. Through a simple electrical stress experiment, bulk charge or "charge at a distance" is introduced and its impact on mobility studied.

In Chapter 4, a thermal desorption experiment on SiO_2/Si NMOSFETs is carried out to controllably introduce N_{it} in SiO_2/Si NMOSFETs and its impact on mobility studied. Conclusions are drawn on its applicability to high-κ based advanced gate stacks. A highly accurate N_{it} scattering limited mobility estimate is also provided.

In Chapter 5, we study the aggressively scaled high mobility metal gate/high-κ gate stacks. The role of metal gate screening on the soft optical phonon modes is investigated theoretically. Ultra low temperature measurements are performed on the above mentioned high mobility stack to separate out the different scattering mechanisms.

Conclusions are drawn on the theoretical limit of the high mobility attainable in high-κ metal gate stacks.

In Chapter 6, we conclude our work. Future directions are chalked out. Novel experimental methodologies enabling direct measurement of metal gate screening are proposed. Lastly, the feasibility of high-κ gate dielectrics making its way into mainstream Si electronics is discussed.

References

[1]. Scott E Thomson, Robert S Chau, Tahir Ghani, Kaizad Mistry, Sunit Tyagi, and Mark T Bohr, "In Search of "Forever" Continued Transistor Scaling One New Material at a Time", IEEE Transactions on Semiconductor Manufacturing, Vol. 18, No. 1, February 2005, pp. 260 – 265.

[2]. Yuan Taur, " CMOS Scaling Beyond 0.1 μm: How far can it go? ", Proc. 1999 International Symposium on VLSI Technology, Systems, Applications (Cat. No. 99TH8453), 1999, pp. 6 – 9.

[3]. K. F Schuegraf, C. C King, and C. Hu, " Ultra-thin dioxide leakage current and scaling limit", VLSI Technology Symposium Digest of Technical papers, IEEE, 1992, pp. 18 – 19.

[4]. The International Technology Roadmap for Semiconductors (2003 edition), available online @ http://public.itrs.net

[5]. Yuan Taur, and Tak. H Ning, "Fundamentals of Modern VLSI Devices", Cambridge University Press, 1999 (Chapter 2, p. 96).

[6]. Eduard Cartier, 2004 Reliability Physics Tutorial Notes, IEEE International Reliability Physics Symposium, Topic 134.

[7]. W. Tsai et. al., "Comparison of sub-1 nm TiN/HfO_2 with Poly-Si HfO_2 gate stacks Using scaled chemical oxide interface", VLSI Technology Symposium Digest of Technical papers, IEEE, 2003, pp. 21 – 22.

[8]. R. Chau, S. Dutta, M. Doczy, B. Doyle, J. Kavalieros, and M. Metz , "High-κ/Metal Gate Stack and its MOSFET Characteristics", IEEE Electron Device Letters, 2004, Vol. 25, No. 6, pp. 408-410.

[9]. S. Saito, K. Torri, M Hiratani, and T. Onai, "Improved Theory for Remote Charge scattering limited Mobility in metal-oxide-semiconductor Transistors", Applied Physics Letters, 2002, Vol. 81, No. 13, pp. 2391 - 2393.

[10].W. Zhu, J. P Han, and T. P Maa, "Mobility Measurements and Degradation Mechanism of MOSFETs made with Ultrathin High-K Dielectrics", IEEE Transactions on Electron Devices, 2004, Vol. 51, No. 1, pp. 98 – 105.

[11]. M. V Fischetti, and S. E Laux, "Monte Carlo Study of electron transport in Si Inversion layers", Physical Review B, 1993, Vol. 48, pp. 2244 - 2274.

[12]. Stefano Villa, Andrea L. Lacaita, Laura M. Perron, and Roberto Bez, "A Physically Based Model of the Effective Mobility in Heavily doped NMOSFETs", IEEE Transactions on Electron Devices, 1998, Vol. 45, No. 1, pp. 110 – 115.

[13]. Nian Yang, W. K Henson, J. R Hauser, and J. J Wortman,"Estimation of effects of Remote Coulomb scattering on electron mobility of n-MOSFETs with ultrathin gate oxides", IEEE Transactions on Electron Devices, 2000, Vol. 47, No. 2, pp. 440-447.

[14]. S. Saito, K. Torri, M. Hiratani, and T. Onai, "Improved theory of remote charge scattering limited mobility in metal oxide semiconductor transistors " , Applied Physics Letters, 2002, Vol. 81, No. 13, pp. 2391-2393.

[15]. F. Gamiz, and M. V Fischetti, "Remote Coulomb scattering in MOSFETs: Screening by electrons in the gate", Applied Physics Letters, 2003, Vol. 83, No. 23, pp. 4348-4350.

[16]. R. J Carter et. al., "Electrical Characterization of High-κ materials prepared by atomic layer CVD", in Proceedings of International workshop on Gate Insulators", Tokyo 2001.

[17]. L. A Ragnarsson et. al., "The impact of sub-monolayers of HfO_2 on the device performance of high-κ based transistors", in IEDM Technology Digest, 2003, pp. 4.2.1-4.2.4.

[18]. M. Hiratani, S. Saito, Y. Shimamoto, and K. Torri, " Effective Electron Mobility reduced by remote charge scattering in High-κ gate stacks", Jpn. Journal of Appl. Phys., 2002, Vol. 41, pp. 4521-4522.

[19]. M.V Fischetti, D.A Neumayer, and E. A Cartier, "Effective Electron Mobility in Si Inversion Layers in Metal-Oxide-Semiconductor Systems with a High-K Insulator: The Role of Remote Phonon Scattering", 2001, Journal of Applied physics, Vol. 90 , No. 9, pp. 4587-4608.

[20]. Z. Ren et. al., "Inversion Channel Mobility in High-κ high performance MOSFETs", in IEDM Technology Digest, 2003, pp. 33.2.1-33.2.4.

[21]. A. R Bhatt et. al, "Reduction of Interface Phonon modes using metal-semiconductor heterostructures", Journal of Applied Physics, 1993, Vol. 73, pp. 2338-2342.

[22]. A. Callegari et. al. "Interface Engineering for Enhanced electron mobilities in W/HfO_2 gate stacks", in IEDM Technology Digest 2004.

[23]. Y. Akasaka et. al., " Impact of electrode side chemical structures on electron Mobility in metal/HfO_2 MISFETs with sub 1 nm EOT", in VLSI Technology 2005, pp. 228-229.

[24]. L. A Ragnarsson et. al., " High performing 0.8 nm EOT HfO_2/TaN low thermal budget n-channel FETs with solid phase epitaxially regrown (SPER) junctions", in VLSI Technology 2005, pp. 234-235.

Chapter 2

Charge Pumping and its impact on Mobility Characterization Methodologies

2.1 Introduction

In this chapter, we start by describing in detail a popular charge trapping characterization methodology namely the charge pumping technique. Principally, the amplitude sweep charge pump technique is given the maximum attention, the more conventional base sweep technique is also discussed for the sake of completeness. It is emphasized that with the amplitude sweep charge pump technique, information about both the interface states and traps away from the interface (namely bulk traps) may be studied, whereas base sweep is only limited to providing information about the interface states (N_{it}). It is instructive to note that there is a voluminous amount of literature on this subject, and charge pump technique has become commonplace over the past few years as a very powerful technique of characterizing charge trapping in high-κ gate stacks. In this chapter, only those portions, which we believe has not been dealt with thoroughly before, are emphasized. More importantly, the connection between the said technique and mobility measurement schemes (for instance the inversion charge pump based mobility estimation scheme or the ICP technique) is discussed in detail. A novel variant of the ICP technique-replacing the well known 1 and 2 contact measurement scheme with a 50/100 technique, is proposed. Charge trapping in metal gate based high-κ gate stacks is also discussed in the end. An attempt is made to connect the understanding acquired in studying charge pump characteristics in ordinary high-κ based MOSFETs with more important metal gate/high-κ based high mobility NMOSFETs discussed in Chapter 5.

2.2 Experimental Methodology

Figure 2.1 schematically describes the charge pumping technique. The source and drain of a high-κ based NMOSFET is tied together and grounded. The gate is pulsed from accumulation to deep inversion by applying a pulse train. And the charge pump response is measured by monitoring the substrate current. For a detailed understanding of this scheme, the interested reader is referred to [1]. The amplitude sweep charge pump characteristics as it applies to the high-κ based devices is described in detail in [2]. Interestingly, though the charge pumping effect in MOSFETs was reported in the late 1960s [3], reliable interpretation of the results only came by in the 1980s [4]. As described above, the applied pulse train to the gate drives the surface potential from accumulation into inversion and vice versa. As mentioned in [1], it is often customary to apply a small reverse bias to the source drain junctions tied together, but from our experience we have observed that this leads to significant increase in junction leakage thereby smearing out completely the charge pumping response in some cases. So in the following discussion, all the data shown and described have source drain junctions tied together and grounded. Going back to the discussion on the charge pumping characterization methodology, the charge pumping technique is based on the recombination process of charged interface (surface) states, N_{it} at the Si/SiO_2 or Si/SiO_2/high-κ interface which contributes to the charge pumping current at the substrate. The additional components of the charge pumping current observed specifically with high-κ will be described in detail in the next paragraph. For instance, the physical processes that take place during one pulse cycle regardless of the nature of the gate dielectric, involve drifting back of the inversion carriers; this is critically dependent on

the device geometry, as well as the rise and fall times of the applied pulses as we will find out in the next paragraph. In addition, trapped electron near the conduction band get thermally emitted to the conduction band, and also drift to the source/drain junction, whereas electrons trapped deeper in the bandgap remain captured in the interface traps (this is typically dependent on the lattice relaxation energy of the gate dielectric chosen and is expected to be different for SiO_2 and high-κ [5]). When the hole barrier is reduced, holes accumulate at the Si substrate, and recombine with trapped electrons giving rise to the charge pumping current. As depicted in Figure 2.2, in the amplitude sweep charge pump measurement, the base level of the applied pulse is kept fixed, whereas the amplitude is varied, whereas in base sweep, the amplitude is kept fixed, and the base level is altered.

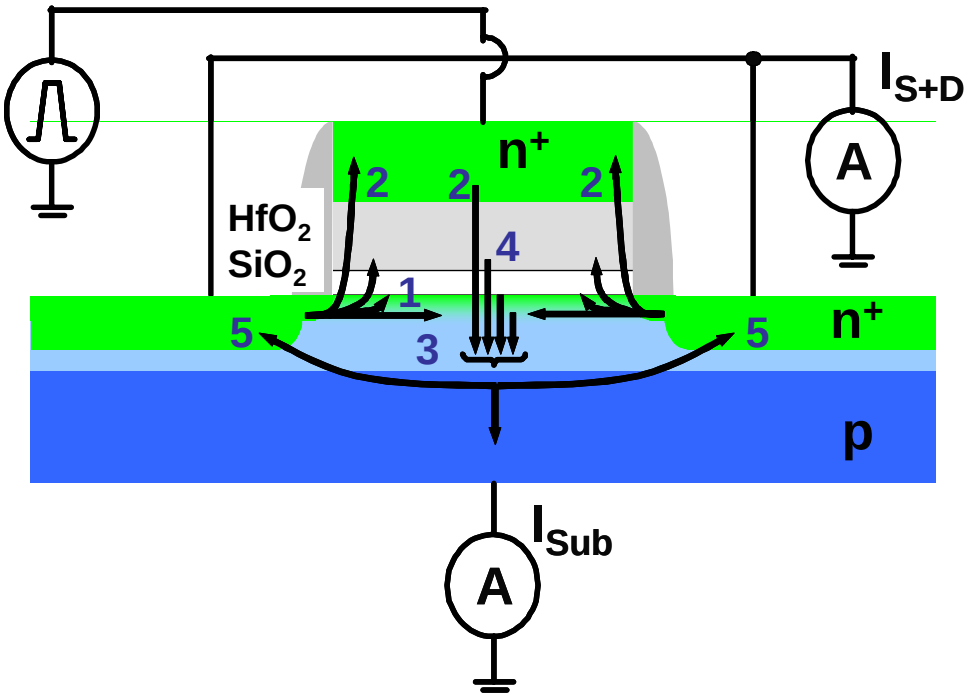


Figure 2.1. Schematic showing the charge pump measurement technique, the numbers and arrows mark the different components of the charge pump current measured at the substrate [2].

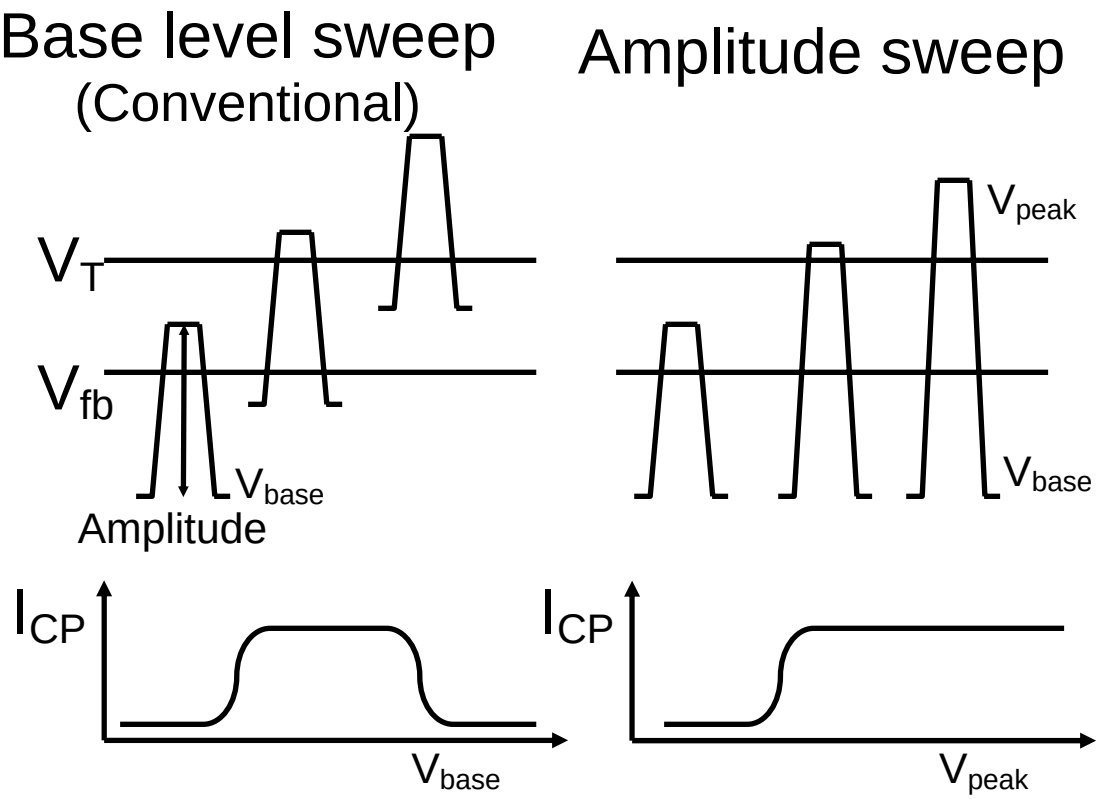


Figure 2.2. Schematic describing the difference between the conventional Base sweep and the Amplitude sweep. Below, the difference between the base sweep and amplitude sweep characteristics for a conventional SiO_2 gate dielectric [2]. (It is worth pointing out that the flat characteristics are not obtained in reality, this is an approximation assuming constant T_{rise} (rise time) and T_{fall} (fall time))

2.3 Charge Pumping as applied to high-κ gate stacks

As pointed out earlier, we will now focus our attention to the different current components giving rise to the charge pumping response at the substrate are described below, it is worth pointing out these are also marked with arrows and numbered in Figure 2.1 [6]. They are classified (as pointed out in [2]) as follows:

(1). Response of N_{it}

(2). Gate Leakage

(3). The geometrical component.

(4). Response of the bulk traps.

(5). Diffusion of injected carriers back to the source/drain junctions.

It is worth noting that the amount of carriers recombining at the substrate is determined by the pulse shape (rise and fall time). Due to the fact that bulk defect density in conventional SiO_2 is very small, no significant contribution from bulk traps to the charge pumping current measured at the Si substrate is expected. If a typical initial trap density of 10^{15} – 10^{16} cm^{-3} is used for thermally grown gate dielectrics, the contribution from bulk oxide traps is expected to be on the order of 10^{9} cm^{-2} [7]. Typical interface trap densities (N_{it}) in MOSFETs with thermally grown gate dielectrics are in the 10^{10} cm^{-2} range as shown in Figure 2.3.

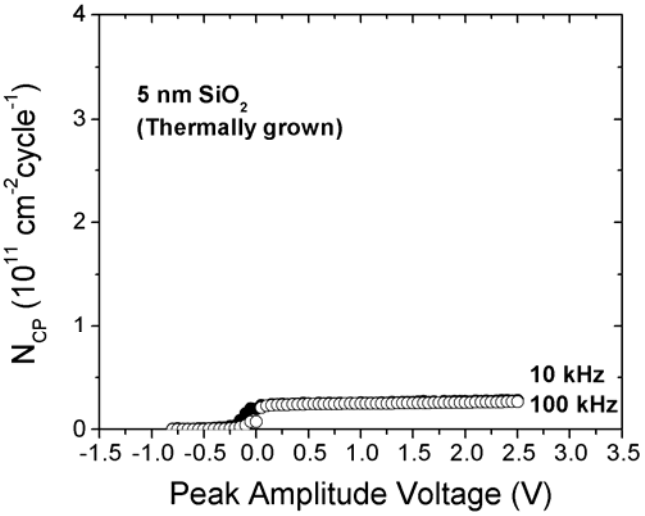


Figure 2.3. Amplitude Sweep Charge Pump Characteristics for a conventional, thick, thermally grown SiO_2 based NMOSFET. Compare and contrast this with Figure 2.4

Thus, due to the low density of bulk traps in conventional gate dielectrics, their contribution to the charge pumping current cannot be easily separated. However, in high-κ gate dielectrics, it is demonstrated as in Figure 2.4 below, the response of the bulk traps cannot be ignored.

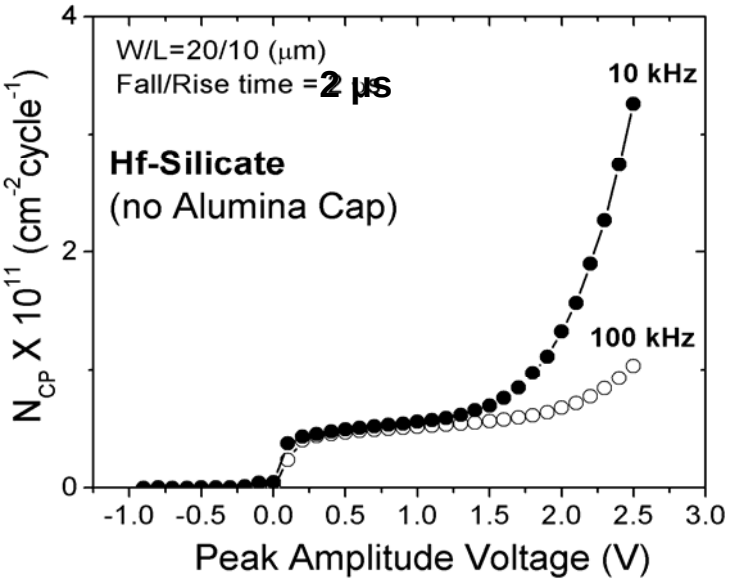


Figure 2.4. Amplitude sweep charge pump characteristics for a representative high-κ based NMOSFET, namely: Hf-silicate deposited by conventional self-aligned CMOS process [2]. These stacks will be revisited in a different context in Chapter 3.

Clearly, when a direct comparison is made between the charge pump characteristics described in Figure 2.3 with the one described in Figure 2.4, we find that the charge pump characteristics in case of the high-κ gate stack shows severe frequency dispersion, this has got to do with the fact that the bulk traps exhibit both temporal and spatial variation. This kind of frequency dispersion is expectedly absent in conventional SiO_2 stack represented in Figure 2.3 where only the interface states are important (N_{it}). It is

instructive to note that higher the dielectric constant of the gate dielectric involved, severe is the bulk trapping. This is very clearly demonstrated in Figure 2.5 where HfO_2 stacks ($\kappa \sim 22$-25) exhibit more bulk trapping than Hf-silicate gate stacks ($\kappa \sim 10$-11). This is because as pointed out in [5], higher the dielectric constant, lower is the dielectric relaxation energy required to re-emit the trapped carrier to recombine with the incoming holes in the substrate. Or in other words, for conventional SiO_2 or SiON based dielectric stacks, this relaxation energy is fairly high, and one needs to go to unreasonably high voltages to observe this kind of trapping. Whereas, for a high-κ based gate stack, this kind of bulk trapping is observed at very low gate biases ($V_g \sim 2$ V). Also as better exemplified in Figure 2.5, the bulk trapping behavior is critically dependent on the interfacial oxide layer thicknesses of the high-κ gate stack under consideration. For instance, the 2.5 nm EOT HfO_2 has lesser bulk trapping than the EOT 1.8 nm one, because the thicker interfacial layer HfO_2 is more SiO_2 like than the thinner one [8]. It is interesting to note here that the Hf-silicate gate stack shows an anomalous rollover behavior at higher gate biases, namely the normalized charge pump current changes direction at around $V_g \sim 1.5$ V or so. **It should be noted that in Figure 2.5, the X-axis represents (V_g+1) V**. This is attributed to tunneling of electrons from the valence band of Si. This is investigated in detail in the subsequent figures. Figure 2.6 monitors the normalized source/drain current along with the normalized charge pump current. Ordinary electron tunneling as also pointed out in [2], should make the charge pump curve to separate out from the source drain current curve as shown in Figure 2.6. And the difference is strongly frequency dependent, tunneling being a slow phenomenon fails to be dominant at higher frequencies (~1 MHz). But the moot point is, the feature observed

in the EOT ~ 2 nm Hf-silicate is different from ordinary tunneling leakage depicted in Figure 2.6.

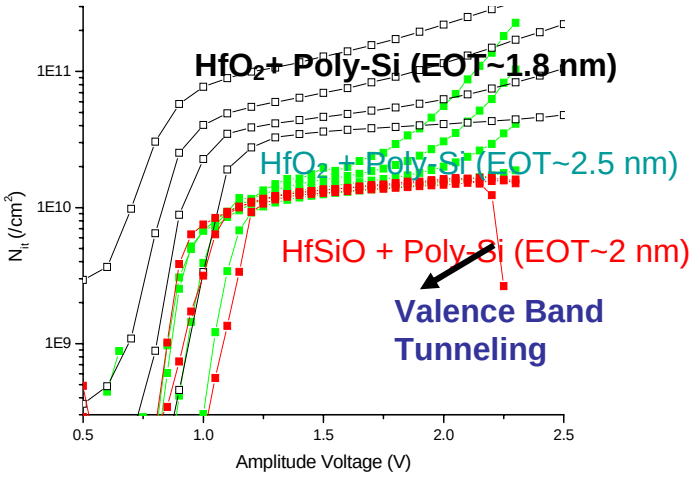


Figure 2.5. Direct comparison of bulk trapping behavior in two representative high-κ gate stacks. All the stacks have poly-Si gate on top and are fabricated by self aligned process [2]. (X-axis: (V_g+1) V)

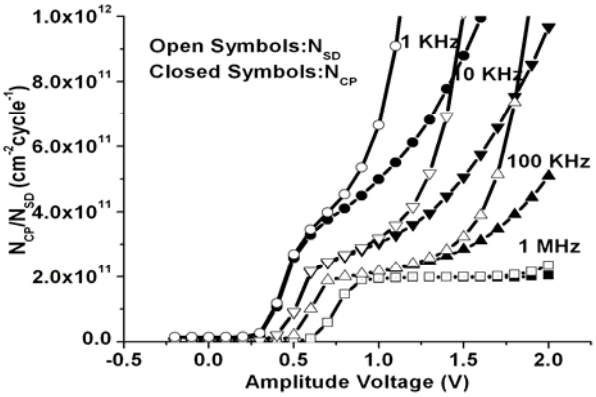


Figure 2.6. Comparison of normalized charge pump and normalized source/drain current. The difference expectedly is strongly frequency dependent.

Interestingly, similar behavior is observed in a thin SiON based NMOSFET as well (Figure 2.7). This is indicative of the fact that this is not a high-κ specific phenomenon at all. For an EOT of 2 nm, the SiON based NMOSFET shows similar turnaround behavior at gate bias of about $V_g \sim 1.5$ V, earlier attributed to valence band tunneling from the Si band. On the other hand, as depicted in Figure 2.8, this is strongly frequency dependent as well, the 10 KHz curves being worst affected by the turnaround behavior, and the 1 MHz curves being the least. This is also indicative of the fact that almost all the features that are attributed to any tunneling phenomenon are present.

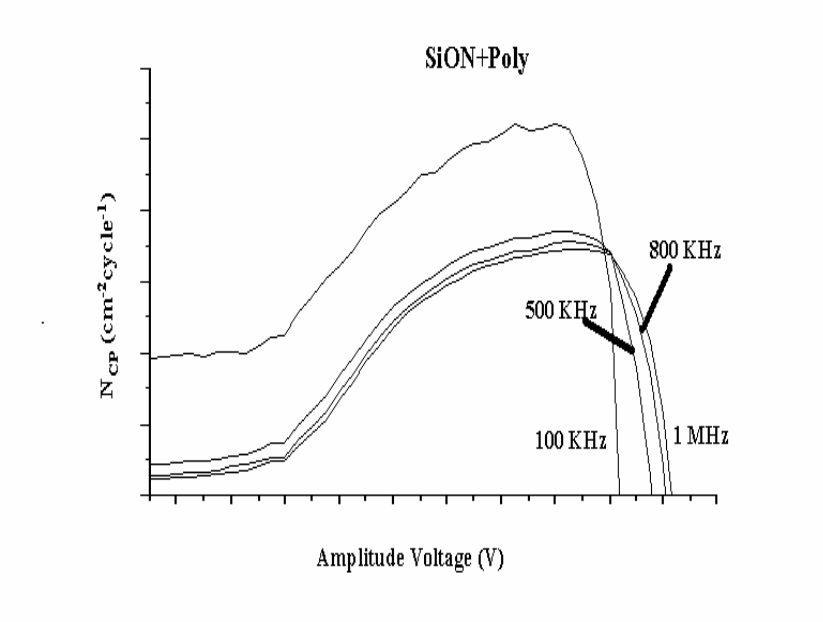


Figure 2.7. An EOT of 2 nm SiON based, poly-Si NMOSFET shows similar turnaround behavior attributed to valence band tunneling leakage.

More evidence on the same are provided by Figure 2.9 where it is clearly observed that if the substrate current is directly measured along with the gate leakage current as a function of gate voltage, around the point where valence band tunneling comes into play,

the substrate current follows the gate leakage (It is instructive to note here that the X-axis of Figure 2.5 is essentially ($V_{g\text{-}peak}$- V_{base}); V_{base} in this case being -1.0 V.)

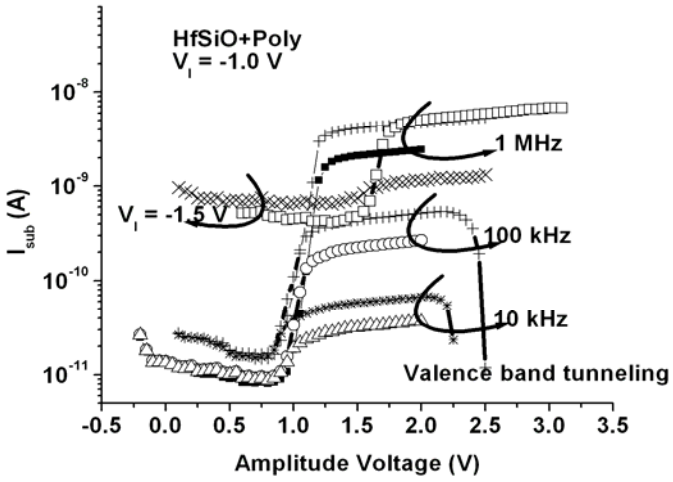


Figure 2.8. Valence band tunneling shows severe frequency dependence consistent with the fact that tunneling is a slow phenomenon.

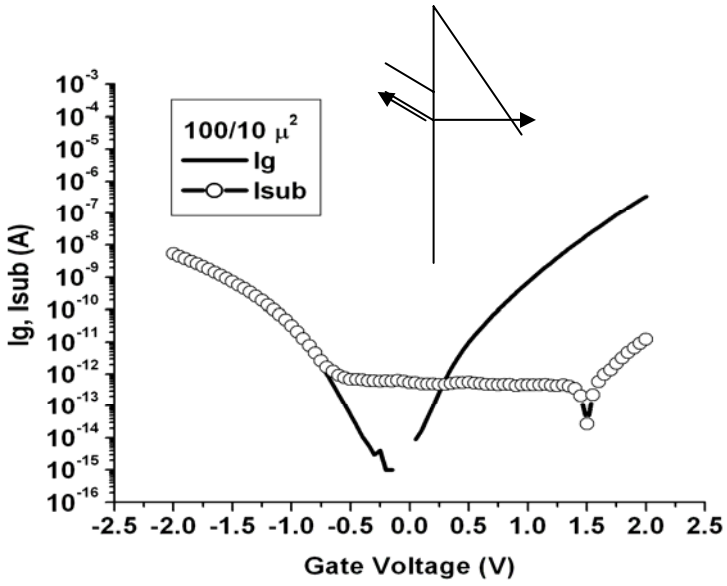


Figure 2.9. By simultaneously monitoring the gate current (I_g) and substrate current (I_{sub}), or charge pump current, it is seen that around V_g~1.5 V, around the bias region where valence band tunneling comes in play, I_{sub} follows I_g.

So far, we have discussed in detail three of the five charge pumping current components mentioned at the beginning of this section-namely, the N_{it} response, the response of the bulk traps, and a specific component of gate leakage (components 1, 2 and 4 mentioned above) viz: the valence band tunneling leakage or electron tunneling from the valence band of bulk-Si in a NMOSFET. The geometrical component was deliberately side stepped because it will be taken up in greater detail in the next section, when we discuss the inversion charge pumping based mobility estimation technique. Before we end this section, we take up the issue of channel length dependence of trapped charge.

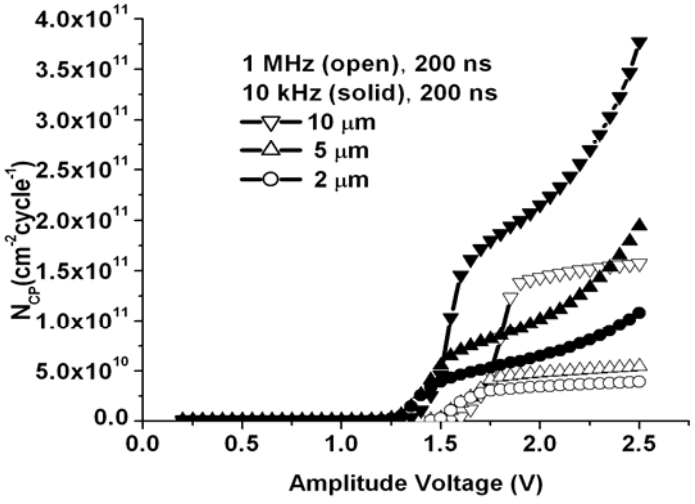


Figure 2.10. Channel length dependence of trapped charge, bulk trapping is found to be less severe in smaller channel length devices than longer channel ones. (HfO_2 + Poly-Si transistor with EOT~2.5 nm vide Figure 2.5)

As depicted in Figure 2.10, we find that a longer channel transistor (10 μm) is more severely affected by bulk trapping (more severe tailing up of the charge pumping curve at higher peak amplitude biases) than shorter channel length (5 μm) transistors. It is worth

mentioning that the minority carriers injected either from the N_{it} or bulk traps have certain lifetime in the bulk-Si channel. In the time range where the emitted carrier remains active, it can either go back to the source drain (where it originated) or can recombine with an incoming hole in the substrate. In the former case, it does not contribute to the charge pump response. Since bulk traps are slow traps (slower than N_{it}), they are likely emitted back into the Si substrate when it is accumulating again (during the downward cycle of the applied pulse at the gate). For a shorter channel transistor, the propensity of the carrier to travel back to the source and drain (source drain back-diffusion [2]) is higher than in a longer channel transistor. Hence the observation in Figure 2.10. Also, for short channel transistors, sidewall oxidation may lead to local thickening of the interfacial oxide layer, thereby contributing to some of the weakening in the trapped charge response, compared to long channel transistors with the same high-κ gate stack.

2.4 Conventional Mobility Estimation Technique – The Split-CV I_d-V_g Technique

Mobility is usually determined by split-CV method, wherein the capacitive response exclusively of the inversion channel is used to estimate the inversion charge [1]. A static I_d-V_g with the drain biased in the linear region determines the drive, and an estimate of the mobility is obtained by combining the two. The carrier mobility so obtained is usually plotted vs. the effective Si field which is either calculated from the substrate doping or obtained from the depletion split-CV. In Figure 2.11, typical split-CV and conventional dc I_d-V_g characteristics are shown for an EOT~5 nm SiO_2 NMOSFET [2]. However, in case of high-κ materials, this technique is not reliable due to the presence of strong

charging effects during the quasi-dc measurements. To overcome this difficulty, an alternative measurement technique is proposed in literature [9], which measures the inversion charge and drive current in the μs time scale. A modification to the original technique proposed in [9] will also be discussed in the following section.

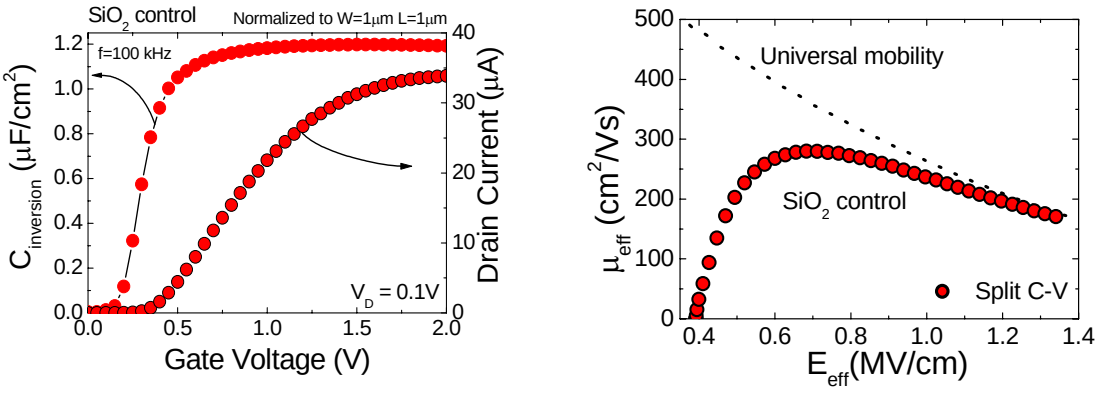


Figure 2.11. The conventional split-CV/I_d-V_g technique-The inversion charge is estimated by integrating the split-CV and the mobility henceforth calculated by combining it with the dc I_d-V_g [2]. (Note: V_d = 10 mV ensures that the transistor is biased in the linear region.) [2]

2.5 The Inversion Charge pump based Mobility Estimation Technique (ICP)

Before we delve into the details of the ICP technique of directly estimating the inversion charge using the geometrical component of the amplitude sweep charge pump of a large geometry device, we will discuss the motivation behind doing this. For a starter, we take a closer look at Figure 2.12 and Figure 2.13. This is a family of split-CV curves measured at different frequencies for a 5 nm SiO_2 control based NMOSFET with extremely high N_{it}~ 1 X 10^{12} cm^{-2} (measured independently by amplitude sweep charge pump technique). The mechanism of N_{it} generation would be discussed in detail in Chapter 4.

However, it is worth noting here that under this kind of high N_{it}, severe frequency dispersion in the split CV characteristics is observed. This translates into significant errors in the mobility versus N_{inv} plots as shown in Figure 2.13. In order to ascertain the right mobility, one needs to benchmark and compare these plots with mobility estimated by a different technique, which is believed to reasonably immune to distortions caused by N_{it} and trapped charge in the integrated split-CV [2].

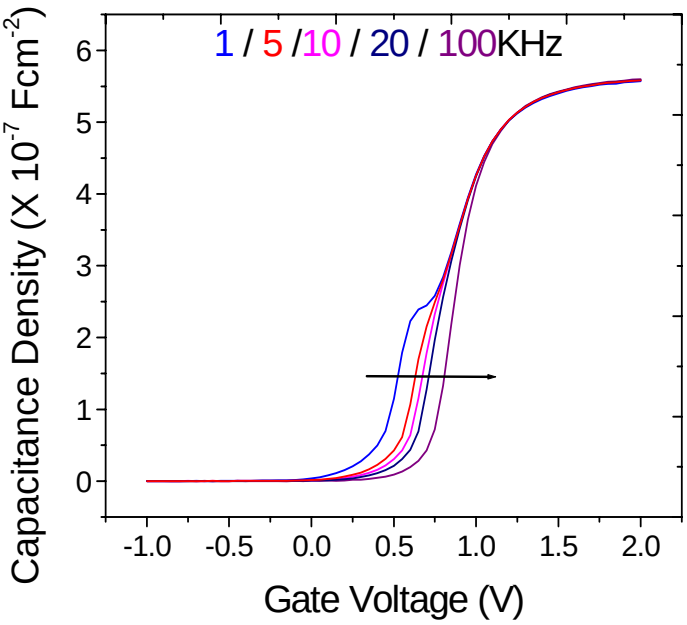


Figure 2.12. Frequency dispersion in Split-CV curves in a conventional SiO_2 NMOSFET with large N_{it}~1 X 10^{12} cm^{-2}

In Figure 2.13, we have also subtly emphasized on one aspect of the mobility estimation problem which we have so far ignored, namely the impact of series resistance/channel resistance and the accuracy with which the effective channel length is measured. This is also depicted in Figure 2.13. Apart from the frequency dispersion observed because of the errors in estimating the trapped charge, different channel length transistors exhibit

different mobilities. Using shift and ratio [10] technique, an accurate estimation of the series resistance was made and subsequently the mobility curves were corrected. This will be discussed in more detail in the context of low temperature measurements in Chapter 5, where it will be shown that this is a more pressing concern at low temperatures. Suffice it to say, that henceforth whenever mobility plots will be reported throughout this thesis, it will be corrected for series resistance, R_s (wherever it is expected to be a concern) and the printed gate length will be corrected subtracting twice the overlap length from the printed gate length, L_g to account for S/D side-diffusion.

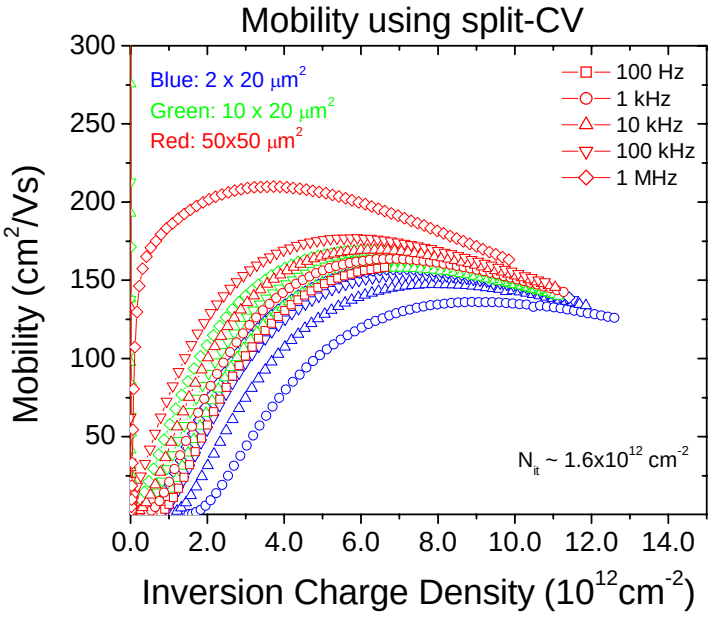


Figure 2.13. Frequency dispersion evident in mobility plots, the severe channel length dependence stems from the fact that series resistance, R_s and channel resistance depends strongly on channel length, L. Also the inaccuracy in mobility estimation stems from the fact that the metallurgical channel length is different from the printed gate length (in this case, R_s~ 65 Ω and $2\Delta L$ ~ 0.165 μm)

Now, we will develop the ICP technique as a means to benchmark the split-CVs for accuracy, so that we can use the split-CV/I_d-V_g technique frequently while we study the N_{it} response on transport (or more specifically calibrate accurately the N_{it} response parameter for the mobility curve) in Chapter 4. It's implications on high-κ transport will also be discussed.

It is worth mentioning that apart from the ICP technique several improvements in the mobility estimation technique has been suggested in literature to skirt around the problem of charge trapping as will be discussed in the following paragraph. Here, we have deliberately chosen a conventional SiO_2 based NMOSFET with high N_{it} as a test case since it is a system which is very well studied.

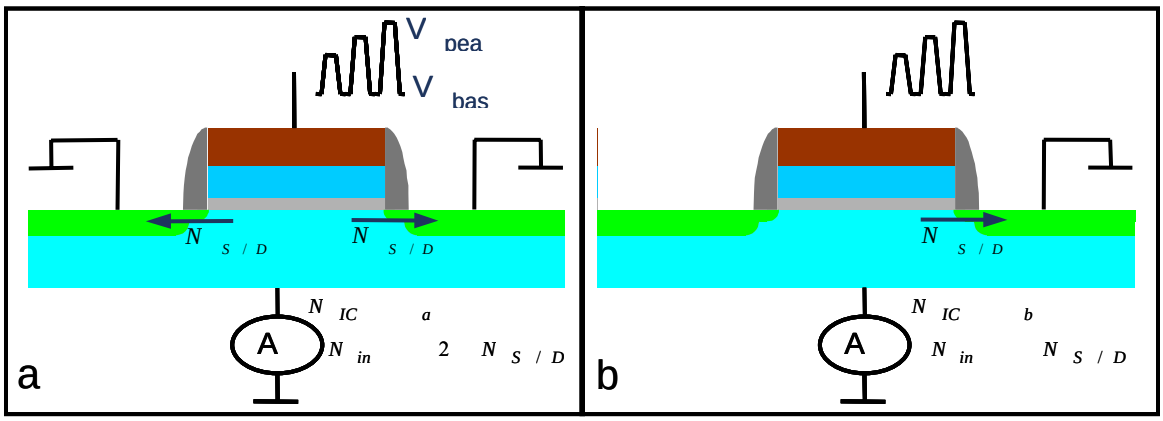
A modified split-CV technique taking into account the role of trapped charge was developed by W. Zhu et.al [11] wherein a measured split-CV is corrected using an ideal split-CV estimated for the same effective oxide thickness (EOT). The EOT is determined from the accumulation capacitance or C_{ox}. The problem with this technique is that the accuracy of the estimated mobility is critically dependent on the accuracy of the estimated EOT. Also the technique is untested on state of the art gate stacks where charge trapping may be low. An alternative way to measure the mobility is to use the Hall technique, a detailed study of which has been carried out by L.A Ragnarssen et. al. [12]. Hall technique was used to determine mobility in Hf-based dielectric structures. However, the accuracy of this scheme hinges on the value of Hall factor used. Ragnerssen et. al. used a value of unity. We believe any uncertainty in the value of Hall

factor can introduce errors into the mobility value so obtained. Over and above that, how the Hall mobility relates to the surface field effect mobility in bulk-Si channels of MOSFETs (where the mobility is completely limited by surface recombination) is unclear and a matter of conjecture.

Thus, we adopt the technique reported for the first time in [9] and tune it to suit our purpose- namely choosing the best split-CV (frequency) and the channel length for future mobility estimation. The motivation behind doing this is to take advantage of the fact that a conventional split-CV/I_d-V_g mobility estimation technique is very simple to carry out, and does not require any special characterization setup to perform (viz: a 4284 CV meter for estimating the split-CV and a 4155 for doing the I_d-V_g) . On top of that, the conventional split-CV/dc I_d-V_g technique has the advantage of being integrated in line with an automated test environment.

Here, we discuss in detail the ICP technique of mobility estimation (based on inversion charge pumping) developed by A. Kerber in his PhD thesis [2, 9]. Inversion charge in mobility estimation is determined from the geometrical component in amplitude sweep charge pump measurements, schematically described in Figure 2.14. It is interesting to note that inversion charge pump component (ICP) or the geometrical component so far was regarded as a parasitic component in amplitude sweep charge pump type measurements, and measurement conditions were optimized in such a way so as to reduce this component. In this case, however, in order to maximize the contribution of the geometrical effect, long channel devices in combination with fast rise and fall times were

used. Figure 2.14 describes the detailed measurement scheme as reported by Kerber et. al. [9]. It is instructive to note here that even though long channel devices were used, some fraction of the inversion charge still escapes back to the source/drain junctions and does not contribute to the substrate current. By making use of the device symmetry, A. Kerber [9] corrected for this. The charge pumping current was measured contacting both junctions in one case and only one junction in the other case. The complete inversion charge was then computed by using the formula shown in Figure 2.14.



$$N_{inv} = 2 \cdot N_{ICP\_b} - N_{ICP\_a}$$

Figure 2.14. Schematic showing the inversion charge pump scheme of directly measuring the inversion charge, the one and two contact scheme of measuring the complete inversion charge correctly is also shown [2]

However, despite the best of our efforts, we never got the one and two contact correction scheme, which accounts for loss of inversion charge to the source/drain during inversion charge pump, to work. We came to the definitive conclusion that the proposed correction scheme is not failsafe in cases where it is used in devices with large junctions. Also, the capacitive coupling between the source and drain can smear out the difference between

the one and two contact inversion charge pump curves. In order to get around this problem, we chose two symmetric devices namely a 50/50 (μm/μm) and a 100/100 (μm/μm) transistor. Amplitude sweep charge pump measurements with fast rise/fall time pulses (to ensure that the inversion charge does not get sufficient time to drift back to the source or drain junctions) were carried out on the two transistor geometries mentioned above. It is assumed that whatever charge escapes, despite using very fast rise and fall times, does so from the edges of the channel. This in turn implies that charge loss is directly related to the width of the transistor under consideration. Hence, for a 100 μm width transistor, the charge loss during an inversion charge pump is twice as large as a 50 μm width transistor. Thus, we may write the following two equations:

$$\mathbf{I_{CP}'(100) = 4I_{CP}(50) - 2\Delta I} \qquad \text{……(2.1)}$$

$$\mathbf{I_{CP}'(50) = I_{CP}(50) - \Delta I} \qquad \text{……(2.2)}$$

To the first order, I_{CP} (50) gives the correct estimate of the inversion current. Once the current magnitudes are normalized using the dimensions of the devices, and frequency, and electronic charge, a quantitative estimate of the correct inversion charge may be obtained from the following simple equation

$$\mathbf{N_{CP} = 2N_{CP}'(100) - N_{CP}'(50)} \qquad \text{…….(2.3)}$$

It is instructive to note that there is a surprising similarity between the original 1 and 2 contact correction equation and the proposed correction technique. This is partly because of the similarity in assumptions made in both cases – namely loss of inversion carriers from the edge of the channel. However, there is a caveat-if charge loss occurs from the body of the channel, rather than from the edges, this assumption breaks down. So the

experimental conditions need to be so optimized as to ensure that charge loss happens only from the edges of the channel and nowhere else.

It has been assumed in the foregoing calculation that the trapped charge or interface state densities (N_{it}) are negligible. How we deal with the trapped charge in the scheme described above will be taken up in the latter part of this section. Before we go any farther, we would like to check the accuracy of the proposed 50/100 scheme over the 1 and 2 contact scheme. We implicitly assume that for low N_{it}, the split-CV/I_d-V_g technique and the ICP technique should yield the same value of corrected inversion charge and subsequently mobility. This is what is depicted in Figures 2.15 and 2.16.

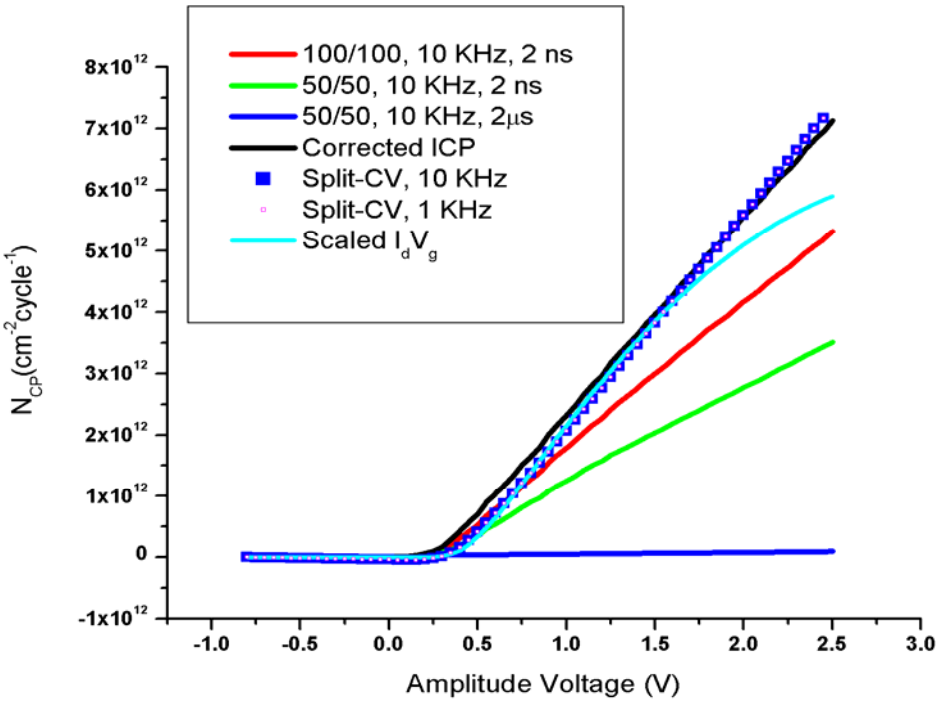


Figure 2.15. Corrected inversion charge obtained from ICP and split-CV/I_d-V_g technique demonstrate an exact match justifying the assumptions of the 50/100 correction scheme.

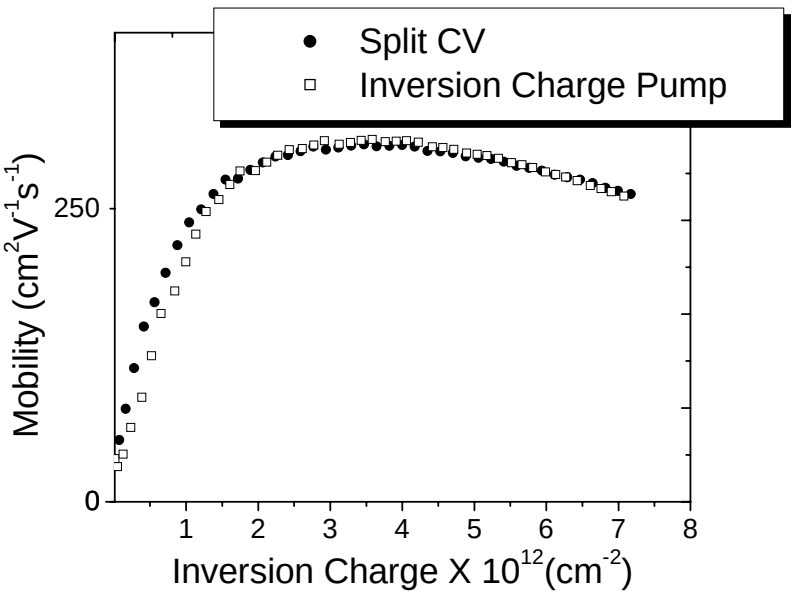


Figure 2.16. Comparison of mobility estimated from the Split-CV/dc I_d-V_g technique and the ICP technique using the 50/100 correction scheme rather than the 1 and 2 contact correction methodology. The match between the two is excellent.

With the benchmarking of the proposed technique complete, we now focus our attention on how to use this technique in presence of trapped charge. It should be noted that with ~ns order rise/fall time, the inversion charge does not get enough time to escape to the source drain junctions, and the channel gets pinched off, with the channel charge subsequently recombining with the incoming holes in the substrate thereby giving the charge pump response at the substrate, referred to as inversion charge pumping (ICP) or the geometrical component.

However, when a longer rise/fall time is used, the channel charge drifts back to the source/drain and the charge pump response contains only the response of N_{it} (and also bulk traps in the case of high-κ dielectric based NMOSFETs). So, longer rise/fall pulses may be used to capture the trapped charge and then, the N_{CP}' ('s) can be individually corrected with the trapped charge information before applying the correction scheme depicted by equation (2.3). However, no matter how long the pulses used are, there is still a possibility that some amount of inversion charge would be still pumped, particularly in case of long channel transistors as is evident in Figure 2.17.

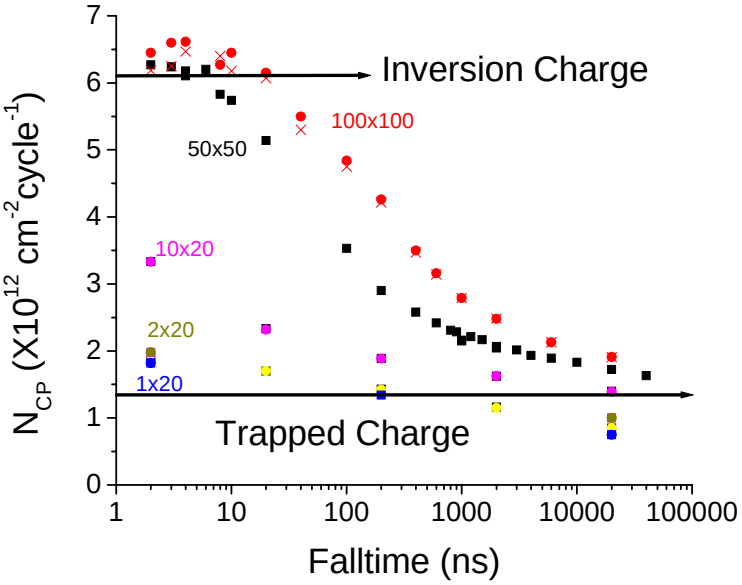


Figure 2.17. Channel Length dependence of Trapped Charge (Note that at higher fall/rise times (~μs), there is still substantial dispersion in charge pump current between the different channel length transistors)

So knowing exactly the right trapped charge is a formidable challenge, and the problem of determining the correct trapped charge is compounded by the fact that it is extremely difficult to get detectible charge pump response in very small channel length transistors. It is also worth mentioning in this context what has been pointed out earlier, that on very small channel length transistors, there may be substantial loss of even trapped charge to the source and drain junctions, and the signature of the entire trapped charge may not be detected in the charge pump response. It is therefore necessary to optimize the measurement conditions and the choice of channel length in such a way that the entire N_{it} response is obtained. One such representative case is depicted in Figure 2.18.

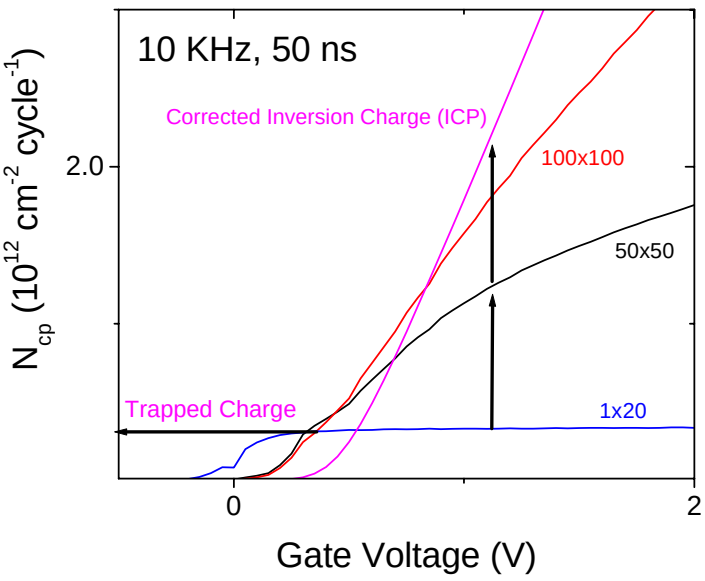


Figure 2.18. ICP measurement carried out on a transistor with 3 x 10^{11} cm^{-2} N_{it}. The signature of the trapped charge is measured on an extremely short, 1 µm transistor, where there is no response of inversion charge in the charge pump response.

Now, the corrected inversion charge determined by this method is compared with two different frequencies (10 and 100 KHz) split-CV curves. It is noticed that the 100 KHz split-CV carried out on a large geometry transistor (in this case W/L: 20/10 µm/µm) gives the inversion charge value closest to that measured by the ICP technique. This is depicted in Figure 2.19. Figure 2.20 shows the mobility curves of the device mentioned earlier determined by the ICP and 100 KHz split-CV technique. On a log scale, it is seen there is excellent match between the ICP and the split-CV mobility curves at low fields where Coulomb scattering is expected to be dominant.

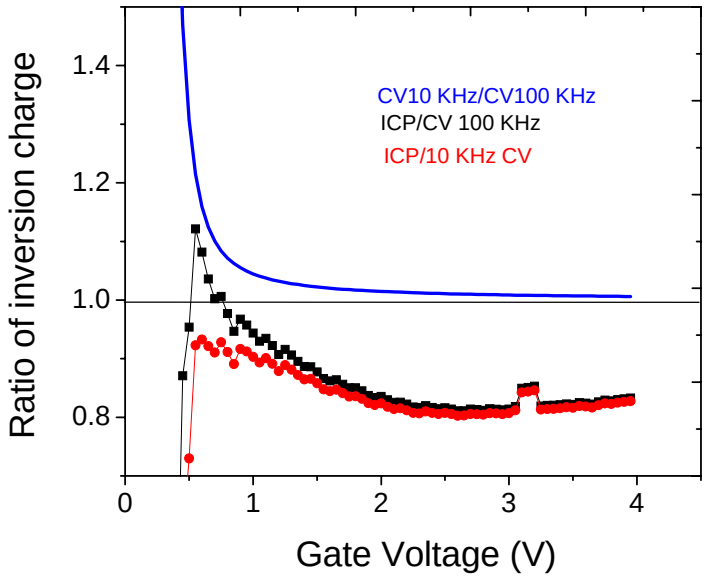


Figure 2.19. Comparison of integrated inversion charge obtained by integrating the split-CVs with two different frequencies (10 and 100 KHz) and that obtained by the ICP technique as described earlier in this section.

Based on the foregoing analysis, it is instructive to note that throughout the rest of the thesis, we will be using a 100 KHz split-CV/dc I_d-V_g technique for estimating the mobilities even in presence of N_{it}. As depicted in Figure 2.19, the subsequent errors in analysis are negligible, particularly at low fields. The difference between the ICP determined mobility and the split-CV technique determined mobility at intermediate fields is in all probability due to the non-quasistaitc nature of the ICP measurement.

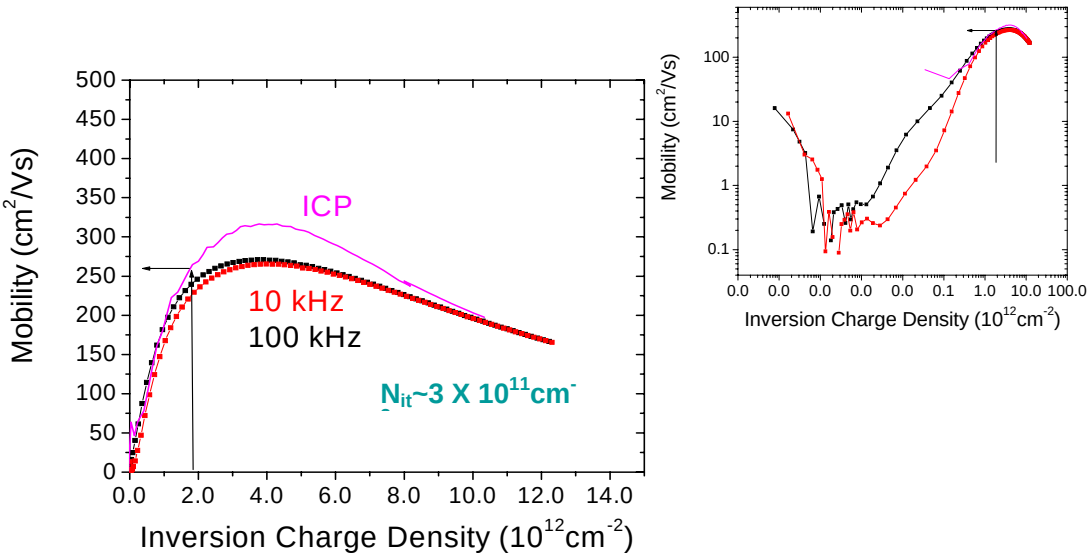


Figure 2.20. Mobility plots obtained by the ICP technique and the 10 and 100 KHz split-CV technique. (Inset: Mobility plotted in the log-scale to emphasize the difference at low fields)

The 100 KHz split-CV/I_d-V_g technique will be used in Chapter 4 to calibrate the interface state response (N_{it}) on mobility, and subsequently apply it to Poly-Si/high-κ and metal gate/high-κ based advanced gate stacks. The errors introduced by not using the ICP technique in our opinion will be minimal as will be demonstrated by the accuracy of the

proposed calibration scheme, and as mentioned earlier, the simplicity of the split-CV/I_d-V_g technique makes it extremely attractive in determining the mobility in MOSFETs.

2.5 Charge Pumping in Metal Gate/High-κ Stacks

Before we end this chapter, we would like to take a closer look at the charge pumping response on metal gate based high-κ gate stacks. In this section we give a very brief overview of charge pumping measurements carried out on metal gate high-κ gate stacks. It is necessary to take into consideration the basic physics of a metal gate high-κ system. It is well known that a poly-Si/high-κ or poly-Si/SiO_2 based transistor in general shows poly-depletion [10]. So if there is a sheet of charge in the high-κ gate stack, the ground plane in the gate electrode where this charge is screened, is pushed farther and farther away, depending on the scale of poly-depletion. Whereas, for a metal gate electrode, because of high carrier density in the gate electrode, a poly-depletion like phenomenon is absent. This in turn implies that the ground plane where the charge in the high-κ gate stack is screened is much nearer than in the case of a poly-Si gate electrode. Thus quasistaic measurements which track inversion phenomena like threshold voltage shift may not be able to detect the full extent of charge trapping. In this context, a direct measurement of trapped charge provided by an amplitude sweep charge pump is extremely useful.

Figure 2.20 shows charge pump characteristics of a representative metal gate high-κ gate stack with reasonably low EOT~1.8 nm (T_{inv} ~ 2.2 nm). The metal gate is TaSiN and the

high-κ used is HfO_2. The stack has been fabricated using non-self aligned gate last process details which may be found in Chapter 2 of [2]. Both base sweep and amplitude sweep are simultaneously plotted, it is instructive to note that base sweep gives information about the interface states only (N_{it}) whereas amplitude sweep gives information about both bulk traps and N_{it}.

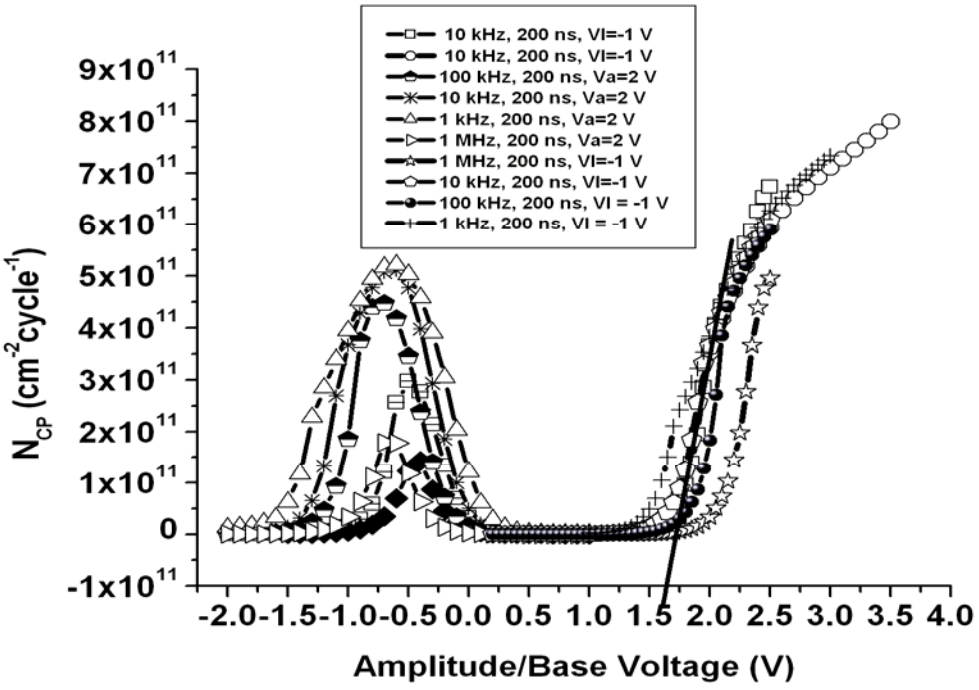


Figure 2.21. Amplitude and Base Sweep simultaneously performed on a TaSiN/HfO_2 gate Stack

This charge pump study on TaSiN/HfO_2 gate stack based NMOSFET is designed to act as a template for studying high mobility aggressively scaled metal gate high-κ NMOSFETs in Chapter 4 and 5 later in this thesis. We will be using the conductance technique in Chapter 5 to estimate N_{it} in presence of high leakage in aggressively scaled metal gate/high-κ NMOSFETs.

2.6 Conclusions

In the foregoing discussion, we have discussed in detail the amplitude sweep charge pump technique as a viable characterization methodology for determining interface state densities (N_{it}) and also bulk traps in advanced high-κ gate stacks. We also discuss in detail an inversion charge pump based mobility estimation technique-also develop a correction scheme to the already existing methodology of inversion charge pump based mobility determination. The proposed correction methodology is found to work extremely well when tested against the more conventional split-CV/I_d-V_g technique performed on a thermally grown SiO_2 based NMOSFET with low N_{it}. Then using the developed methodology, we benchmarked the split-CVs to choose the most correct split-CV to be used later in estimating the mobilities in SiO_2 and high-κ based NMOSFETs in presence of interface states.

In the next chapter, we identify two components of trapped charge in a representative high-κ based gate stack namely: stable charge and unstable or transient charge. By a novel experimental methodology, a semi-quantitative estimate of the location and magnitude of trapped charge is made. A separate experimental technique namely the in-situ sense at stress technique is introduced in order to quantify the full extent of transient charging. The impact of transient charge and remote stable charge on transport will also be studied.

References

[1]. D. K Schroder, “Semiconductor Material and Device Characterization”, Second Edition, 1998, John Wiley and Sons, Inc.

[2]. A. Kerber, “Methodology for Electrical Characterization of MOS devices with Alternative Gate Dielectrics”, PhD Thesis, T.U Darmstadt, and IMEC Belgium, 2003.

[3]. J. S Brugler, and P.G.A Jespers, “Charge Pumping in MOS Devices ” , IEEE Transactions on Electron Devices, 1969, Vol. 16, p. 297.

[4].G. Groeseneken, H. E Maes, N. Beltran, and R. F De Keersmacker, “A Reliable Approach to Charge Pumping Measurements in MOS Transistors ” , IEEE Transactions on Electron Devices, 1984, Vol. 31, No. 1, pp. 42 – 53.

[5]. H. Resisinger et. al., “A Comparative Study of Dielectric Relaxation Losses in Alternative Dielectrics”, IEDM Technology Digest, 2001, pp. 12.2.1 – 12.2.4.

[6]. A. Kerber, Private communication.

[7]. D. Ielmini, A. S Spinelli, A. L Lacaita, and A. Modelli, “ A Statistical Model For SILC in Flash Memories”, IEEE Transactions on Electron Devices, 2002, Vol. 49, No. 11, pp. 1955 – 1961.

[8]. A. Kerber et. al., "Strong Correlation between Dielectric Reliability and Charge Trapping in SiO_2/Al_2O_3 gate stacks with TiN electrodes", VLSI Technology Symposium, Honolulu, 2002.

[9]. A. Kerber et. al., "Direct Measurement of the Inversion Charge in MOSFETs: Application: Application to Mobility Extraction in Alternative Gate Dielectrics", VLSI Technology Symposium, 2003, pp. 159 – 160.

[10]. Yuan Taur, and Tak H. Ning, "Fundamentals of Modern VlSI Devices", 1998, Cambridge University Press.

[11]. W. Zhu, J. P Han, and T. P Maa, "Mobility Measurements and Degradation Mechanism of MOSFETs made with Ultrathin High-κ Dielectrics", IEEE Transactions on Electron Devices, 2004, Vol. 51, No. 1, pp. 98 – 105.

[12]. L.-Å Ragnarsson, N. A Bojarczuk, J. Karasinski, and S. Guha, "Hall Mobility in Hafnium Oxide Based MOSFETs: Charge Effects", IEEE Electron Device Letters, 2003, Vol. 24, No. 11, pp. 689 – 691.

Chapter 3

Location and Magnitude of Trapped Charge – The Impact of Remote Charge on Mobility

3.1 Introduction

The magnitude and the location of trapped charge in Al_2O_3 capped Hf-silicate (HfSiO) gate stacks with Poly-Si gate electrodes is studied with 'stress-and-sense' type measurements using standard CV measurements to monitor the flatband voltage shifts as a function of injected charge. A simple quantitative model is used to estimate the charge location within the multilayer gate stack. To the first order, the magnitude of trapped charge is found to be on the order of 5 X 10^{12} cm^{-2}, independent of the cap layer thickness, and the charge is found to be located within the Hf-silicate layer with the charge centroid situated 1 nm away from the Al_2O_3/HfSiO interface. The motivation behind using Al_2O_3 cap layer and its utility in advanced gate stack is properly justified. It is also emphasized that though this experiment has been carried out with Hf-silicate gate stacks, it is applicable to more commonly used HfO_2 gate stacks as well. It is clearly pointed out that these stacks have two components of charge, namely, stable charge and unstable charge. The former component is detected and quantified by the method described above and the latter component is detected by the in situ sense at stress technique. The impact of transient charging on mobility is studied. A novel experimental methodology is proposed by which stable charge is introduced inside the gate stack (away from the interface or "charge at a distance") by electrical stressing. Direct experimental evidence is given on the impact of remote charge on mobility. Theoretical

conjectures on impact of screening of remote charge by the carriers in the gate is also discussed. It is conclusively shown that charge away from the interface has very limited influence on transport in the channel.

3.2 Motivation: Why Al_2O_3 (Alumina) Cap?

The utility of Alumina cap on high-κ gate stacks is summarized below:

(1) The well known pFET V_t shift observed in poly-Si and high-κ based gate stacks is often attributed to Fermi level pinning on the gate side of the poly-Si gate electrode as illustrated in Figure 3.1 [1].

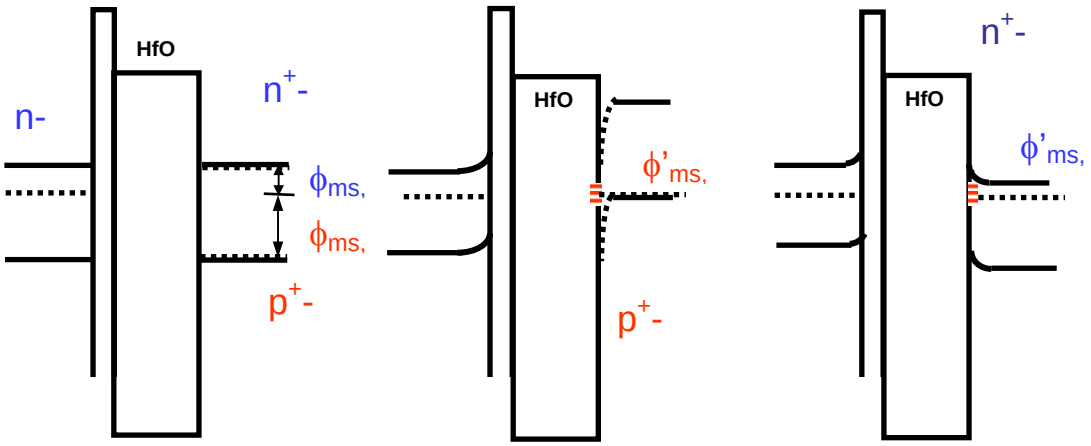


Figure 3.1. Band Structure of poly-Si, HfO_2, (100)-Si illustrating Fermi level pinning leading to the pFET V_t shift [1].

This phenomenon of Femi level pinning leads to increase in pFET threshold voltage (V_t) thereby making the pFET V_t asymmetric with respect to the nFET V_t.

A number of solutions have been proposed in literature to counter this problem, T. Iwamoto et. al. tried to eliminate the pFET V_t problem by channel engineering. Figure 3.2 conclusively shows that this method of attempted V_t –adjustment failed to solve the problem [2].

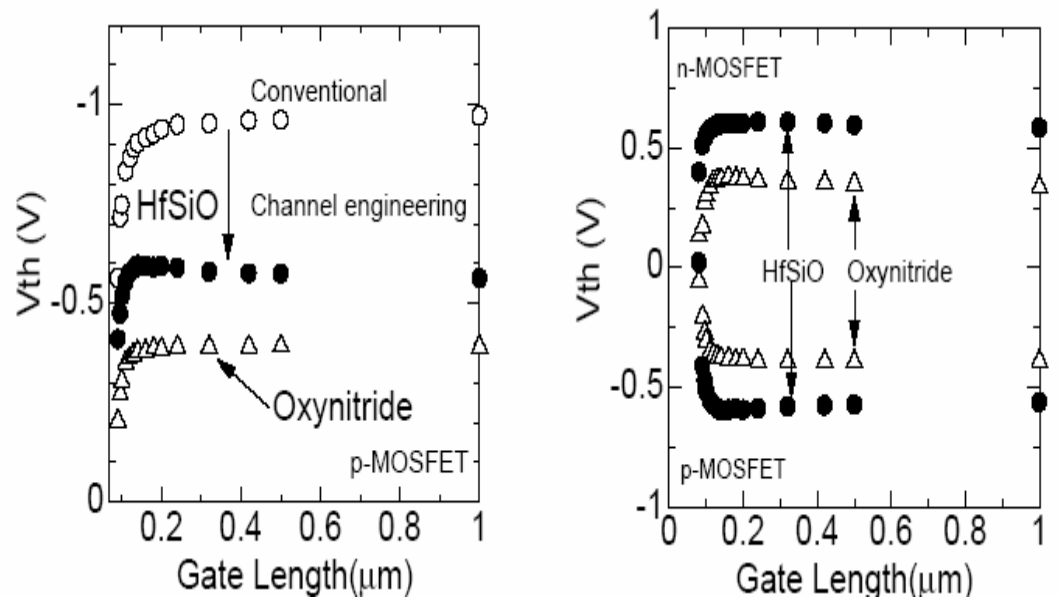


Figure 3.2. Taking oxynitride as the baseline, it is demonstrated that careful halo engineering cannot eliminate the pFET V_t shift [2].

T. Iwamoto et. al. [2] also showed that with a dual layer symmetric structure, with SiO_2 layers on either side of the high-κ structure, the pFET V_t shift may be eliminated. However, with this structure, a severe EOT penalty has to be paid (Figure 3.3) rendering this technique unattractive.

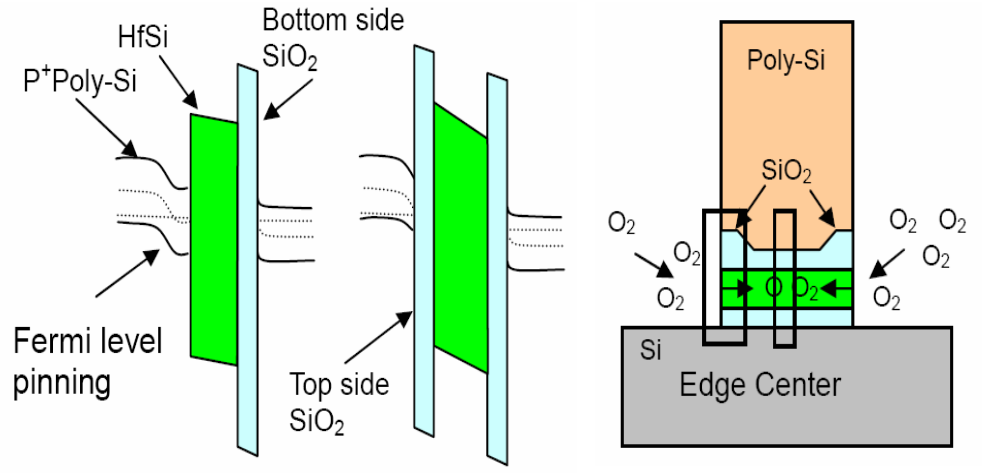


Figure 3.3. Dual layer symmetric structure eliminates pFET V_t shift, but a severe EOT penalty is paid affecting scalability. The schematic on the right illustrates the mechanism of oxygen diffusion resulting in increase of EOT [2].

There is consensus in the literature [1,2,3] that the pFET V_t shift is due to Fermi level pinning on the gate side, possibly between the p+ poly-Si gate and the high-κ gate stack. In this context, putting a capping layer between the gate electrode and the underlying dielectric appears attractive, so long it does not adversely affect the EOT. SiN as a capping layer was tried in [3] with limited success. Whereas, Al_2O_3 cap on Hf-silicate appears to be a reasonably credible solution [3]. It is to be noted that this capping layer is deposited by atomic layer deposition technique (see section on processing details) and growth of the capping layer from cycle to cycle appears to be strongly nonlinear (Figure 3.4). This has an inherent advantage, upto ~5 cycles, there is limited EOT change, but V_t moves back to a value, reasonably symmetric with nFET V_t (Figure 3.5). Thus , a symmetric V_t configuration may be achieved without affecting scalability.

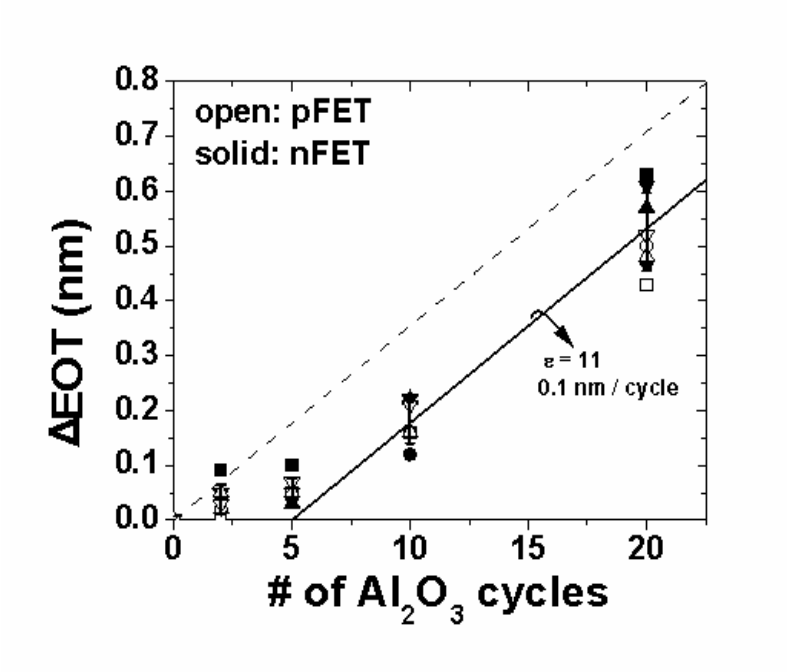


Figure 3.4. EOT growth during ALD cycle highly nonlinear thereby ensuring that the EOT change during the first few cycles to be nominal (Dotted line represents linear growth assuming a growth rate of 0.1 nm/cycle)

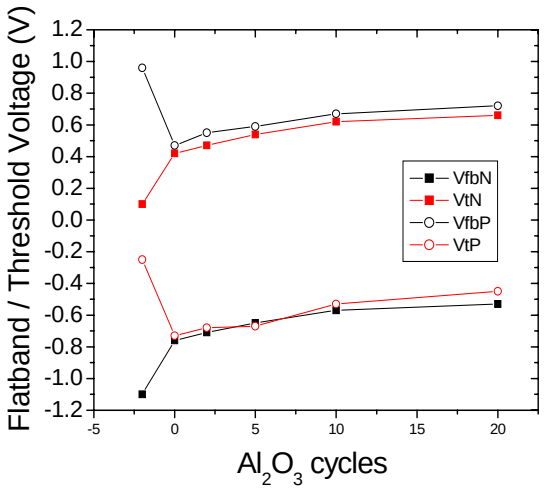


Figure 3.5. Reasonably symmetric V_t configuration achieved after ~5 cycles of ALD Al_2O_3 with minimal change of EOT [3]

(2) Recent reports [4] in literature with $HfAlO_2$ based gate stacks both with poly-Si and FUSI gates also show elimination of pFET V_t shift.

(3) Al_2O_3 cap on HfO_2 may be a credible way to increase (by making use of Hf-Al intermixing) the dielectric constant of the gate stack thereby extending scalability into the sub-1 nm regime [5].

(4) Al_2O_3 cap is known to simplify process integration and is also known to yield pFET like workfunction on SiO_2 [6] (vide Figure 3.6)

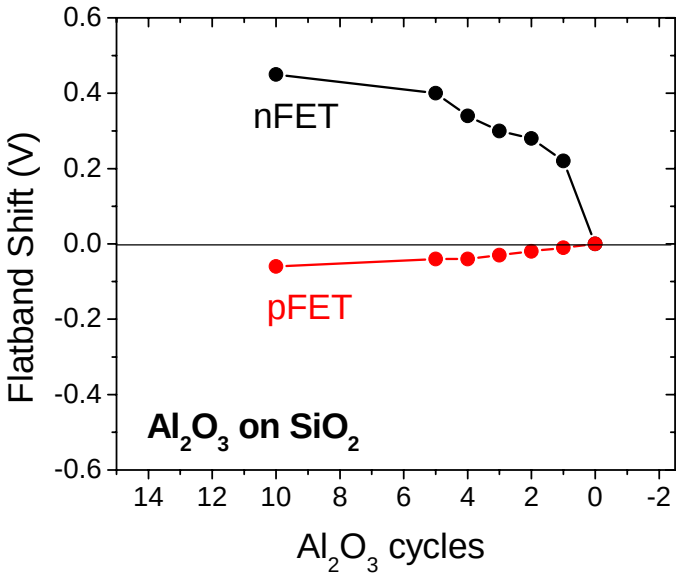


Figure 3.6. Flatband shift with Al_2O_3 cap on SiO_2 minimal, preserving pFET V_t to its original value, but increasing the dielectric constant [6].

(5) A recent work by a group at SEMATECH [7] has shown that TiN metal gates with Al_2O_3 capped HfO_2 can yield pFET like workfunction. Thus, this capping layer approach may be used for workfunction tuning on high-κ.

(6) Al_2O_3 cap on Hf-silicate shows excellent leakage reduction (Figure 3.7) and if it is properly integrated with poly-Si stack can be used as a credible ultra low power solution.

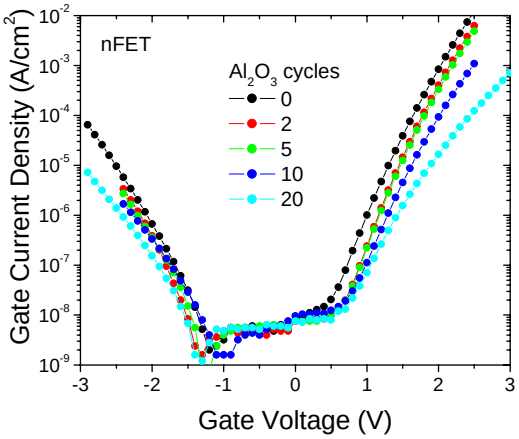


Figure 3.7. Excellent leakage reduction with Al_2O_3 capped HfSiO nFETs. Asymmetry in leakage characteristics relate to the asymmetry in the band structure in dual layer stacks as pointed out in [8].

(7) The samples as will be described in the section on processing details, which includes varying cycles of ALD Alumina cap on HfSiO and provides an opportunity to study transport in the channel in possible presence of remote charge in the Hf-silicate layer itself because the ground plane where the remote charge is screened namely the poly-Si gate electrode is pushed farther and farther away from the bulk-Si channel.

However, it is well known that Al_2O_3 is known to increase charge trapping in Si MOSFETs [9]. So using it as a capping layer to take advantage of the factors stated above

may come with some negative effects. In this work, we focus on investigating two such aspects:

(1) Identifying the sources of charge trapping with Al_2O_3 cap, namely locating stable charge, and also investigating the role of transient charging [10] as a limiting factor in these stacks.

(2) Incorporating by controlled electrical stressing remote charge or "charge at a distance" far away from the channel and thereby studying its impact on mobility in the channel.

3.3 Processing Details and Description of samples

Silicon oxynitride layers with a nominal thickness of 1 nm were used as a starting layer for CVD Hf-silicate (4 nm) deposition. For Al_2O_3 cap layer deposition, samples were transferred to an ALD reactor and 2, 5, 10 and 20 cycles of Al_2O_3 were deposited and the layers were capped with LPCVD poly-Si. n and p-FET devices were fabricated in a standard CMOS process using 1000 ^{0}C for gate activation. A final forming gas anneal was performed at 500 ^{0}C for device passivation. All measured devices have T_{inv}'s in the range of 2.5-3.0 nm which is of interest for low power applications [11] in present day VLSI technology. This is evident from the CV-plot in Figure 3.8.

The impact of Al_2O_3 cap layers on the effective oxide thicknesses (EOT's) was already depicted in Figure 3.4. As can be seen, the cap thickness does not increase linearly with ALD cycle number. Only after about 10 cycles does EOT increase at a rate consistent with the layer thickness increase of 0.1 nm per ALD cycle and with a dielectric constant

of ~ 11 as expected for Al_2O_3. As pointed out earlier, the data suggests that the growth is initially inhibited and that the cap is physically closed only for caps having more than ~10 cycles of Al_2O_3. This growth behavior is also reflected in the threshold voltage instability, as will be shown in a later section. It is instructive to compare and contrast this growth behavior with growth of Al_2O_3 cap layer on SiO_2 as described in [6], where growth is completely linear. If used for flatband voltage adjustment, the scaling penalties are quite small for less than about 10 cycles of Al_2O_3.

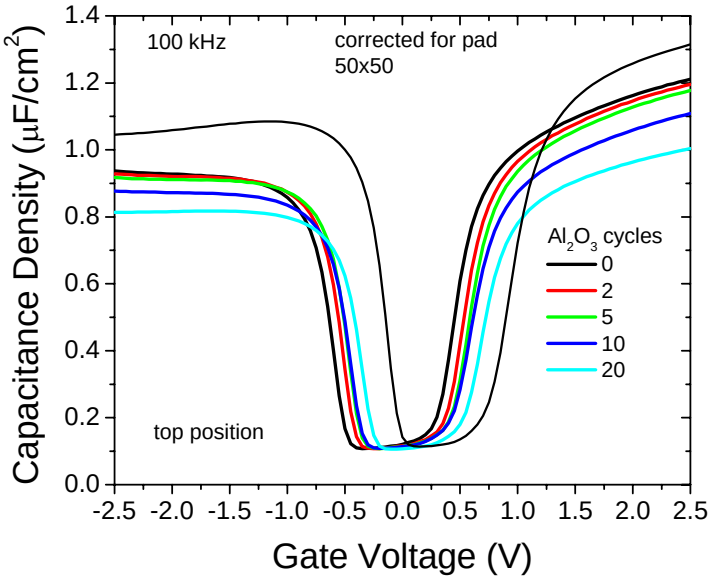


Figure 3.8. CV-plots showing T_{inv} 's in the range of 2.5 – 3.0 nm [3].

A representative TEM for the 5 cycle Al_2O_3 based HfSiO stack is shown in Figure 3.9. The high-κ (Hf-silicate) is found to be completely poly-crystalline. Alumina cap is not discernible, this is consistent with growth inhibition observed in the early ALD cycles as illustrated in Figure 3.4. The oxynitride interfacial layer appears to be significantly thicker than the targeted 1 nm thickness. The actual thickness is on the order of 2 nm, this appears to be either due to regrowth during the high temperature annealing step or more

likely due to TEM induced thickening. Depending on the exposure time of the electron beam on the gate stack, some thickening may be possible. It is difficult though to make a conclusive statement on the occurrence of TEM induced thickening.

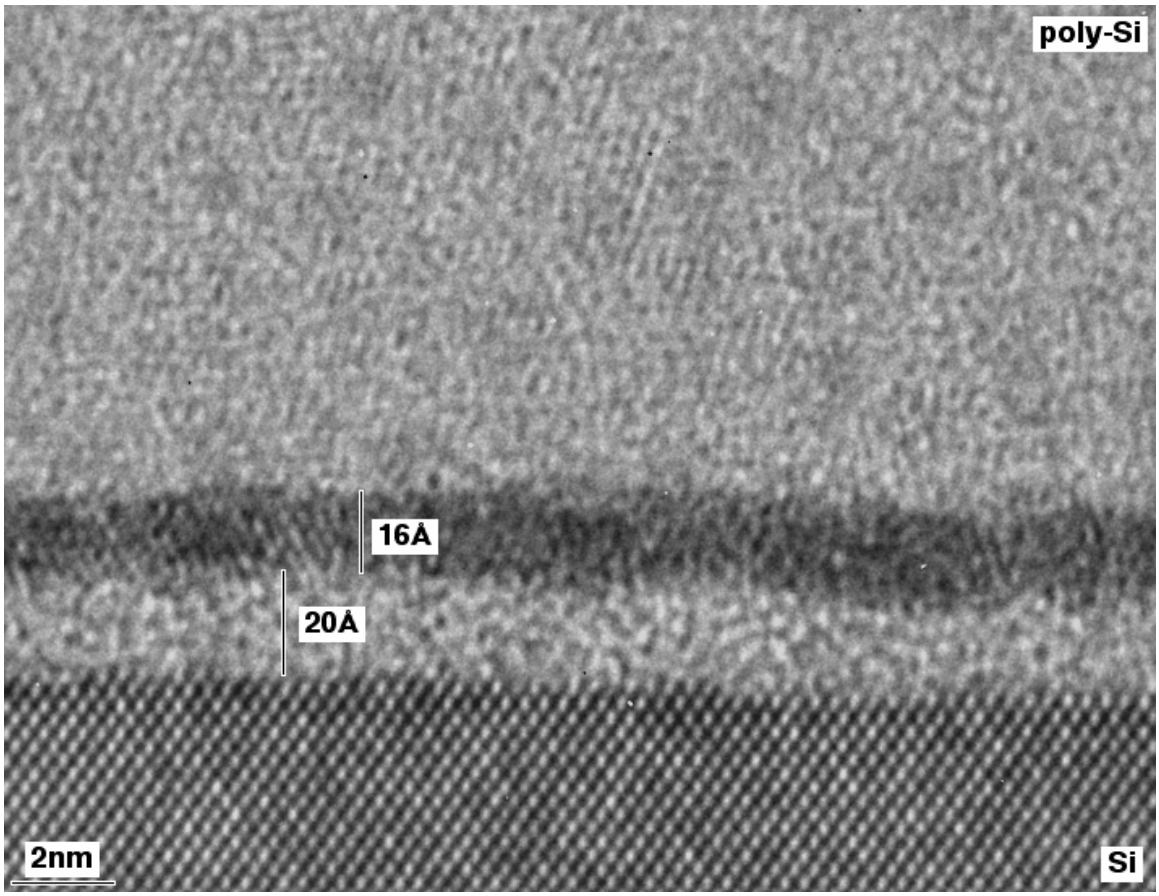


Figure 3.9 A representative TEM of the 5 cycle Al_2O_3 capped Hf-silicate

In order to investigate the charge trapping behavior of these stacks, we start by using conventional charge trapping characterization technique described in [12,13], also known in literature as the stress and sense characterization technique [14].

3.4 The "Stress and Sense" Characterization Technique

As pointed out earlier, this technique has been used extensively in literature to study the instabilities in conventional SiO_2 based gate dielectrics, and is known as the "stress and sense" method. In this case, an initial sense measurement is carried out prior to stressing the devices. The stress is then interrupted periodically and a sense measurement is performed. A schematic drawing of the bias sequence in "stress and sense" procedure is shown in Figure 3.10 below.

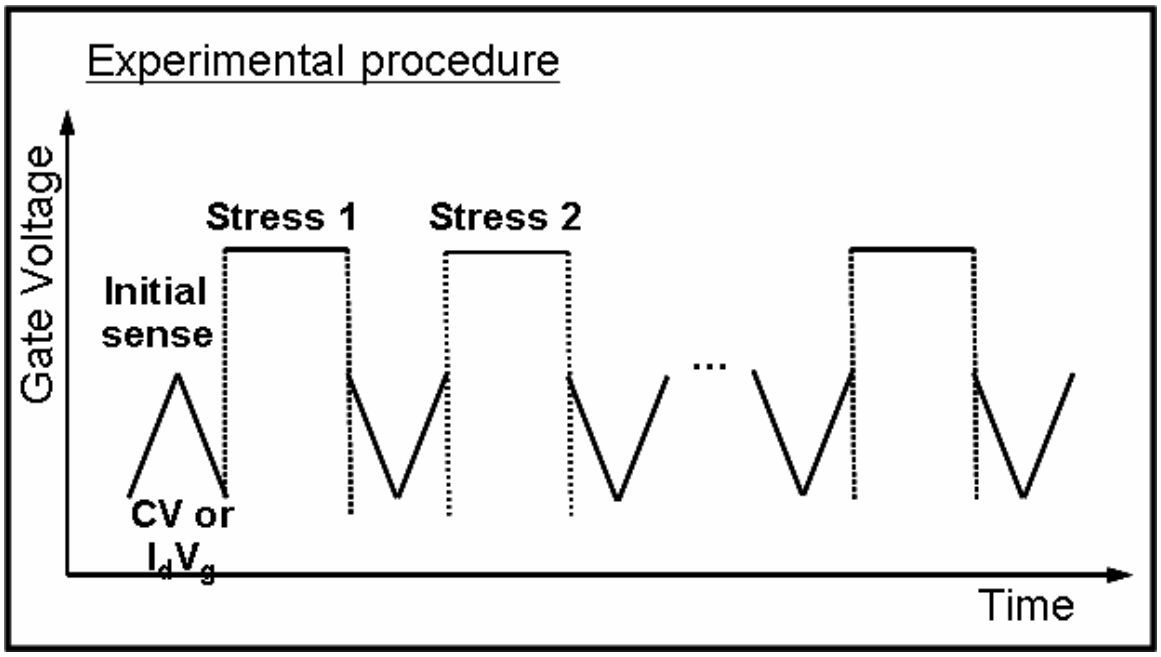


Figure 3.10. Schematic describing the stress and sense technique, CV or I_d-V_g sense sweeps are used in this thesis to monitor the shift in flatband/threshold voltage.

The instability of the device is extracted by comparing the sense measurement after stress with the initial device characteristics. Predictions to operation conditions are usually

made when combining the time dependence with the voltage dependence of the instability.

One of the known issues of the stress and sense procedure is the inherent time delay between stressing and sensing. In case, some recovery of the instability occurs at the time periods on the order of ~ 10 to 100 ms, this procedure will not capture its full extent. This will be aptly illustrated in section 3.6. Despite this limitation, this technique is chosen as a unique tool for detecting the full impact of stable trapped charge in high-κ based gate stacks.

3.5 Location and Magnitude of Trapped Charge

The threshold voltage instability for Hf- based dielectrics has been shown to be limited by electron trapping under positive gate bias (substrate injection) [10]. In this study, the charge trapping under positive gate bias is measured with a conventional 'stress-and-sense' procedure described earlier. The stress induced flatband voltage shift in capacitors on n-type substrates (pMOS) is monitored after each stress pulse using a conventional CV sweep measurement. Essentially, identical results were obtained if the V_t instability was measured on nMOSFETs, as expected. The evolution of the sense CV as a function of injected charge are shown in Figure 3.11 for 0 and 20 cycles of Al_2O_3 on the Hf-silicate. The reduction of the maximum capacitance with 20 cycles of ALD Al_2O_3 cap is indicative of the increase in EOT with Al_2O_3 cap thickness as summarized in Figure 3.4, and shows the scaling limitations of a cap layer approach for V_t adjustments. As can be seen from the data in Figure 3.11, no CV distortions are observed for substrate injection (CVs are shifted parallel) and the charge induced flatband voltage shifts, δV_{fb}, as a function of injected charge are readily extracted. No interface states are generated in

these gate stacks for the gate biases and injected charge magnitudes studied. The injected charge is obtained by integrating the gate leakage current,

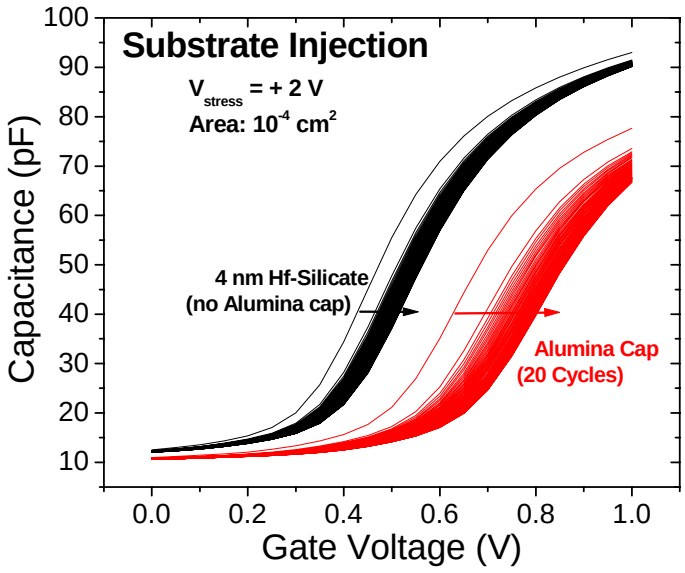


Figure 3.11. Comparison of the sense CV's measured in between successive stress pulses for a gate stack with no cap layer (solid lines) and with 20 cycles of ALD deposited Al_2O_3 on the 4 nm Hf-silicate layer. Positive gate bias is used for charge injection from the substrate.

The flatband voltage shifts for substrate injection (positive gate bias) as a function of injected charge are summarized in Figure 3.12. As can be seen, the flatband instability is substantially enhanced for the layers with thicker Al_2O_3 caps. To quantify the amount of trapped charge and to obtain information on the charge location, it is assumed that the Al_2O_3 cap does not alter the trapping behavior of the stack. The Al_2O_3 cap is assumed to only change the physical location of the charge centroid with respect to poly-Si gate electrode, as illustrated in Figure 3.13.

Thus, the measured flatband voltage shift changes with cap layer thickness as,

$$\delta V_{fb} = Q_{tr} / C_X \quad (3.1)$$

where Q_{tr} is the magnitude of the trapped charge per unit area and C_X is the capacitance density of a hypothetical capacitor of physical thickness X, covering the volume from the poly-Si gate to the charge centroid [15].

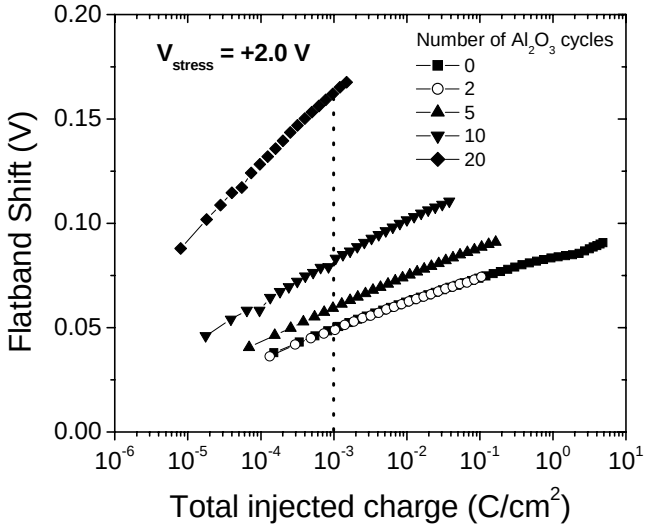


Figure 3.12. Flatband voltage shifts versus total injected charge for Hf-silicate gate stacks with various ALD Al_2O_3 cap layers (0, 2, 5 10 and 20 cycles) as extracted from data like that shown in Figure 3.11.

These simple assumptions allow a good quantitative description of the experimental data, as illustrated in Figure 3.14, in which the measured flatband voltage shifts for the various layers at a constant injected charge magnitude of 1 mC/cm^2 (from Figure 3.12) are compared with a model curve using the equation in Figure 3.13. The calculated curve corresponds to a charge magnitude of 5 X 10^{12} cm^{-2} with a charge centroid located at a

distance of 1 nm from the Hf-silicate/Al_2O_3 interface. This analysis is independent of the amount of injected charge. To illustrate this point, we calculated the differential trapping probability, $\delta N_{tr}/\delta Q_{inj}$, for the various layers studied using the charge centroid derived from the equation in Figure 3.13. The calculated data in Figure 3.15 support the use of the simple trapping model. A cap layer independent trapping behavior is obtained at all levels of injected charge, showing that to first order, the differential trapping probability only depends on the Hf-silicate layer and not on the cap thickness. Larger voltage shifts are only observed because the location of the charge centroid changes with respect to the poly-Si gate, thereby implying that the varying levels of ALD Al_2O_3 cap act as amplifiers. These findings justify our assumption that the cap layer does not change the trapping behavior substantially. Trapping remains controlled by the Hf-silicate trapping properties. The trapping probability of the Hf-silicate layer is found to be high for small amounts of injected charge and to decrease with $\sim Q_{inj}^{-1}$, a behavior frequently reported for Hf-based gate stacks with poly-Si gates [14]. With such high trapping probabilities, changes in the trapping behavior due to intermixing may not be detected easily. Also, extra trapping in the thin cap is close to the gate and has little impact on the net voltage shifts.

The magnitude and location of the trapped charge in Al_2O_3 capped Hf-silicate high-k gate stacks has been measured. It is shown that the threshold voltage instability can be substantially enhanced by such cap layers. However, no evidence is found that the presence of the cap layer dramatically alters the trapping properties of the gate stack. The reduced threshold and flatband stability for thicker cap layers is predominantly caused by

the change of the location of the centroid of the trapped charge with respect to the gate. The trapped charge is found to be located in the Hf-silicate layer within 1 nm of the Hf-silicate/Al_2O_3 interface. It is likely that the trapping behavior of the gate stack is changed somewhat due to intermixing, and charge trapping in the cap is also expected to be possible. However, based on the analyses presented here, the contributions of these effects to the threshold voltage instability under positive gate bias (substrate injection) appear to be small. If used for flatband voltage adjustments in pFETs, Al_2O_3 cap layers should be kept thin to minimize the distance of the trapped charge to the gate.

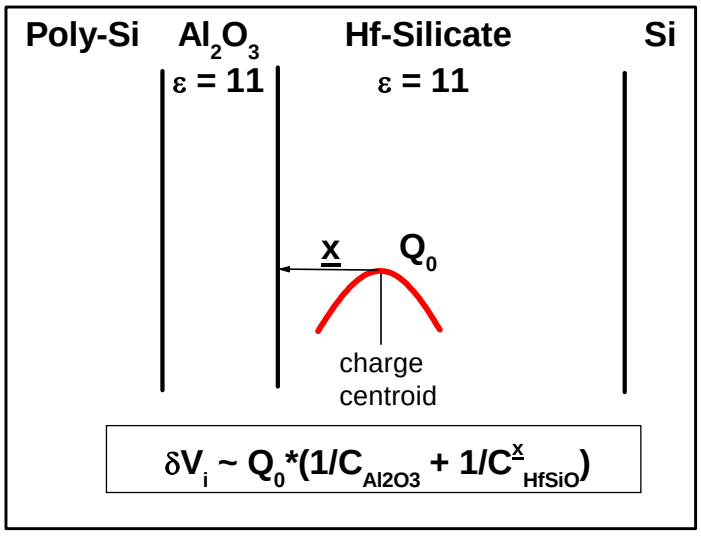


Figure 3.13. Schematic representation of charge trapping in Hf-silicate/Al_2O_3 dual layer gate stack used to calculate the location and magnitude of trapped charge. The charge centroid is assumed to be located within the Hf-silicate layer @ a distance x from the Hf-silicate/Al_2O_3 interface.

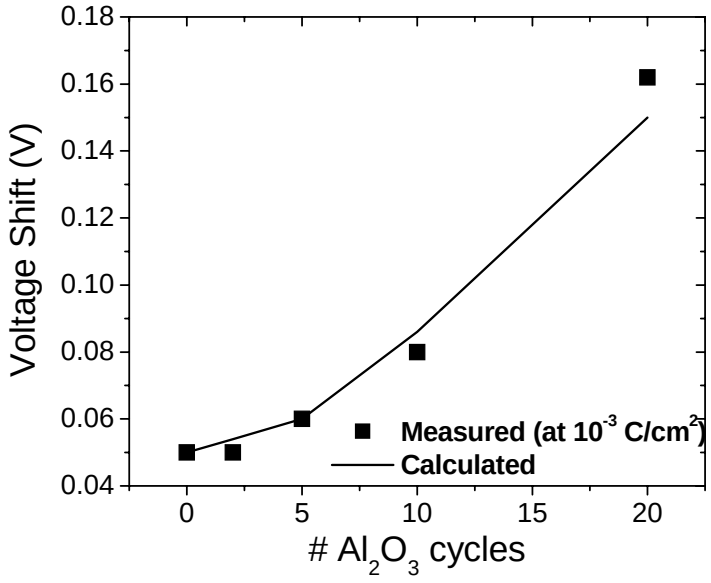


Figure 3.14. Theoretical voltage shifts versus measured data.

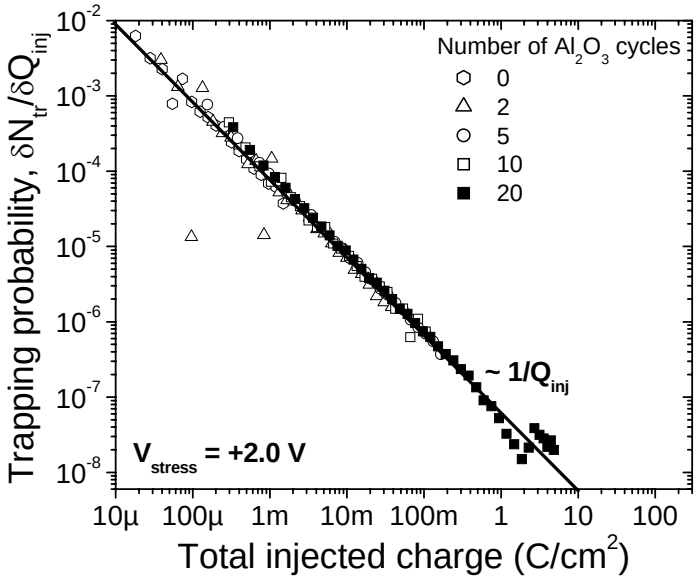


Figure 3.15. Trapping probability as a function of injected charge for Hf-silicate layers with Al_2O_3 caps of various thicknesses (0, 2, 5, 10, 20 ALD cycles).

In the foregoing analysis, we have shown through a simple analysis of electrical characterization data, a quantitative estimate of the location and magnitude of trapped charge may be obtained. It is instructive to note here that the simple analysis stated above does not determine uniquely both the location and magnitude of trapped charge, but what it uniquely determines is the first moment of the charge centroid. However, it accomplishes one important goal: It should be noted here that there is no restriction here on the sign of $\underline{x}$. $\underline{x}$ can either be positive or negative, in case of the latter, charge would be located in the Al_2O_3, and not in the Hf-silicate. So a positive $\underline{x}$ completely precludes the possibility of charge being located in the Al_2O_3 capped layer. This is turn implies that if Al_2O_3 capped layers are used for workfunction tuning on high-κ, charge trapping initiated by the capping layers should not be a cause of concern.

If I_g-V_g sweeps are used as a secondary stress monitor, the charge location and magnitude may be more convincingly determined [14]. However, we anticipate two problems with this approach:

(1) Most of the gate stacks used today have EOT's in the direct tunneling regime. In order to credibly use the I_g-V_g sweeps as a parallel stress monitor, the stress current relaxations in the I_g-V_g sweeps needs to be converted into voltage shifts. This implies dealing with direct tunnel models which can be inaccurate [16].

(2) On top of that, as illustrated in Figure 3.16, there may be substantial charge loss between two successive I_g-V_g sweeps between stresses, which opens up the broader question of transient charging and charge loss during conventional

measurements, which we discuss when we take up the in situ sense at stress technique in the following section.

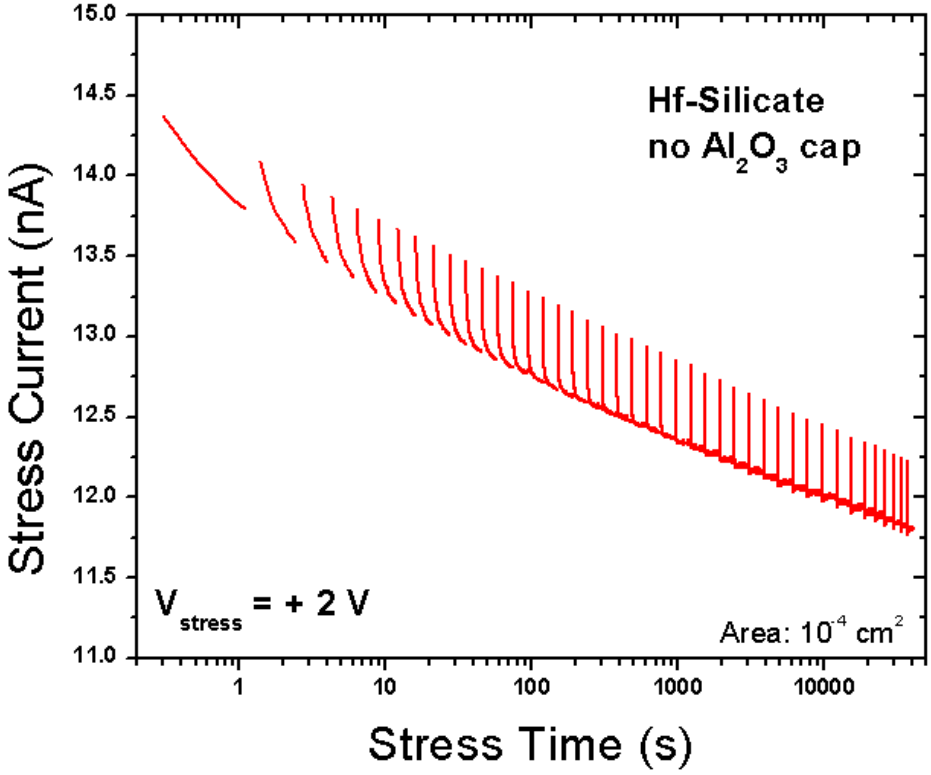


Figure 3.16. Severe stress current relaxation is observed at small time scales, suggesting fast/transient charge loss.

When the flatband voltage shifts are plotted as a function of time, rather than as a function of injected charge as shown earlier in Figure 3.12, we notice that bulk of the voltage shift appears at very short time scales, namely in the order of ~ms. This is markedly different from the case of SiO_2 where such fast transient charging is completely absent. This is aptly illustrated in Figure 3.17, where a direct comparison between flatband shifts as a function of charge injection time and injected charge is made. It is

seen that in the ~ms time scale, charge injection is on the order of mC/cm^2, a very large magnitude.

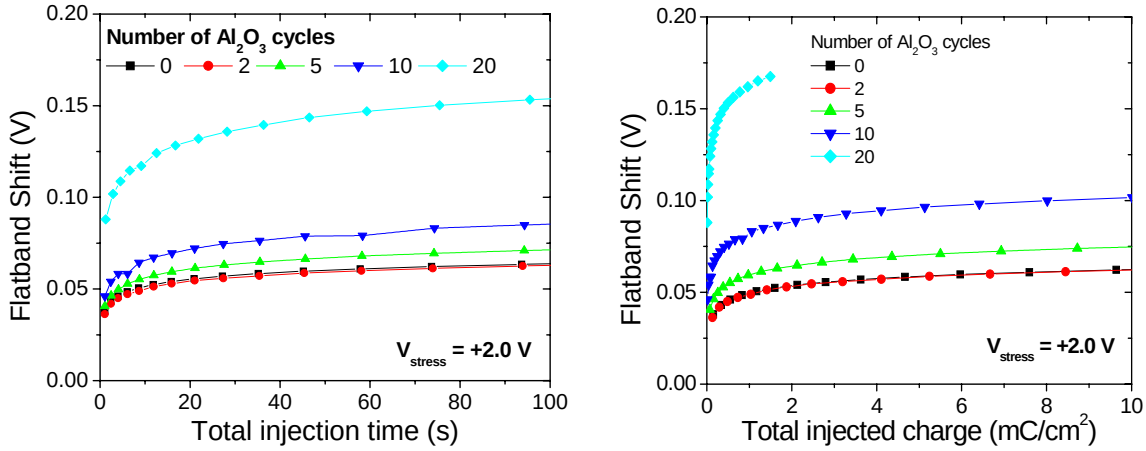


Figure 3.17. A direct comparison between flatband voltage shifts as a function of total injection time and total injected charge.

3.6 The In Situ Sense-at-Stress Method

For high-κ gate stacks, where there is fast transient charging, a better way to quantify the instability is by monitoring drain current degradation as and when the transistor is stressed. The instability in the drain current can come from either of the following two sources:

1) loss or gain in inversion charge due to charge trapping effects.

2) Enhancement in the scattering mechanism due to generation of scattering centers like interface states.

In the following analysis, we will conclusively show that the second factor namely impact of transient charging on transport in the channel or mobility is not a cause of concern.

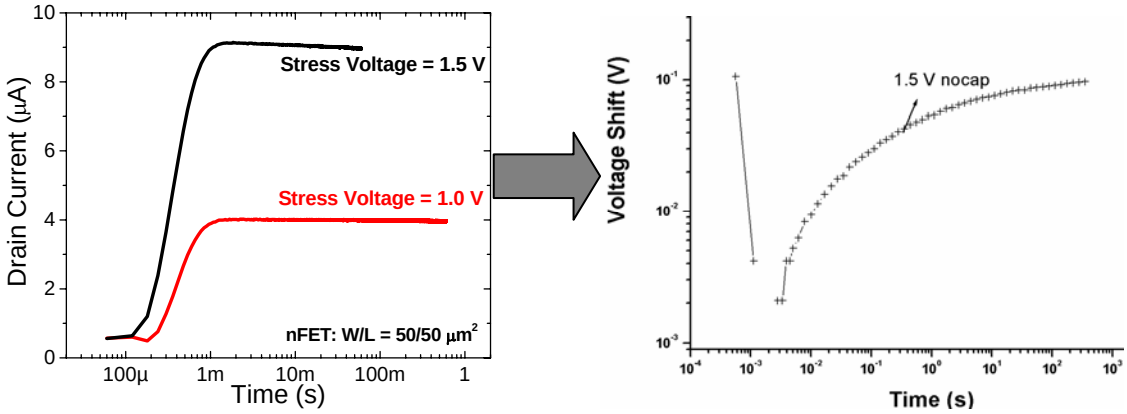


Figure 3.18. On the 4 nm Hf-silicate sample, the in situ sense at stress technique is carried out, and the subsequent drain current degradation is converted into voltage shift (only the voltage shift corresponding to the 1.5 V stress is shown)

The drain current degradation depicted in Figure 3.18 is converted into a voltage shift by using the transconductance of the MOSFET, as depicted by equation (3.2),

$$\Delta V_t = \Delta I_d \times (\delta I_d/\delta V)^{-1} \qquad \text{.....(3.2)}$$

It is assumed that the transconductance, g_m does not change during stress, a fair assumption at low stress voltages. A second implicit assumption made during the calculation of voltage shift from in-situ sense at stress measurement technique is constancy of mobility during the stress, an assumption which might lead to errors in the value of voltage shifts so obtained. However, this technique illustrates the fast voltage

shift that occurs at very short time scales namely ~ ms. This is better exemplified in Figure 3.19, where it is seen 90% of the voltage shift occurs in the first 1000 seconds.

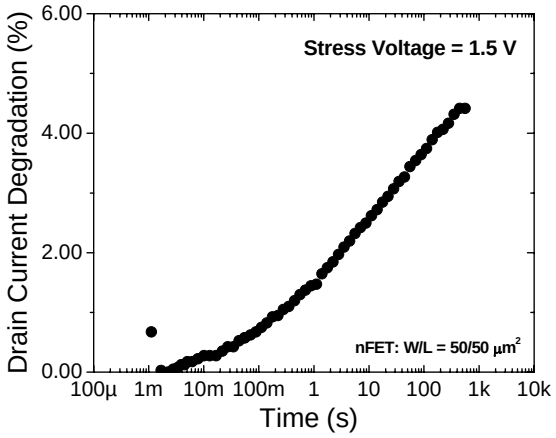


Figure 3.19. Nearly 90% of drain current degradation is observed in the first 1000 seconds with this technique.

A direct comparison of the stress and sense and the in-situ technique in Figure 3.20, where we find out that there is about 100X difference between the stress and sense determined voltage shift and voltage shift determined by the in situ technique.

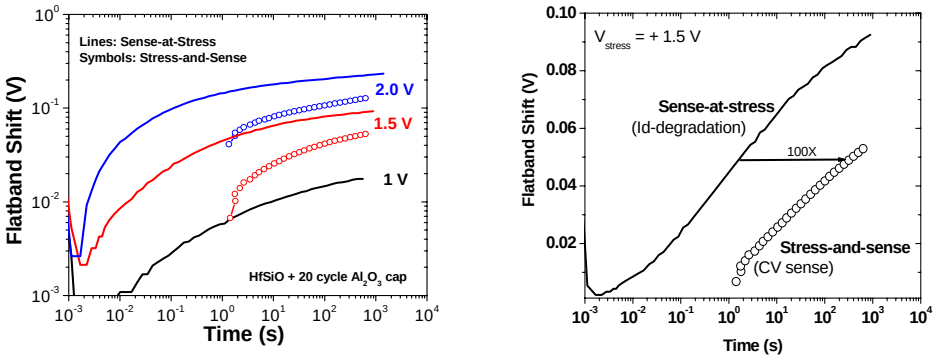


Figure 3.20. Substantial difference between stress and sense determined voltage shift and that determined by the in-situ technique (20 Cycle Alumina cap).

This kind of prompt voltage shift (~40 mV at 1.5 V in about 1 s) could be a limiting factor for these stacks. This transient trapping instability is better captured in charge pumping measurements when using ~μs order rise and fall times, which has been identified earlier in Chapter 2 as bulk trapping. This is demonstrated in Figure 3.21 below.

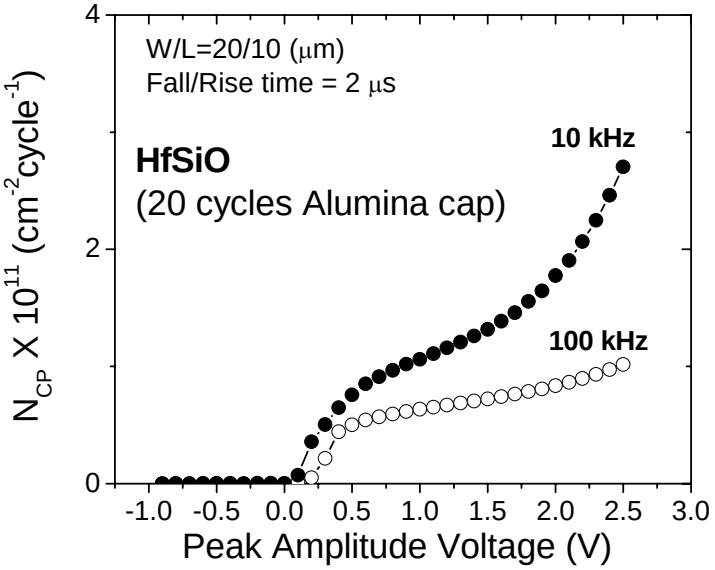


Figure 3.21. Charge Pumping characteristics of 20 cycle Al_2O_3 capped Hf-silicate based NMOSFET with ~μs rise and fall times. Bulk trapping at around 1.5 V can be directly linked with the transient trapping instabilities evident with the in situ technique discussed earlier.

Now, we focus our attention on the second point-namely the impact of transient charging on transport in the channel. As illustrated in Figure 3.22, by monitoring the evolution of mobility with gate voltage simultaneously with the amplitude sweep based charge pumping curve, we see the region of the mobility curve where bulk trapping comes into play (namely the high field region of the mobility curve at higher gate voltages), mobility

is already expected to be limited by surface roughness scattering. If transient charging impacts transport in the channel, it has to be through a Coulombic interaction which is important only in the low field region of the mobility curve.

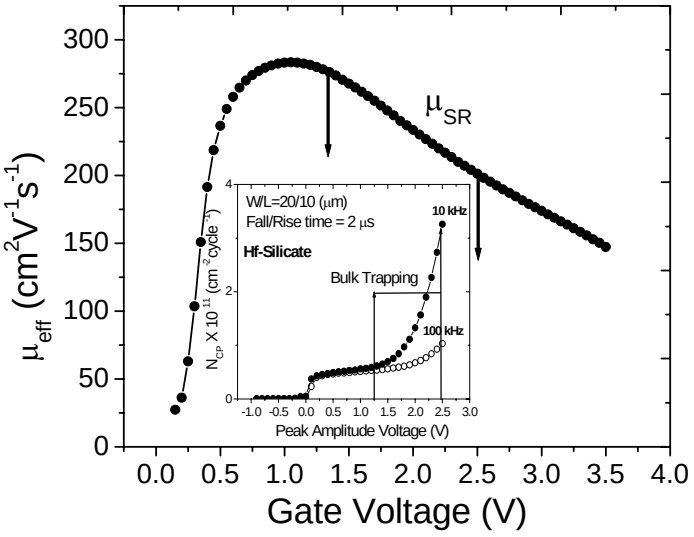


Figure 3.22. Mobility already limited by surface roughness when bulk trapping comes into play.

It is instructive to note that this is a representative case for a representative high-κ gate stack. We have observed the same phenomenon with all other high-κ gate stacks. So we may safely and definitively conclude that transient trapping as manifested by bulk trapping is not expected to adversely impact the transport in the channel.

3.7. Impact of Remote Charge on Mobility

We have shown conclusively in the foregoing analysis in this chapter, that traps in the high-κ gate stack can affect the device quasistatics like threshold voltage and flatband voltage by trapping charge during substrate injection. This component of the charge was

identified as stable charge. We have also identified a limiting factor for these stacks namely transient charging or charging at very small time scales. We have also shown that this kind of fast transient charging is not a concern for transport in the channel. Now, we try to investigate the impact of remote stable charge on transport in the channel. For instance, by electrical stressing we would introduce remote charge in the high-κ gate stack itself and then study the impact on mobility. It should be emphasized at this point

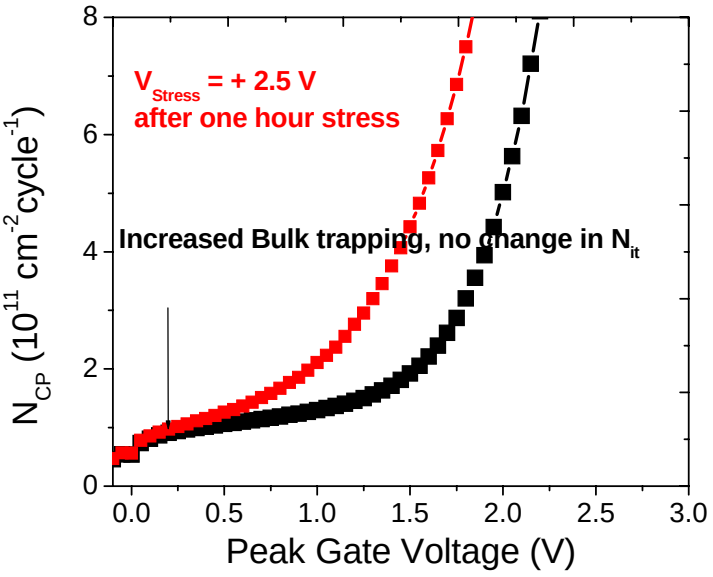


Figure 3.23. Increased Bulk Trapping with positive stress under negligible change in interface state density (N_{it}) for Hf-silicate NMOSFETs.

that as pointed out in [14] under positive stress (substrate injection), bulk traps are created under minimal change of N_{it}. This is suitably demonstrated in Figure 3.23 for the Hf-silicate NMOSFET, where by direct measurement of trapped charge, we show indeed under positive stress, there is marked increase in bulk trapping. However, in context of Figure 3.15, the trapping probability tends to go towards zero at larger time scales. This is attributed to the fact that under the voltage stress conditions used so far in this study,

namely in the range of 2.0-2.5 V, there is filling up of pre-existing bulk traps, rather than creation of new ones. Despite the best of our efforts, with the stress and sense measurement technique, it was difficult to stress the samples with a stress voltage higher than 2.5 V controllably. So we chose a different experimental methodology, as described in the flowchart in Figure 3.24.

First, a routine CV upsweep and downsweep is carried out on the virgin device in the inversion split-CV mode. Then, the device is stressed at a very high voltage (4.0 V) and the measurement paused immediately for a full minute, and the measurement resumed. The downsweep is followed immediately by an upsweep and the CV recorded. This engineers a CV shift of about 160 mV between pre and post stress conditions as displayed in inset of Figure 3.25. The parallel shift in CV (no stretchout) clearly suggests generation of stable bulk charge only and no N_{it}. DC I_d-V_g's were carried out on an identical adjacent device under the same measurement conditions, and mobility pre and post-stress computed. The 160 mV shift in CV can be easily converted into charge, by assuming the charge to be located in the middle of the high-κ gate stack for instance. By this method, the estimated charge is found to be fairly high on the order of 1 X 10^{12} cm^{-2}. This is seen to have minimal impact on mobility.

As briefly discussed in Chapter 1, the presence of remote charge is expected to have little impact on mobility [17,18]. The effect of remote charge is further minimized by screening of the remote charge by the carriers in the gate as described in [19]. By the same token, the impact of remote charge on mobility should be less severe on the 0 cycle

Al_2O_3 cap Hf-silicate than on the 20 cycle one , because of the proximity of the poly-Si gate to high-κ stack in the former than the latter (Figure not shown)

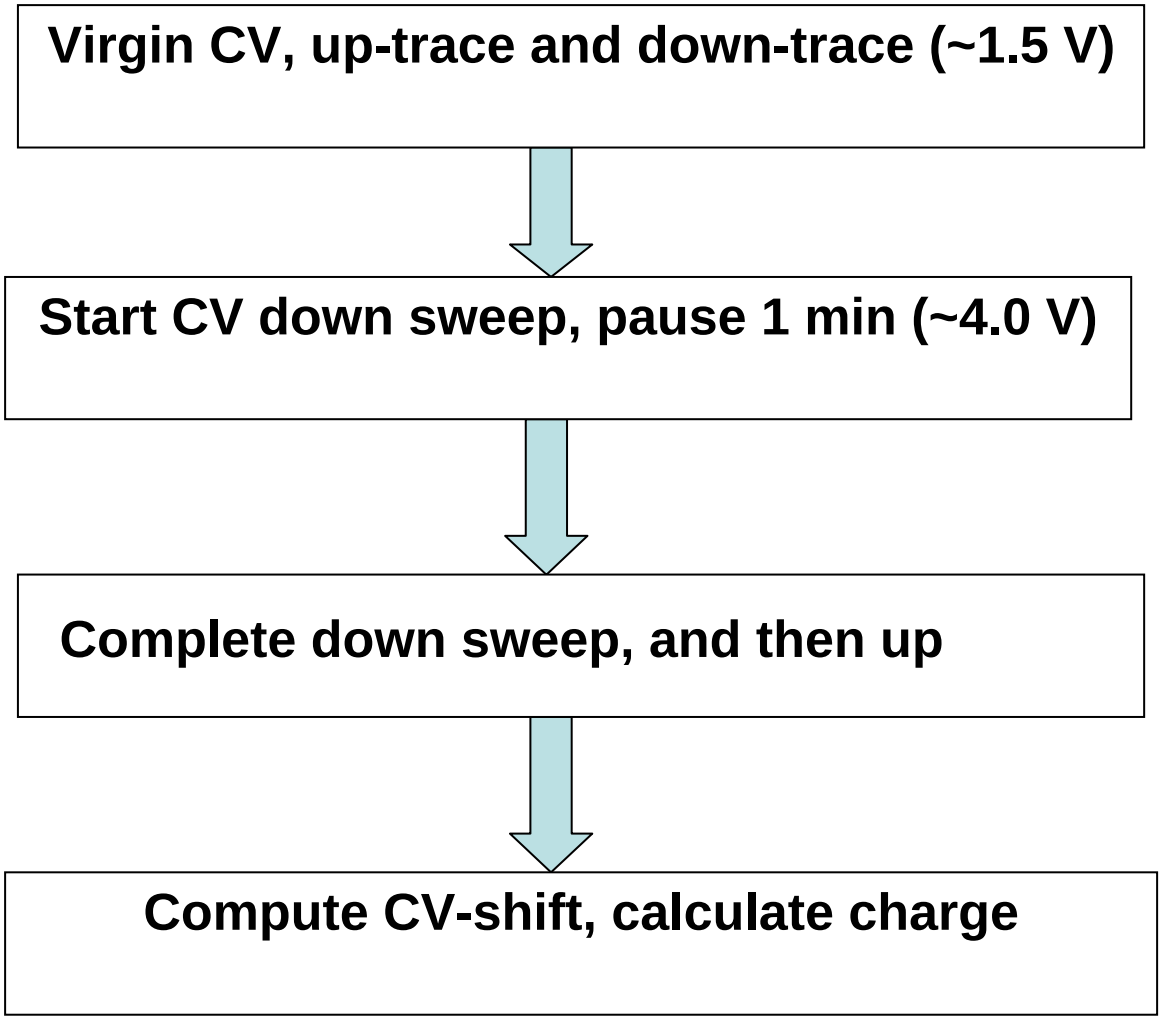


Figure 3.24. Flowchart showing the measurement scheme used to introduce bulk remote charge in the high-κ gate stack

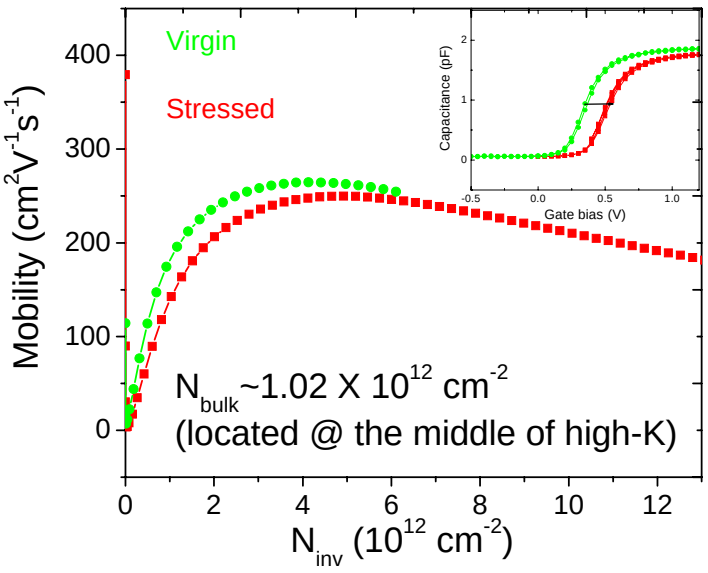


Figure 3.25. Remote charge has nominal impact on mobility on the Hf-silicate based NMOSFET (**Inset:** Stress induced CV-shift is used to calculate remote bulk charge)

3.8. Conclusions

A simple methodology based on standard electrical characterization techniques is proposed to track the location and magnitude of trapped charge, whence, the first moment of the charge centroid is uniquely determined. It is conclusively shown that charge is in the Hf-silicate gate stack itself, and charging from the Al_2O_3 capping layer is not a cause of concern, which in turn implies that this approach may be used for V_t adjustment with high-κ if it can be scaled properly. The oxygen barrier properties of Al_2O_3 also makes it attractive for pFET like workfunction tuning with metal gates under minimum EOT

penalty [20]. Concerns remain about intermixing of the Al_2O_3 cap with the underlying dielectric and subsequently altering the phonon spectrum and hence the phonon limited mobility. This will be discussed in Chapter 4.

By a novel electrical experiment, we confirmed the theoretical projections about the impact of remote charge on mobility in the bulk-Si channel. For the first time, it is definitively shown that charge far away from the channel has little impact on mobility. However, the role of interface states or N_{it} in this context assumes added significance, since they are adjacent the channel electrons. In the next chapter we investigate thoroughly the role of N_{it} on mobility.

References

[1]. C.C Hobbs et. al., "Fermi Level pinning at the poly-Si/metal-oxide interface-Part I", 2004, IEEE Transactions on Electron Devices, Vol. 51, No. 6, pp. 971-977.

[2] T. Iwamoto et. al, "A highly manufacturable low power and high speed HfSiO CMOS FET with dual poly-Si gate electrodes", 2003, IEDM Tech. Dig., pp. 27.5.1-27.5.4.

[3] E. A Cartier et. al., "Systematic study of pFET V_t with Hf-based gate stacks with poly-Si and FUSI gates", 2004, VLSI Technology Symposium, pp. 44-45.

[4] M. Kadoshima et. al., "Fermi Level pinning engineering by Al composition Modulation and doped partial silicide for $HfAlO_x$ (N) CMOSFETs", 2005, VLSI Technology Symposium, pp. 70-71.

[5] S. H Bae, C.H Lee, R. Clark, and D.L Kwong, "MOS characteristics for ultra thin CVD HfAlO gate dielectrics", 2003, IEEE Electron Device Letters, Vol. 24, No. 9, pp. 556-558.

[6] C.C Hobbs et. al., "Fermi Level pinning at the poly-Si/metal-oxide interface-Part II", 2004, IEEE Transactions on Electron Devices, Vol. 51, No. 6, pp. 978-984.

[7] Private Communication (SEMATECH)

[8] A. Kerber, E. Cartier, G. Groeseneken, H. E Maes, and U. Schwalke, "Stress induced charge trapping effects in SiO_2/Al_2O_3 gate stacks with TiN electrodes", 2003, Vol.94 No. 10, pp. 6627-6630.

[9] L. A Ragnarsson, S. Guha, N. Bojarczuk, E. Cartier, M.V Fischetti, K. Rim, J. Karasinski, "Electrical characterization of Al_2O_3 nMOSFETs with Al gates ", 2001, IEEE Electron Device Letters, Vol. 22, No. 10, pp. 490-492.

[10] A. Kerber et. al., "Origin of threshold voltage instability in SiO_2/HfO_2 dual layer Gate Dielectrics", 2003, IEEE Electron Device Letters, Vol. 24, No. 2, pp. 87-89.

[11] ITRS Technology Roadmap, 2004 (http://public.itrs.net)

[12] S. Zafar, A. Callegari, M. V Fischetti, and E. Gousev, "Charge Trapping related Threshold voltage instabilities in high permittivity gate dielectrics", 2003, Journal of Applied Physics, Vol. 93, No. 11, pp. 9298-9303.

[13] K. Maitra et. al., "On location and magnitude of trapped charge in ALD Alumina capped Hf-silicate gate stacks", 2005, ECS meeting, Ottowa.

[14] A. Kerber, "Methodology for Electrical Characterization of MOS devices with Alternative Gate Dielectrics", PhD Thesis, T.U Darmstadt, and IMEC Belgium, 2003.

[15] M. V Fischetti, "Generation of positive charge in SiO_2 during avalanche and tunnel Electron injection", 1985, Journal of Applied Physics, Vol. 57, No. 8, pp. 2860-2879

[16] W. C Lee, and C. Hu, "Modeling CMOS tunneling currents through ultrathin gate oxide due to conduction and valence band electron and hole tunneling", 2001, IEEE Transactions on Electron Devices, Vol. 48, No. 7, pp. 1366-1373.

[17]. S. Saito, K. Torri, M Hiratani, and T. Onai, "Improved Theory for Remote Charge scattering limited Mobility in metal-oxide-semiconductor Transistors", Applied Physics Letters, 2002, Vol. 81, No. 13, pp. 2391 - 2393.

[18].W. Zhu, J. P Han, and T. P Maa, "Mobility Measurements and Degradation Mechanism of MOSFETs made with Ultrathin High-κ Dielectrics", IEEE Transactions on Electron Devices, 2004, Vol. 51, No. 1, pp. 98 – 105.

[19]. M. V Fischetti, and S. E Laux, "Monte Carlo Study of electron transport in Si Inversion layers", Physical Review B, 1993, Vol. 48, pp. 2244 - 2274.

[20]. E. A Cartier et. al., "Role of V_{FB}/V_t stability of pFET metals on HfO_2", VLSI Technology Symposium, 2005, pp. 230-231.

Chapter 4

Impact of Interface States (N_{it}) on Mobility in Advanced High-κ Gatestack based NMOSFETs

4.1 Introduction

In chapter 3, we have conclusively shown that remote charge or "charge at a distance" has minimal impact on transport in the channel. However, we have also emphasized that in this context, the role of interface states (N_{it}) becomes extremely critical. Because of the proximity of the N_{it} to the channel electrons, the N_{it} can easily create additional scattering events via Coulombic coupling with the channel electrons than remote charge induced Coulombic coupling. It is instructive to note that the principal scattering mechanisms in the channel can be classified into three broad categories, namely, Coulomb, phonon, and surface roughness. In this chapter, we focus on the Coulombic scattering mechanisms. The Coulombic scattering mechanisms can be further sub-divided into three categories- response of the N_{it}, response of the dopants in the channel or ionized impurity scattering and remote charge. Remote charge induced scattering mechanisms have already been dealt with in Chapter 3, and it has been shown that this mechanism can largely be ignored. We trace the early SiO_2 literature on oxide charge scattering, and ionized impurity scattering in the next section. Then we give the motivation behind studying the N_{it} limited scattering mechanisms in the context of advanced high-κ/metal gate stacks. We also emphasize that SiO_2 is a perfect system to study the impact of N_{it} on the channel electrons. We also allude to the ESR data in literature to draw a parallel between N_{it} in SiO_2 based NMOSFETs and HfO_2 based NMOSFETs. A novel thermal desorption experiment is done to calibrate the N_{it} response on mobility , wherein N_{it} are introduced

at the SiO_2/(100)-Si interface and its impact on mobility properly calibrated. A test of accuracy to properly study its impact in high-κ based NMOSFETs is performed. The calibrated N_{it} response is finally applied to advanced high-κ/metal gate MOSFETs and the mobilities in absence of N_{it} under different processing conditions are compared. The role of nitrogen induced fixed oxide charge, which behaves very much like N_{it}, on mobility of channel electrons in high-κ insulator based NMOSFETs is also studied on a controlled set of samples.

4.2 Motivation

We consider a very typical situation observed in aggressively scaled high mobility metal gate high-κ NMOSFETs (Figure 4.1). The high-κ gate stack in question is HfO_2 (T_{inv}~1.6-1.7 nm, EOT ~ 1.2-1.3 nm) and the metal gate is TaSiN (n+ metal) with FUSI NiSi cap [1]. As briefly described in Chapter 1, it is seen that the as deposited stack (the processing details of this sample is given in [1]) has low mobility, but as soon as a high temperature treatment is given (1000 ^{0}C, 5 seconds) the mobility shoots up significantly with no change in EOT and V_t. This has been discussed in detail as mentioned earlier in [2]. However, the physical origin of this high mobility and the significance of the high temperature process step is not clearly understood. Carefully monitoring the charge pump characteristics (Figure 4.1) indicates that these stacks (the as deposited and 1000 ^{0}C annealed one) have different interface state densities (N_{it}). It is not clear whether this mobility difference can be understood from the difference in N_{it} alone. In order to answer that question clearly, it is necessary to know exactly the impact N_{it} has on mobility and hence factor out the N_{it} response completely and accurately and see whether the mobility

curves merge together. To do this, we consider a conventional 5 nm thick SiO_2 based NMOSFET and introduce N_{it} controllably and study its impact on the channel electrons. The justification behind doing this is given in Figure 4.2.

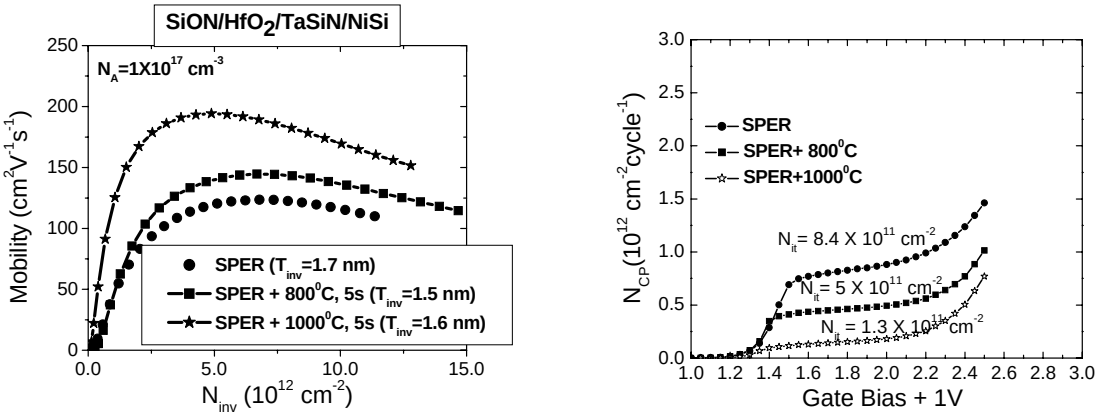


Figure 4.1. Impact of high temperature on mobility in aggressively scaled metal gate/high-κ NMOSFETs. It is seen that the different samples have different interface state densities (N_{it}).

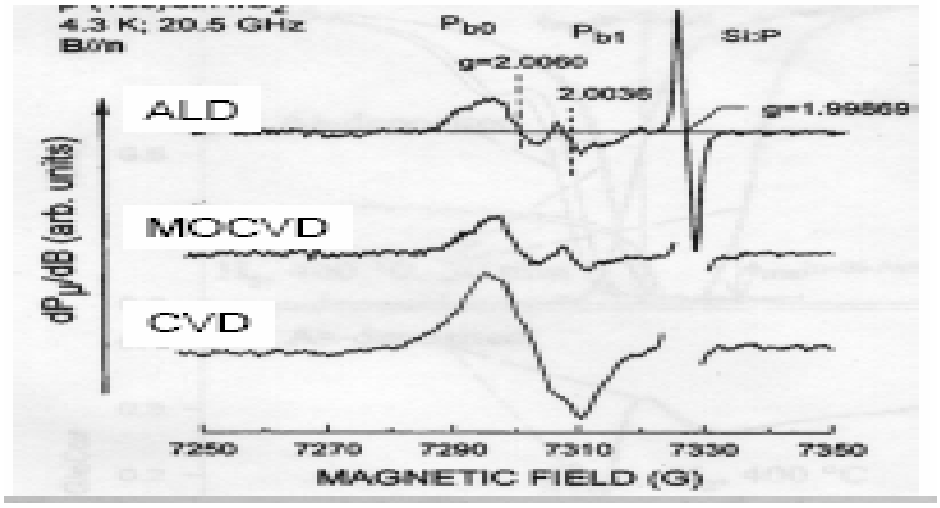


Figure 4.2. g-tensor values obtained from ESR spectroscopy show that regardless of deposition process, interfacial oxide layer thickness and nature, the interface between HfO_2 and (100)-bulk Si behaves like a poor quality SiO_2/(100)-Si interface [3].

As depicted in Figure 4.2, it is seen (based on ESR data as summarized in [3]), that a HfO_2/(100)-bulk Si and SiO_2/(100)-Si interfaces are very much alike. Or in other words, the N_{it} associated with the high-κ-HfO_2 and (100)-bulk Si interface and the SiO_2/(100)-bulk Si interface are one and the same. It is expected substantial insight into the N_{it} limited Coulombic mobility in high-κ based NMOSFETs may be obtained by the studying the N_{it} response of mobility in SiO_2 based NMOSFETs. This will be dealt with in more detail in section 4.3. In the following section, we trace the early literature on ionized impurity scattering, since as pointed out earlier, in order to study the impact of interface states exclusively on the Coulombic component of mobility, it is necessary to come up with a simple mechanism (preferably an analytical closed form equation depicting the role of ionized impurity) to decouple the role of ionized impurity and N_{it} within the Coulombic scattering limited mobility expression. With this goal in mind, we do a brief literature survey of the relevant SiO_2 literature in the next section. An excellent summary of the development of analytical techniques to model Coulombic scattering in MOS channels in presence of SiO_2 gate dielectric may be found in [4]. In the following section, we give a summary of the review in [4], and use the approach developed in [4] to calibrate the impact of ionized dopants in the Si channel.

4.3 The Ionized Impurity Scattering

The scattering of inversion layer electrons by charge centers in its vicinity is generically termed as Coulombic scattering. Possible sources of charge centers include ionized impurities in the channel, interfacial charge, fixed oxide charge, and mobile oxide charge.

Since we are only interested in studying and modeling ionized impurity scattering, in this section, Coulombic scattering will be synonymous with impurity scattering [4].

The three parameters that affect impurity scattering in MOS inversion layers are ionized impurity concentration, carrier density and temperature. Charge carriers respond to an electrostatic potential in such a way as to reduce its strength. This effect is known as screening, and it is typically proportional to the density of mobile carriers. In the limit of low carrier concentrations, scattering is essentially due to bare Coulomb potential and is termed unscreened Coulomb scattering. With the increase of carrier density in the channel, as is the case in today's transistors (which are heavily doped to minimize short channel effects like punchthrough and DIBL (drain induced barrier lowering)) screened Coulomb scattering is of prime relevance today regardless of the nature of gate dielectric on top (either SiO_2 or high-κ).

Following the work of Brooks and Herring [5], and earlier Conwell and Weisskoff [6], three dimensional Coulomb mobility in presence of screening was computed. Though they gave analytical closed form solutions, they fell short on predicting the experimental data correctly. Ando, Fowler and Stern [7] presented a theoretical treatment of Coulombic scattering in MOS inversion layers assuming that only the lowest sub-band i.e the ground state was occupied. However, given the rigor of the treatment, this unfortunately did not yield a closed form solution. Prior to Ando et. al., Sah et. al. [8] calculated closed form solution for unscreened Coulomb mobility in MOS inversion layers. However, the work only considered scattering from fixed oxide charges and not

scattering from ionized impurities. Subsequently, Ning and Sah expanded on this work [9] and included screening among other effects in their work, but it still dealt with fixed oxide charge scattering rather than ionized impurity induced scattering. Other works have appeared in literature [10-14] which treated impurity scattering in a quantum well rather than MOS inversion layers. While the structure is different, the problem is essentially similar, since in both treatments, the electron gas is treated as a quasi-two-dimensional system. However, the major shortcoming of these analyses is that they treat Coulomb scattering almost exclusively in the low temperature limit. Gamiz et. al. [15] calculated the Coulomb scattering limited mobility but his treatment though exhaustive and detailed did not yield an analytical result. Syed Mujtaba Aon in 1995 [16] gave a concrete direction on the analytical technique necessary to calculate the ionized impurity scattering limited mobility. He conclusively showed that in presence of screening, the ionized impurity scattering limited mobility is directly proportional to the density of inversion carriers in the channel (N_{inv}) and inversely proportional to the doping density (N_D). In the following section, based on this work, and an experiment conducted on SiO_2 based NMOSFETs, we properly calibrate the proportionality factors.

4.3 The Thermal Desorption of Hydrogen experiment – Calibrating N_{it} Response on Coulombic Mobility

Impact of interface states (N_{it}) on mobility in SiO_2/Si NMOSFETs is studied. Varying levels of N_{it} are incorporated into the SiO_2/Si interface controllably by vacuum annealing the NMOSFETs at different temperatures. As opposed to electrical stressing, or irradiation, vacuum annealing creates interface states only, and not traps away from the SiO_2/Si interface. This study thus allows us to study exclusively the impact of N_{it} on

mobility. Coulomb component of the effective mobility and the N_{it} limited mobility are extracted using Matthiesen's rule. Theoretical fits for the Coulomb component and the N_{it} limited component are obtained using a parameterized equation in. Careful calibration of the fitting parameters, in particular the N_{it} limited scattering factor, β, in the equation is carried out. Implications of this work on understanding mobility degradation in advanced high-κ gate stacks are also discussed.

In this section, we carefully calibrate the impact of interface states (N_{it}) on mobility in 5 nm thick SiO_2 based NMOSFETs. Varying levels of N_{it} are incorporated into the devices by vacuum annealing the NMOSFETs either at different temperatures or for different times at a fixed temperature. Extensive studies have been carried out in the past on the mechanism of N_{it} generation by hydrogen depassivation in Si/SiO_2 interfaces (viz: by exposure of the Si "dangling bonds" or P_b centers [17,18,19]) through vacuum annealing. This method is preferred over other methods of interface state generation existing in the literature (viz: electrical stressing [20,21,22], and irradiation [23]) because this gives precise control on the location of the N_{it}'s viz: at the Si/SiO_2 interface. With the other two techniques, it is likely that traps may also be introduced away from the interface [24, 25] (bulk defects generated by hot carrier injection into the dielectric, the hot carriers being produced either by electrical stressing or irradiation) . Vacuum annealing under "controlled circumstances" (temperature, pressure, time) is an effective way of depopulating the P_b centers at the Si/SiO_2 interface. Interface states are electrically measured by amplitude sweep charge pump. Mobility measurements are carried out using a 100 KHz inversion split capacitance voltage (split-CV) (inversion charge is estimated by integrating the split-CV curve, refer to Chapter 2 and discussion

on mobility measurement techniques) , and dc I_d-V_g technique (the drive current is estimated from the I_d-V_g's). Coulomb component of the mobility is estimated using Matthiesen's rule.

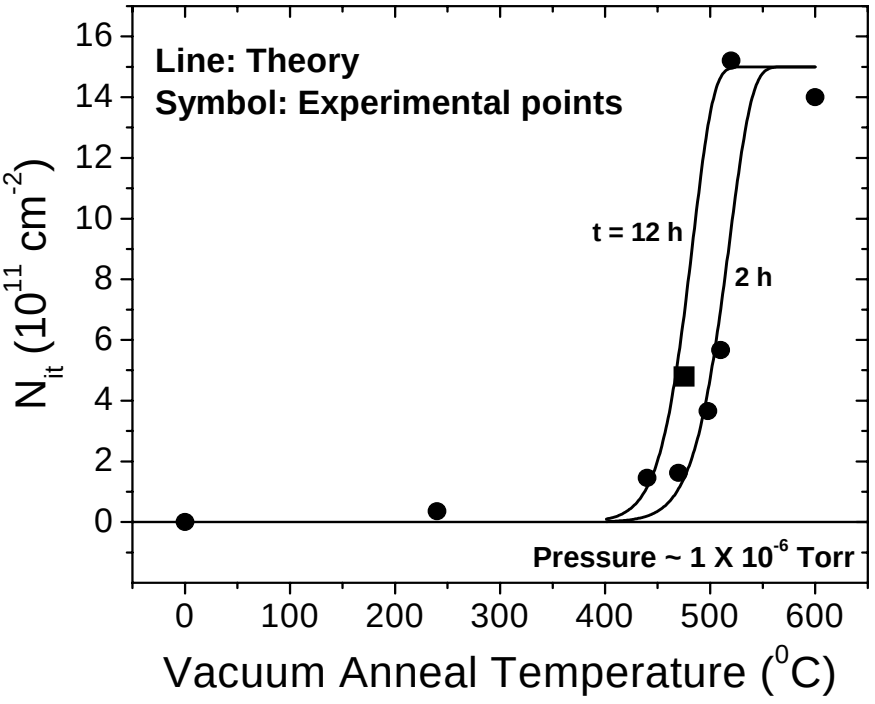


Figure 4.3. Method of Interface State Generation-Density of interface states as a function of vacuum anneal temperature with vacuum anneal time as parameter. The experimental curves are in close agreement with the theoretical curves generated from first order rate kinetics (N_{it} determined by charge pumping.)

NMOSFETs with poly-Si gates were fabricated on heavily doped (N_A=3 x 10^{17} cm^{-3}) (100) Si substrates using a standard self-aligned process. The devices were made directly probeable on silicide to ease hydrogen depassivation. Fairly thick SiO_2 (5 nm) was

chosen as the gate dielectric, to minimize remote charge scattering due to impurities in the gate [26]. At the chosen gate oxide thickness, channel mobility is independent of gate oxide thickness [27]. Full passivation of the "dangling bonds" (corresponding to N_{it}= 2.5 x 10^{10} cm^{-2} in Figure 4.3) is attained by annealing the transistors in forming gas at 460 ^{0}C. Variable N_{it}'s are introduced by annealing the transistors at different temperatures (470-600 ^{0}C) under low pressure (~1.0 x 10^{-6} Torr), with full depassivation (corresponding to N_{it} = 1.7 x 10^{12} cm^{-2} in Figure 4.3) attained at 600 ^{0}C. As a proof of the fact that only N_{it} are introduced, we also study the evolution of the subthreshold slope as a function of N_{it} and find the linear dependence of subthreshold slope on N_{it}. (Figure 4.4). It is also shown in Figure 4.4 that the threshold voltage does not change much as a function of N_{it} so generated.

N_{it} measurements (shown in Figure 4.3) are carried out using amplitude sweep charge pump (CP). Short transistors with channel length (L) of 10 μm and width (W) of 20 μm are chosen for CP measurements in order to evade the geometrical component. Slow fall time pulses (~2 μs) with frequencies of 10 KHz are applied at the gates with source and drain of the transistors grounded, the charge pumping current being measured at the substrate.

Standard 100 KHz split-CV technique is employed to get an estimate of the channel inversion charge [28] (capacitive response measured between the gate and the source + drain connected together, the substrate being grounded). dc I_d-V_g's are used (with drain biased to 50 mV) to obtain drive current in the linear region. Mobility is determined using the formulae in [20] (see Chapter 2, Figure 2.11).

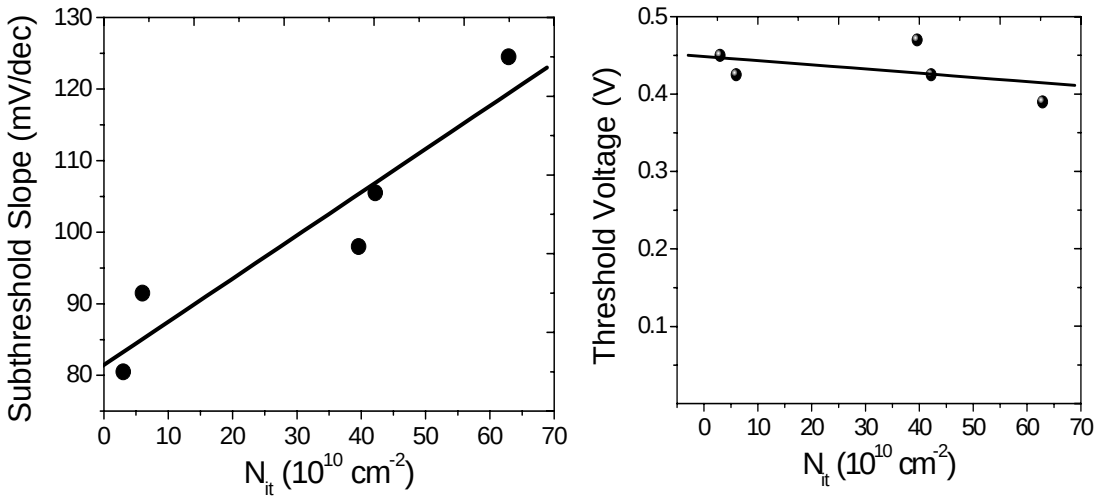


Figure 4.4. Evolution of subthreshold slope and threshold voltage as a function of generated N_{it}.

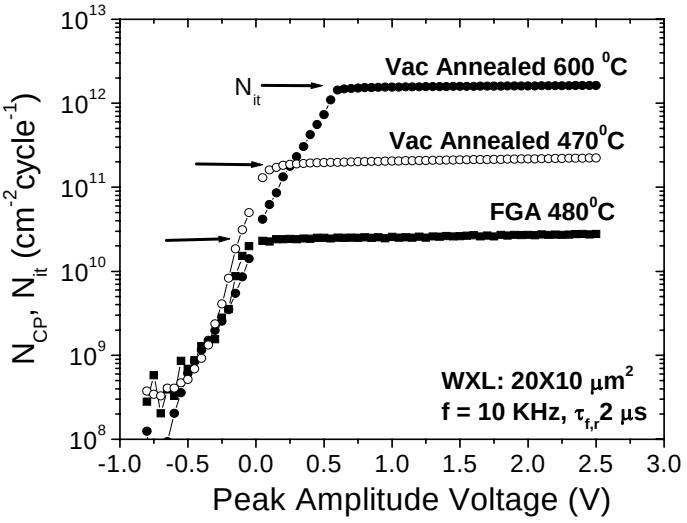


Figure 4.5. Amplitude Sweep charge pump performed to estimate the density of interface states.

Figure 4.3 depicts the evolution of N_{it}'s with vacuum anneal temperature. Anneals of 2 hour and 12 hour were performed at different temperatures. The N_{it} values so obtained follow the theoretical double exponential curve in [19] obtained from first order rate kinetics. The results of the CP measurements are given in Figure 4.5. Three representative curves are shown. It is instructive to note that once the curves saturate, they become completely independent of the peak amplitude voltage indicating that time dependent charge trapping is not present. It was also observed that the split CV curves (figure not shown) on devices with large N_{it} showed no hysteresis. This is a clear indicator of the fact that slow charge trapping is absent. Figure 4.6 depicts the mobility curves for different N_{it}. The inset shows the peak mobility calibration curve. Over 40% degradation in peak mobility is observed by changing the N_{it} value from 2.5 x 10^{10} cm^{-2} to 1.7 x 10^{12} cm^{-2}. It is a well known fact that the low field mobility is dominated by Coulomb scattering. Assuming the mobility to be dominated by three principal scattering components viz: Coulomb (Coul) , surface roughness (SR), and phonon (Ph), we can express the effective mobility using Matthiesen's rule as in equation (4.1),

$$\mu_{eff}^{-1}=\mu_{Coul}^{-1}+\mu_{SR}^{-1}+\mu_{Ph}^{-1} \qquad \text{……..(4.1)}$$

The surface roughness (equation (4.2)) and the phonon component (equation (4.3)) are modeled using the semi-empirical expressions from [12].

$$\mu_{SR}=\delta \exp[-(T/T_0)^2]/(E_{eff})^2 \qquad \text{.....(4.2)}$$

$$\mu_{Ph}=1470[1+(E_{eff}/7X10^{4})^{0.2}]^{-1} \quad \text{……..(4.3)}$$

where, $\delta = 8.8 \text{ X } 10^{14} \text{ Vs}^{-1}$, $T_0 = 500$ K and $T = 300$ K

where the effective electric field in the channel E_{eff} is expressed in V cm^{-1}. E_{eff} is calculated using formulae in [20]. The resulting Coulomb component extracted using equations (4.1), (4.2), and (4.3) is modeled by Villa's parameterized form [29] as depicted in equation (4.4). The Coulomb component is assumed to be dominated by two scattering components, ionized impurity due to the dopants in the Si substrate (α), and N_{it}'s present at the Si/SiO_2 interface (β). A parameter γ is introduced to capture the dependence of the Coulomb component on inversion charge (N_{inv}). To the first order, the Coulomb component is found to vary linearly with inversion charge (N_{inv}). The parameter α is the thermal length with a value of 10^{-11} cm (Theoretical studies show that electrons in a quantized state scatter with the charge centers located within a thermal length [30]).

$$\mu_{Coul}=\gamma N_{inv}/(\alpha N_A+\beta N_{it}) \quad \text{……..(4.4)}$$

The parameter γ and the N_{it} scattering factor, β are obtained by non-linear regression analysis as shown in Figure 4.7 ($\gamma = 1.992 \text{ X } 10^{-3}$ $cm^3V^{-1}s^{-1}$, $\beta = 1.9 \text{ X } 10^{-5}$ +/- $2.5 \text{ X } 10^{-6}$). The deviation of the experimental points from the theoretical curves at low N_{inv} values is attributed to the limitations of the measurement technique involved. As expected, N_{it}'s are seen to impact the Coulomb component of mobility very strongly. In order to decouple the impact of N_{it} and ionized impurity, the Coulomb component for a fixed

value of N_{inv} (= 4 X 10^{11} cm^{-2}) is plotted as a function of N_{it} (inset of Figure 4.7). It is instructive to note that at low N_{it}, Coulomb component of mobility is limited by ionized impurity scattering, whereas at high N_{it}, it is dominated by scattering from the interface states.

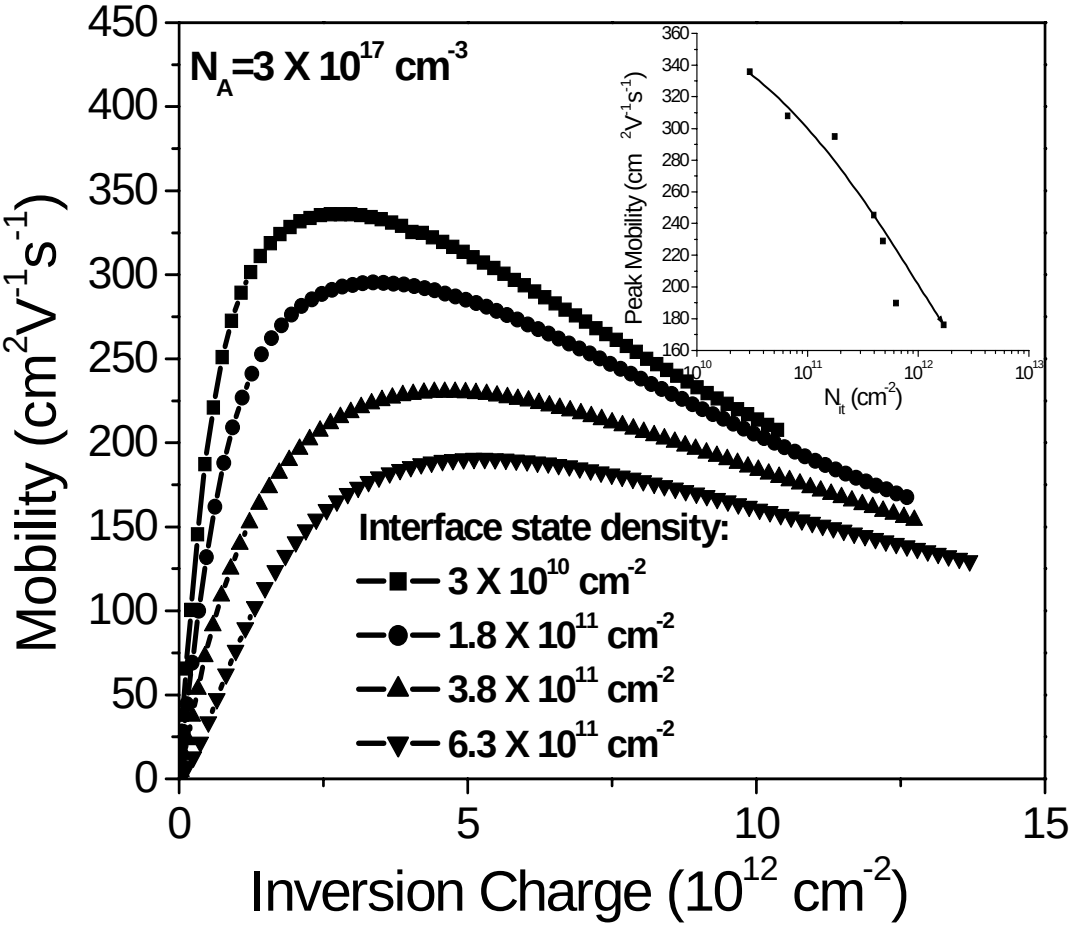


Figure 4.6. Mobility as a function of inversion charge for different interface state densities (**Inset:** Calibration Curve of peak mobility vs. interface state density)

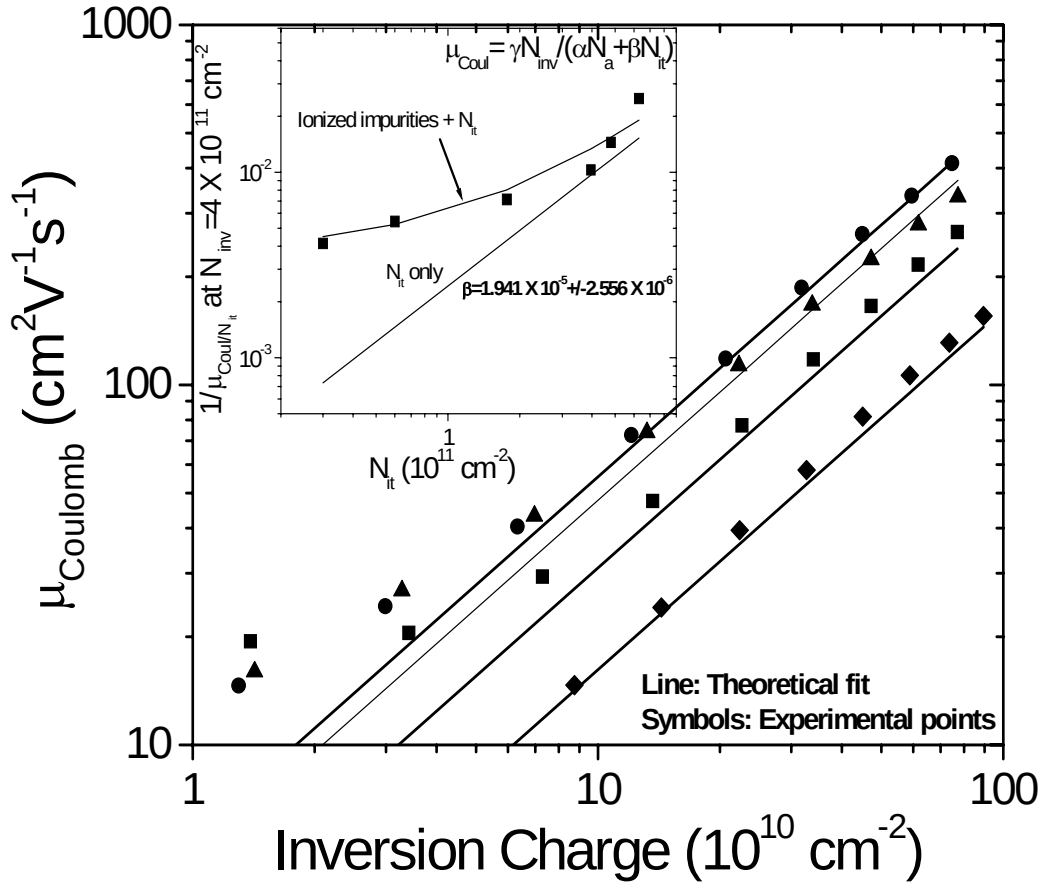


Figure 4.7. Coulomb components of Mobility extracted from Matthiesen's rule for different N_{it}'s (Circle: 3 X 10^{10} cm^{-2}, Triangle: 6 X 10^{10} cm^{-2}, Square: 1.76 X 10^{11} cm^{-2}, Diamond: 4.8 X 10^{11} cm^2). (**Inset:** Inverse of the Coulomb component and the Interface-state limited component of mobility plotted as a function of Interface state density. The parameterized forms of the Coulomb component (μ_{Coul}) and the N_{it} limited component (μ_{Nit}) are also given. (Lines: Theory, Symbols: Experimental points)

The connection between the present study and understanding charge trapping related mobility degradation in high-κ dielectrics stems primarily from the following two factors: (i) g-tensor studies based on ESR spectroscopy indicate that the N_{it} at the high-κ/SiO_2/(100)-Si are SiO_2/(100)-Si like [31,32,33]. (shown in Figure 4.2)

(ii) Charge trapping in high-κ may be broadly classified into two categories: response from N_{it} and response from the bulk traps, i.e traps located away from the interface [34].

Experimental evidence described in (i) justifies the assumption that interaction between inversion channel electrons and N_{it} in a high-κ/SiO_2/(100)-Si system is very similar to the same in a SiO_2/(100)-Si system or in other words, the response of N_{it} on the Coulomb component of mobility in high-K/(100)-Si NMOSFETs may be assumed to be identical to that in a SiO_2/(100)-Si NMOSFET. Similar studies conducted in the past did address the impact of N_{it} on mobility [20], however, as pointed out earlier in the text, interface states were generated either by electrical stressing or irradiation, which would also introduce traps away from the SiO_2/Si interface. In light of (ii) above, decoupling the impact of N_{it} and bulk traps on the Coulomb component of mobility in high-κ gate stacks is extremely important. The accurately calibrated N_{it} response parameter, β will provide fundamental insight exclusively into the role of N_{it} on mobility degradation in high-K dielectrics. Recent studies [35] also indicate that while studying mobility in high-K gate stacks, knowing N_{it} is important. This study will help compare mobilities in high-κ based advanced gate stacks with different N_{it}. In the next section, we test the applicability of the proposed scheme to high-κ gate stacks.

4.4. Applicability of N_{it} calibration technique to high-κ gate stacks

The interface property between the Si substrate and the gate dielectric is a crucial parameter for the MOSFET performance. Therefore, the Si/SiO_2 interface has been investigated intensely since the early stages of MOSFET fabrication. The high density of interfacial defects inherent due to the mismatch of the Si substrate and the thermally grown SiO_2 can be passivated by H_2/N_2 anneal. Values typically reported for a conventional SiO_2 stacks are in the low 10^{10} cm^{-2} range. However, when applying an electrical stress to the MOS device, the formation of N_{it} is commonly observed and is attributed to the release of Hydrogen from the Si/SiO_2 interface.

In dual layer high-κ stacks, formation of N_{it} is observed under electrical stress. Under negative electrical stress, even at very low fluences, significant generation of N_{it} is observed in Hf-silicate/SiON based NMOSFET with poly-Si gates as shown in Figure 4.8. The physical mechanism of interface state generation is similar to the one described in [36]. For gate injection, electrons injected from the poly-Si gate into the conduction band of the Hf-silicate layer is expected to reach the Si substrate with high energy, because of ballistic transport through the thin SiON layer. It is well known that such hot carriers are the cause of N_{it} generation in conventional SiO_2 gate stacks. The mechanism of N_{it} generation in this case may also be attributed to depassivation of the dangling bonds.

As shown in the inset of Figure 4.8, charge pumping measurements are used to track the evolution of N_{it} under negative stress polarity. Mobility before and after stress are

measured in a conventional stress and sense sequence described earlier in Chapter 2. From the two different mobility curves, using equation (4.4), the N_{it} calibration factor, β is determined, and it turns out to be of the same order as that determined from the thermal desorption experiment. This gives independent confirmation of our earlier assertion that N_{it} in high-κ/bulk Si interface is very much SiO_2/Si like.

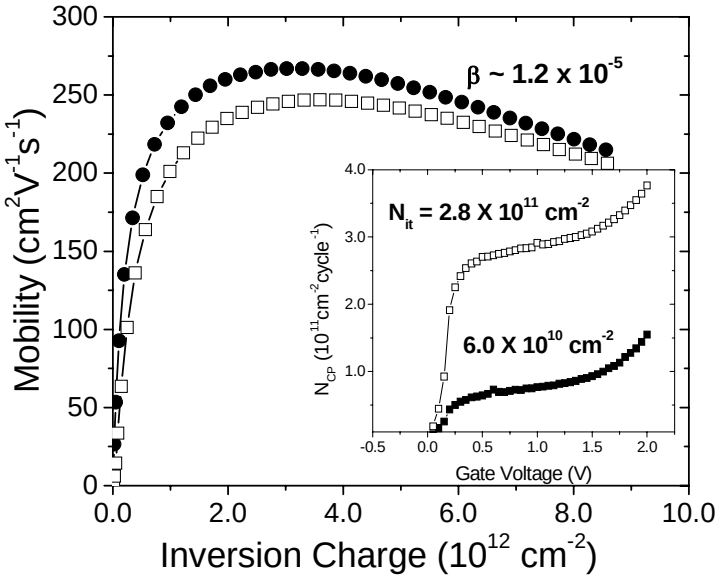


Figure 4.8. N_{it} calibration factor estimated from electrical stress experiment, mobility estimated by split-CV/I_d-V_g technique in between stress (**Inset**: N_{it} are determined from amplitude sweep charge pump measurement technique, where the charge pump is carried in between stress sequences with negative polarity of -2.5 V)

4.5 Limitations of Matthiessen's rule

To summarize the N_{it} limited mobility calibration mechanism, the thermal desorption of hydrogen experiment yields a value of $\beta \sim 1.9 \text{ X } 10^{-5}$, the electrical stress experiment gives a value of $\beta \sim 1.2 \text{ X } 10^{-5}$. Ning and Sah's work in the early seventies [8], based on calculations from perturbation theory predicted $\beta \leq 10^{-2}$. It is important to take a closer look at the possible origin of these discrepancies.

As early as 1980, Frank Stern [37] and Cham and Wheeler [38] in two back to back seminal articles in Physical Review Letters, showed that the method of separation of different scattering components using Matthiessen's rule has its limitations particularly at room temperature. It is well known that when mobility is computed for different individual scattering rates, the mean time of scattering is ascertained first for a given scattering mechanism (τ). That is the averaging over the range of energies is over τ, rather than over $1/\tau$. However, since mobility (μ) is proportional to τ, and in Matthiessen's rule, the different scattering components add inversely as illustrated in equation (4.1), there is no certainty that the left hand side will be equal to the right hand side in equation (4.1) unless spread in energy is very small, which happens only at low temperatures.

Frank Stern [37] calculated that even at a temperature of 40 K, the discrepancy between the left hand side and the right can be as high as 15%.

So, in the context of this study, we would like to emphasize that Matthiessen's rule can be extended thus far and no farther, particularly when dealing with it under room temperature conditions.

4.6 Application of N_{it} calibration equation to high mobility aggressively scaled TaSiN/HfO_2 gate stacks

We now go back to where we started in Section 4.2, where we have shown that for a FUSI capped TaSiN (metal gate), HfO_2 (high-κ) based NMOSFET, the N_{it} between the as deposited and the high mobility stack which saw high temperature process, are different. And in that context set out to develop the N_{it} calibration equation. Now, we apply the N_{it} correction to the said stack as depicted in Figure 4.9.

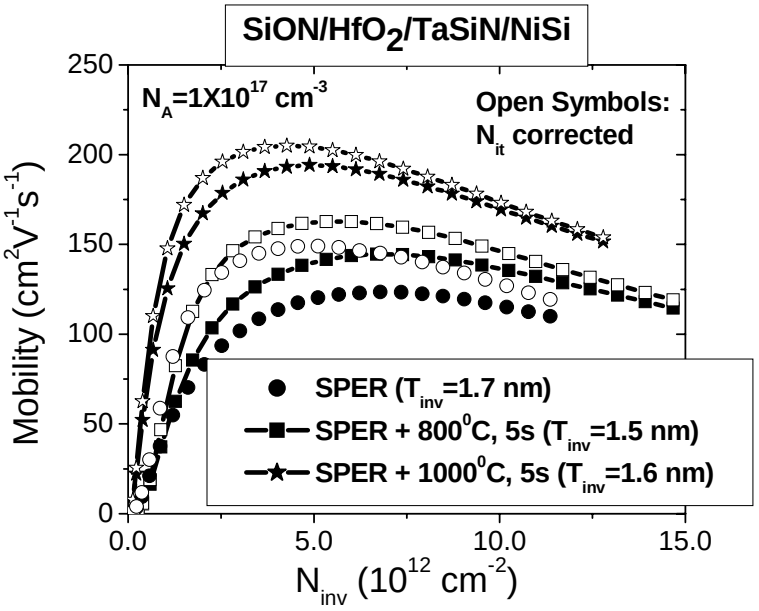


Figure 4.9. The N_{it} correction applied to aggressively scaled high mobility TaSiN/HfO_2 gate stack.

Interestingly, even after accounting for the difference in N_{it} between the as deposited and high temperature stack, the mobility curves do not merge together. Or in other words, the difference in mobility between the as deposited and the high temperature stack may not be explained by N_{it} alone.

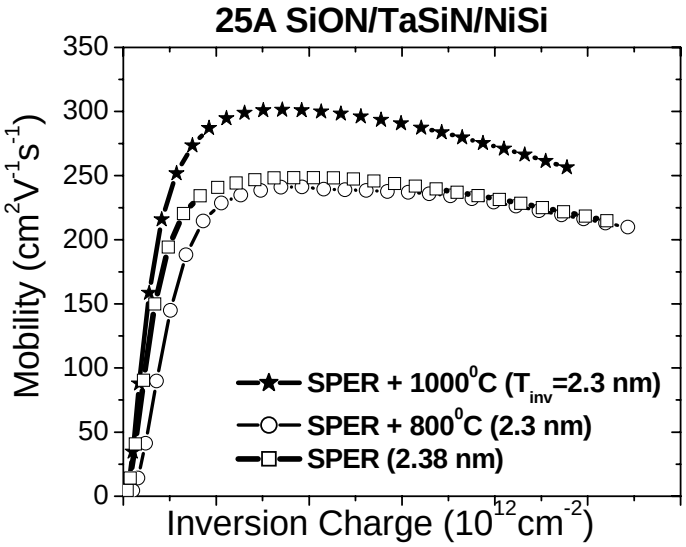


Figure 4.10. The oxynitride control (FUSI/TaSiN/SiON) for the TaSiN/HfO_2 stack shows similar behavior after high temperature treatment. Even after correcting for the different N_{it}, the curves do not merge together.

A careful look at the thicker oxynitride control for the FUSI/TaSiN/HfO_2 stack, namely, FUSI/TaSiN/SiON (T_{inv}~2.3 nm) stack demonstrates almost identical behavior under high temperature treatment as the high-κ gate stack. Even after correcting for the varying N_{it}, the mobilities do not come together. Comparing Figure 4.10 with 4.9, we question whether the enhancement in mobility is completely an interface related phenomenon. In order to answer that question, we make the following assumption: Since both the high-κ and SiON have identical interfaces (SiON interface for the high-κ stack), we assume that

whatever physical change occurs at the interface of the high-κ under high temperature treatment, also happens at the interface of the thicker SiON stack. This is schematically described in Figure 4.11.

Figure 4.11. The interfaces between the high-κ and bulk-Si, and the SiON and bulk-Si are assumed to be identical. And they are assumed to undergo the same physical change under high temperature treatment.

So, we seek additional scattering mechanisms that may be exclusively high-κ induced. In order to accomplish that, we compute an additional scattering factor, or a mobility component (μ_{Add}) given by equation (4.5),

$$\mathbf{1/\mu_{Add} = 1/\mu_{High\text{-}K} - 1/\mu_{SiON}} \qquad \text{……..(4.5)}$$

μ_{Add} for the as deposited and the high temperature anneal stack are plotted in Figure 4.12. It is instructive to note that μ_{Add} for the stacks are completely different. Qualitatively, a parallel shift is observed between the as deposited stack and the high temperature

annealed stack additional mobility component. Clearly, this is an exclusive response of the high-κ part of the stack. Given the shape of the curve, and the intermediate field region where the mobility component is estimated, it is safe to conclude that the phonon spectrum of the two stacks namely the as deposited stack and the high temperature annealed stack are different. This may be induced by structural change in the stack under high temperature treatment as pointed out in [2].

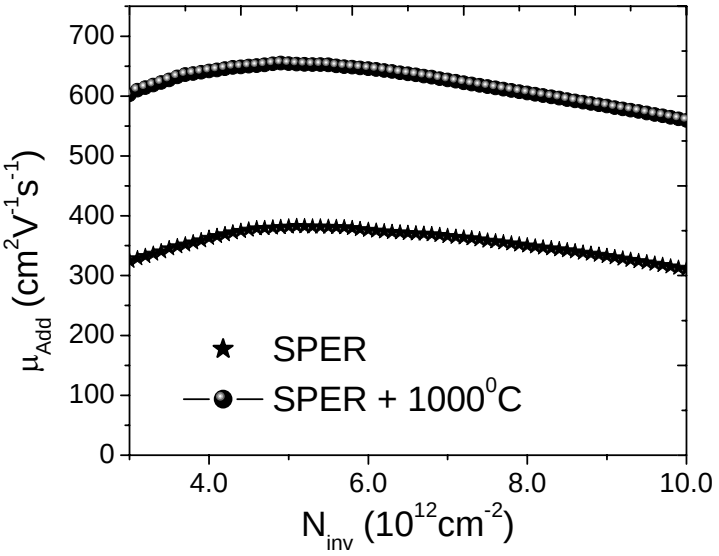


Figure 4.12. μ_{Add} for the as deposited stack and the as deposited + 1000 ^{0}C stack. An almost parallel shift is observed between the two. Qualitatively the shape of the curve is suggestive of different phonon response on the mobility in between the two stacks.

4.7. Role of Nitrogen induced fixed charge at the interface

Based on the foregoing analysis, the role of nitrogen at the interface becomes extremely important. It has been pointed out earlier in literature [8,9] that nitrogen induced fixed oxide charge in the context of Coulomb scattering limited mobility can be treated in a similar fashion as the interface states (N_{it}). Following up on this hypothesis, we extend

the already developed N_{it} calibration formulae to nitrogen induced fixed oxide charge induced scattering in high-κ gate stacks. We describe below the hardware used and the methodology of studying the nitrogen induced fixed oxide charge scattering below.

Poly-Si-gated nFETs (activation anneal 1000°C, 5 s), implementing SiON (optical thickness 26 Å, $8x10^{14}$ N/cm^2) or high-k gate dielectrics are fabricated [39]. The processing details from [39] are given here. Si(100) is first H-terminated by an RCA/HF-last clean. Thin Si_3N_4 is formed by annealing in NH_3 at 650°C. Optionally, O is then introduced into the nitride film by annealing in NO at 700 or 800°C. For comparison, interfacial SiON layers are formed thermally (11 Å, $7x10^{14}$ N/cm^2). HfSiO (~20-30 Å) with low Si content is then deposited by MOCVD, and optionally subjected to plasma nitridation. Blanket films serve for thickness measurement by ellipsometry and for composition/profile analysis by nuclear reaction analysis (NRA; Si(O)N) or medium-energy ion scattering (MEIS; HfSiON).

NO treatment increases the optical thickness of an RTNH_3 layer (no NO: 9.5 Å; 700°C: 12.5 Å; 800°C: 14.5 Å), while N areal density is nearly unaffected (~$2x10^{15}$ cm^{-2}). NO can thus be employed for oxidation of the nitride initially formed, providing independent control of N and O content.

After HfSiO deposition onto such NH_3-NO interface layers, and onto 11 Å SiON control substrates, total optical thickness (assuming refractive index n = 2.0, approximately correct for Si_3N_4 and HfSiO) increases according to NH_3 < 11 Å SiON < NH_3-

NO(700°C) < NH_3-NO(800°C) (Figure 4.13). Electrical thickness in inversion (T_{inv}, Fig. 4.13) is highest with the 11 Å SiON interface, due to the much lower dielectric constant of low N SiON.

nFET electron mobility with interfacial Si_3N_4, as measured using the split C-V technique, is degraded by 20-25% compared to low N content interfacial layers (Figure 4.14, black symbols). Upon introducing O into the interface by NO anneal, mobility recovers only marginally, resulting in an unfavorable T_{inv}-mobility tradeoff (Figure 4.14).

Four scenarios, or a combination of those, may explain the observed N-induced mobility trends: (a) Charge trapping induces hysteresis which distorts the inversion charge and mobility measurement; or (b) trapped charges, (c) slow interface states (areal density N_{it}), or (d) fixed charges cause Coulomb scattering of channel electrons.

To address scenarios (a) and (b), we have measured dual-sweep nFET current-voltage (split C-V) characteristics to high positive stress voltage, to highlight potential instabilities (-0.5 $\leftrightarrow$ +2.2 V). 30 Å HfSiO/Si_3N_4 stacks exhibit hysteresis (Figure 4.15), confirming that charge trapping occurs. Reduced hysteresis upon introduction of O atoms likely is mostly due to gate leakage reduction, since charge pumping measurements (not shown) demonstrate that a similar density of bulk traps is populated with all interfaces. However, hysteresis is negligible for 20 Å HfSiO/Si_3N_4 stacks (Figure 4.15). Also, when reducing stress to +1.6 V (corresponding to $N_{inv} \sim 10^{13}$ cm^{-2} where high-field mobility is

measured), hysteresis is also small for 30 Å HfSiO stacks (not shown). Charge trapping can therefore not account for the observed mobility trends.

To address scenario (c), we measure N_{it} by amplitude-sweep charge pumping (Figure 4.16). Clearly, NH_3-NO interfaces exhibit much higher N_{it} than control SiON interfaces. N_{it} corrections to the mobility were made using Matthiessen's rule in conjunction with an independent calibration with thermally generated Si dangling bonds at SiO_2/Si interfaces, as described earlier in Section 4.3. Clearly, even after eliminating the N_{it} impact on mobility, the mobility trends are unchanged, demonstrating that other mechanisms are responsible for mobility loss.

By contrast, scenario (d) is supported by C-V measurements: V_t extracted from the split C-V up-sweep (Figure 4.16) is ~0.1 V lower with the Si_3N_4 interface than with the control SiON interface, and no significant recovery occurs upon O introduction. N areal density thus determines the areal density of positive fixed charge (upper 10^{11} cm^{-2} range, consistent with [40]), largely independent of the chemical environment (i.e. bonding to N vs. O).

In conclusion, nitride interfaces in poly-Si/high-k stacks degrade mobility primarily by Coulomb scattering at fixed charges. Areal density of such defects is only marginally reduced by introduction of O, resulting in an unfavorable mobility-T_{inv} tradeoff. A slight mobility increase observed upon O incorporation is consistent with fixed charges spatially moving away from the channel, reducing the Coulomb scattering cross section.

Figure 4.17 (a) shows MEIS spectra recorded from identical 30 Å HfSiO films after nitridation at different temperatures (250 vs. 400°C) and plasma pressures. Nitrogen concentrations (N/(N+O)) in the gate stack of up to ~21% are detected. Employing low nitridation temperature and plasma pressure, the N signal (~82 keV) is centered at the highest ion energy, indicating N near the top surface of the HfSiO film.

nFET (and pFET) gate stacks incorporating such HfSiON show N-induced T_{inv} reduction by up to 1 Å. Despite the high N content, V_{fb} and V_t are shifted to more negative values by only 0.01-0.02 V, indicating that only little positive fixed charge is created near the channel. Trap density also remains low, as indicated by low C-V hysteresis (<0.01 V during sweeps to ±2.2 V). Mobility loss with appropriate HfSiO nitridation is minimal, with peak and high-field nFET mobility for HfSiON at 90-100% of an SiON control (Figure 4.17 (b)).

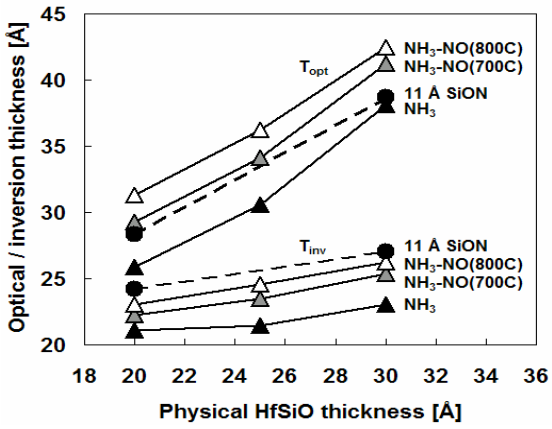


Figure 4.13. Total optical thickness (assuming n = 2) and nFET T_{inv} for 20-30 Å HfSiO deposited onto Si(O)N layers [39]

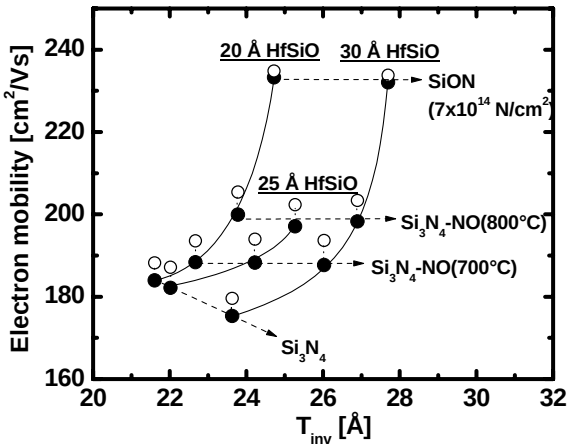


Figure 4.14. nFET electron mobility at high field ($N_{inv} = 10^{13}$ cm^{-2}) as a function of T_{inv} for various Si(O)N processes and HfSiO thicknesses. Black symbols: as measured. White symbols: after N_{it}-correction (see section 4.3).

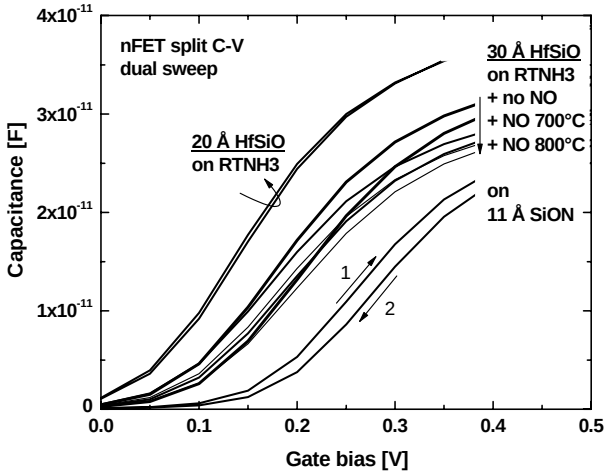


Figure 4.15. Dual-sweep (-0.5 ↔ +2.2 V) nFET split C-V characteristics with HfSiO/Si(O)N gate dielectrics [39]

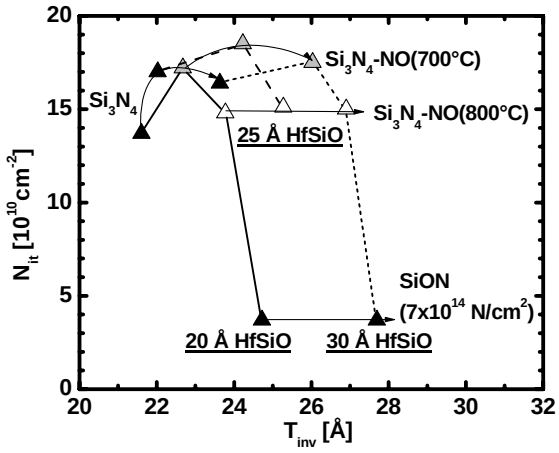


Figure 4.16. N_{it} as a function of T_{inv} for various Si(O)N processes and HfSiO thicknesses, as measured by amplitude-sweep charge pumping.

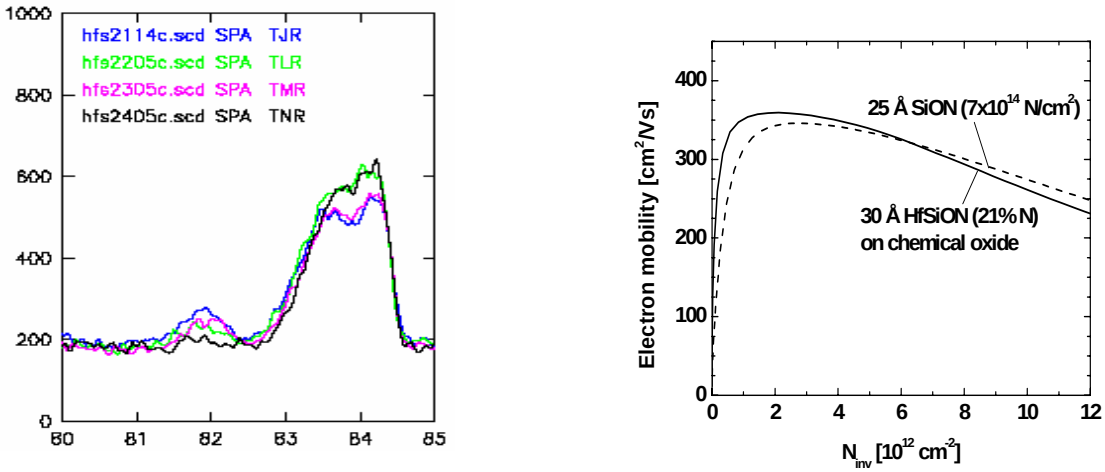


Figure 4.17 (a) MEIS O and N signals for HfSiO nitrided at various temperatures and plasma conditions. (b) nFET electron mobility with 30 Å HfSiON containing 21% N introduced by a low-pressure/low-temperature plasma, compared to low N content SiON [39].

To address the role of fixed oxide charge scattering due to nitrogen at the interface, we consider exclusively the 3 nm HfSiO stack and take a closer look at the mobility values in Figure 4.14. We estimate the oxide charges from CV shift in Figure 4.15, which is found to be on the order of 8-8.5 X 10^{11} cm^{-2} . Using this value, we apply the same correction as described in section 4.3 further. As evident in Figure 4.18, we see that the mobility corresponding to $RTNH_3$ interface comes to the same level as the SiON interface thereby confirming the following two points:

(i) As pointed out in early SiO_2 literature [8,9,10], nitrogen induced fixed oxide charge scattering may be treated in the same way theoretically as interface state induced scattering in the bulk-Si channel.

(ii) The nitrogen induced fixed oxide charge may be very close to the channel given its strength of coupling with the channel electrons.

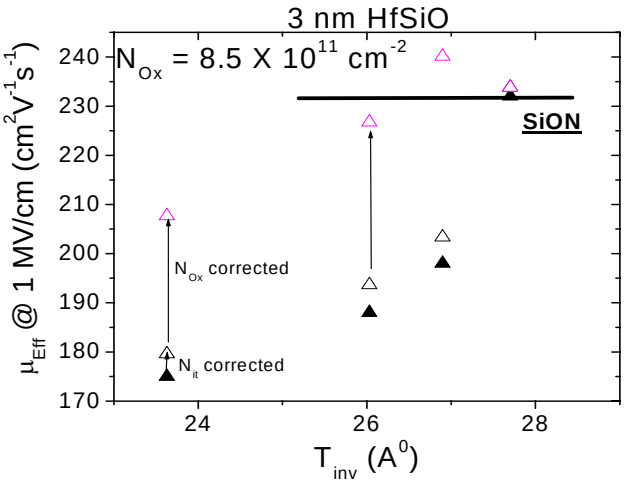


Figure 4.18. N_{Ox} induced fixed oxide charge may be corrected in the same fashion as N_{it} induced scattering.

4.8 Conclusions

In the foregoing chapter, we developed a technique of exclusively studying the impact of N_{it} on mobility in the channel of bulk-Si NMOSFETs with advanced high-κ gate stack. It is also conclusively shown that with an aggressively scaled metal gate high-κ gate stack, the high mobility attained after high temperature treatment, cannot be accounted for by difference in N_{it} alone. By a novel technique, the response of the nitrided interfacial layer is convincingly excluded, to study the influence of the high-κ portion of the stack before and after high temperature anneal. It is shown that the phonon spectrum before and after the high temperature treatment are completely different, thereby confirming partially the earlier claim in literature about high temperature process step bringing about structural change in the stack [2]. Also the role of nitrogen induced charge trapping was systematically studied on a controlled set of hardware and its impact on mobility thoroughly investigated.

In the next chapter, we take up the role of phonons (particularly the role of soft optical phonon in high-κ insulators) and study their impact on mobility in the channel.

References

[1]. V. Narayanan, K. Maitra et. al., "Process optimization for high electron mobility in aggressively scaled metal gate high-κ gate stacks", 2005 (submitted to IEEE Electron Device Letters).

[2]. A. Callegari et. al. "Interface Engineering for Enhanced electron mobilities in W/HfO_2 gate stacks", in IEDM Technology Digest 2004.

[3] John F. Conley, "IRPS Tutorial Lecture", IRPS 2004.

[4]. Syed Aon Mujtaba, "Advanced mobility models for design and simulation of deep sub micrometer MOSFETs", 1995, PhD dissertation, Stanford University.

[5]. H. Brooks, "Scattering by ionized impurities in semiconductors", 1951, Phys. Rev., Vol. 83, pp. 879-880.

[6] E. Conwell, and Weisskoff, "Theory of impurity scattering in semiconductors", 1950 Phys. Rev., Vol. 77, No. 3, pp. 388-390.

[7] T. Ando , A. B Fowler, and F. Stern, " Electronic properties of two dimensional systems", 1982, Rev. of Modern Phys., Vol. 54, No. 2, pp. 437-672.

[8] C. T Sah, T. H Ning, and L. L Schopp, " The scattering of electrons by surface oxide

charges and by lattice vibrations at the Si/SiO_2 interface", 1972, Surface Science, Vol. 32, pp. 561-575.

[9] T. H Ning, and C. T Sah, "Theory of scattering of electrons in a non-degenerate semi-conductor surface inversion layer by surface oxide charges", 1972, Phys. Rev. B, Vol. 6, No. 12, pp. 4605-4613.

[10] A. Gold, "Electronic transport properties of a two dimensional electron gas in a Si Quantum well structure", 1987, Physical Rev. B, Vol. 35, No. 2, pp. 723-733.

[11] J. Lee, H. M Spector, and V. K Arora, "Impurity scattering limited mobility in a Quantum well heterostructure", 1983, Journal of Applied Physics, Vol. 54, No. 2 , pp. 6995-7004.

[12] D. Monroe, Y. H Xie, E. A Fitzerald, P. J Silverman, and G. P Watson, "Comparison of mobility limiting mechanisms in high mobility $Si_{1-x}Ge_x$ heterostructures" , 1993, Journal of Vac. Sci. Technol., Vol. 11, No. 4, pp. 1731-1737.

[13] K. Hess, "Impurity and phonon scattering in layered structures", 1979, Appl. Phys. Lett., Vol. 35, No. 7, pp. 484-486.

[14] A. Gold, "Scattering time and single particle relaxation time in a disordered two-dimensional electron gas", 1988, Phys. Rev. B, Vol. 38, No. 15, pp. 10798-10811.

[15] F. Gamiz, J. A Lopez Villaneuva, J. A Jimenez Tezada, I. Melchor, A. Palma, "A comprehensive model for Coulomb scattering in inversion layers" , 1994, Vol. 75, No. 2, pp. 924-934.

[16] S. A Mujtaba, S. Takagi, and R. W Dutton, " Accurate Modeling of Coulombic scattering, and its impact on scaled MOSFETs", 1995, VLSI Technology Symp. , pp. 99-100.

[17] R. Stalbusch, and E. A Cartier, "Interface defect formation in MOSFET by atomic Hydrogen exposure", 1994, IEEE Trans. Nuc. Sci., Vol. 41, No. 6, pp. 1844-1853.

[18] K. L Brower, "Dissociation Kinetics of hydrogen passivated (111) Si-SiO_2 interface 1990, Phys. Rev. B, Vol. 42, pp. 3444-3453.

[19] J. H Stathis, "Dissociation Kinectics of hydrogen passivated (100) Si-SiO_2 interface Defects", 1995, J. Appl. Phys., Vol. 77, No. 12, pp. 6205-6207.

[20] S. Manzini, "Effect of Coulomb scattering in n-type Si inversion layers", 1985, Journal of Appl. Phys., Vol. 57, No. 2, pp. 411-414.

[21] S. Takagi, M. Iwase, and A. Toriumi, 1988, IEDM Tech. Digest, 398.

[22] A. T Krishnan et. al., 2003, IEDM Tech. Digest, 14.5.1.

[23] A. Kang, P. Lenahan, and J. Conley, "The radiation response of high dielectric constant HfO_2/Si system", 2002, IEEE Trans. On Nuc. Sci., Vol. 49, No. 6, pp. 2636-2642.

[24] J. W McPherson, and H. C Mogul, "Impact of mixing of disturbed bonding states on TDDB on SiO_2 thin films", 1997, Appl. Phys. Lett., Vol. 71, pp. 3721-3723.

[25] T. R. Oldham, and A. J Lelis, 1996, MRS Bulletin, 248.

[26] J. Koga, T. Isihara, and S. Takagi, 2003, IEEE Elec. Dev. Lett. , Vol. 24, pp. 354-357.

[27] J. Koga, T. Isihara, and S. Takagi, 2004, Jpn. J. Appl. Phys., Vol. 43, pp. 1699-1703.

[28] D. K Schroder, "Semiconductor Material and Device Characterization", John Wiley and sons, second edition, 1998.

[29] Stefano Villa, Andrea L. Lacaita, Laura M. Perron, and Roberto Bez, "A Physically Based Model of the Effective Mobility in Heavily doped NMOSFETs", IEEE Transactions on Electron Devices, 1998, Vol. 45, No. 1, pp. 110 – 115.

[30] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the universality of inversion layer mobility in Si MOSFETs: Part I – Effects of substrate impurity concentration", (and references therein), 1994, IEEE Trans. On Elec. Dev., Vol. 41, No. 12, pp. 2357-2362.

[31] V. V Afanesev, and A. Stesmans, J. Appl. Phys., 2004, Vol. 95, p. 2518.

[32] V. V Afanesev, and A. Stesmans, Appl. Phys. Lett., 2003, Vol. 82, p. 4074.

[33] V. V Afanesev, and A. Stesmans, Appl. Phys. Lett., 2002, Vol. 80, p. 1957.

[34] A. Kerber et. al., "Charging instability in n-channel MOSFETs with SiO_2/HfO_2 Gate dielectric", IEEE SISC, 2002, San Diego.

[35] E. Cartier, Topic 134, IRPS Tutorials, IRPS 2004

[36] A. Kerber, "Methodology for electrical characterization of MOS devices with alternative gate dielectrics", 2003, PhD Dissertation, TU Darmstadt and IMEC Belgium.

[37] Frank Stern, "Calculated temperature dependence of mobility in Si inversion layers", 1980, Physical Rev. Lett., Vol. 44, No. 22, pp. 1469-1471.

[38] K. M Cham, and R. G Wheeler, "Temperature dependent resistivities in Si inversion layers at low temperatures", 1980, Physical Rev. Lett., Vol. 44, No. 22, pp. 1471-1473.

[39] Martin Frank, K. Maitra et. al., "Nitrogen in poly-Si/Hf-silicate gate stacks- Carrier mobility impact of traps and fixed charge" (Accepted IEEE SISC 2005)

[40] X. Shi, M. Shriver, Z. Zhang, T. Higman, and S. A Campbell, 2004, Journal of Vac. Sci. Technology, A 22, p. 1146.

Chapter 5

Aggressively Scaled High Mobility Metal Gate High-κ NMOSFETs – Role of Phonon Scattering

5.1 Introduction

So far in this thesis, we have thoroughly investigated the role of charge trapping and its impact on mobility in the channel of bulk-Si NMOSFETs. We also pointed out in Chapter 1 that high-κ gate stacks have an intrinsic limitation which could potentially limit mobility attained in high-κ insulator based NMOSFETs [1]. In this chapter, we describe a methodology of determining phonon component of mobility (surface phonon or soft optical phonon limited mobility) accurately, and apply it to a cross-section of high-κ gate stacks (both Hf-silicate and HfO_2) with both poly-Si and metal gates, and compare the phonon components of the aforementioned stacks. We also revisit the Al_2O_3 capped Hf-silicate stacks we discussed in Chapter 3 and show how intermixing and stack chemistry is expected to alter the phonon spectra of these stacks. The high mobility, aggressively scaled metal gate high-κ gate stacks are described next including processing details. It is emphasized that the choice of metal gates is a critical factor in device performance. Two factors guide the choice of metal gates: (i) setting the right V_t, or threshold voltage of the transistor, (ii) impact of the metal gate on transport in the channel. In order to address (i), we carried out a simulation experiment to understand the feasibility of midgap workfunction metal gates on the performance of 25 nm bulk MOSFETs. In order to understand (ii), we do careful physics inspired modeling and novel ultra low temperature measurements to understand the beneficial impact metal gates have on electron mobility in the bulk-Si NMOSFET. The NMOSFETs chosen were

ultra thin EOT (T_{inv}) high mobility, aggressively scaled metal gate HfO_2 transistors with mobility better than that reported in [2]. We end the chapter by investigating the impact of interfacial oxide layer thickness on phonon limited mobility, and also making projections about the best mobility attainable in high-κ based bulk Si NMOSFETs.

5.2 Accurate extraction of surface phonon component of mobility

Based on the methodology described in Chapter 2, we can now accurately model the Coulomb component of mobility, both the response of the dopants in the channel and the interface states or N_{it}. Following [3], we can factor out the surface roughness component of mobility in high-κ gate stacks by using the same treatment as in SiO_2 based NMOSFETs. Thus by using Matthiessen's rule, we can estimate the phonon component of mobility. Now, for SiO_2 based devices, the phonon component is dominated by one component, namely, the bulk phonon response of the Si channel. However, for the high-κ case, the phonon component consists of two parts, namely the bulk phonon component, and the surface phonon component, the second part being an exclusive high-κ response. It is easy to factor out the bulk phonon response by learning from the SiO_2 phonon response, and the remaining portion of the high-κ mobility is the surface phonon response. Table 5.1 shows some high-κ stacks with different metal gates and T_{inv}'s with the processing details described elsewhere. Figure 5.1 shows a comparison of the surface phonon limited mobility for the stacks described in Table 5.1.

#	Gate Stack	T_{inv} (nm)	N_{it} (cm^{-2})
I	HfO_2 + FUSI [4]	2.2	2.36×10^{10}
II	HfSiO + FUSI [4]	2.0	3.73×10^{10}
III	HfSiO + Poly [Ch. 3]	2.6	6.17×10^{10}
IV	HfSiO + Al_2O_3 cap + Poly [Ch. 3]	3.2	7.14×10^{10}
V	HfO_2 + FUSI [4]	2.2	1.00×10^{12}
VI	HfO_2 + W [5]	1.65	1.2×10^{11}
VII	HfO_2 + W [5]	1.95	2.7×10^{10}

Table 5.1. Different high-κ gate stacks with poly-Si and metal gates used in determining the surface phonon components of mobility shown in Figure 5.1. The processing details of the stacks may be found in the references marked in parenthesis. (Ch. 3 refers to Chapter 3 and relates to the Al_2O_3 capped Hf-silicate stacks described there)

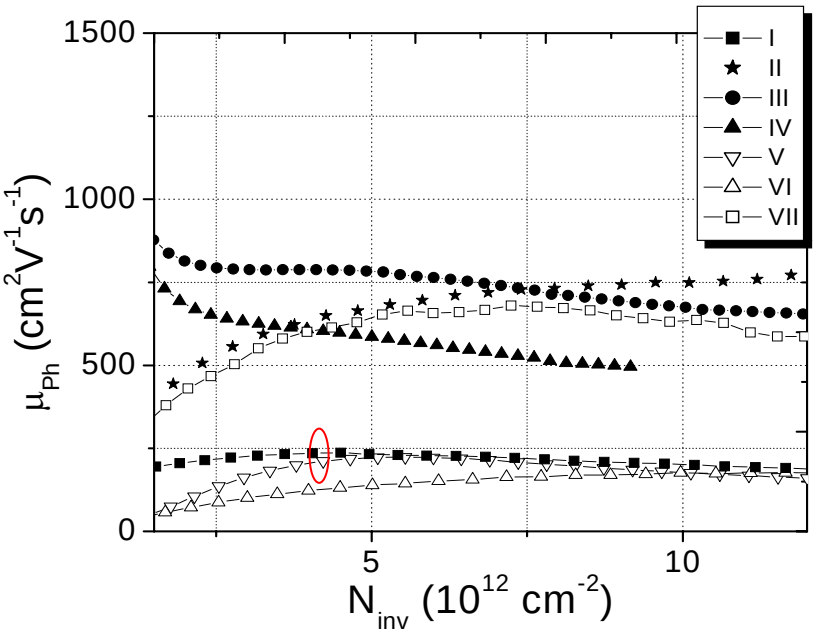


Figure 5.1. Surface phonon limited mobility for different stacks described in Table 5.1 (Stacks I and V have been clubbed together and colored for emphasis)

Three important points from the foregoing analysis may be distinguished:

(i) Stacks I and V are the same stack with different N_{it}, achieved by partial passivation under different temperature H_2/N_2 (FG-forming gas) anneal. After correcting for the N_{it}, it is seen that these two mobility come together thereby affirming the accuracy of the N_{it} calibration scheme developed in Chapter 4.

(ii) Stacks III and IV are Hf-silicate stacks with and without 20 cycles Al_2O_3 capping layer described earlier in Chapter 3. It is instructive to note that the

phonon components of the two stacks are different. This may be because of Al updiffusion observed from MEIS (medium energy ion scattering) data depicted in [6]. It is conjectured that Al up diffusion alters the stack chemistry and subsequently the phonon response of the stack, keeping in mind the fact that Al updiffusion can subtly change the dielectric response of the high-κ insulator without really affecting the charge trapping significantly (Chapter 3).

(iii) Phonon response VI and VII are of stacks from [5] described in Chapter 1 (Figure 1.10). Response VI is of the as deposited stack, and VII is the same for the high temperature annealed stack. Their N_{it} response has been factored out by the treatment already developed in Chapter 4. It is clearly seen even with these stacks, the phonon response between the as deposited stack and high temperature annealed stack is completely different.

In the following section, we describe process optimization for high mobility in aggressively scaled metal gate/high κ gate stacks. Attainment of high mobility has been attributed to high temperature process step as described in Chapter 4. We revisit the issue of possible structural change altering the phonon response of the high-κ insulator gate stack here.

5.3 High Mobility, aggressively scaled metal gate high-κ gate stacks- Process optimization and comparison of mobility in different metal gate stacks

Aggressively scaled metal-gated high-k stacks have been shown to exhibit significantly reduced electron mobility, thereby limiting device performance [7,8]. However, recent

reports have demonstrated high mobility at low T_{inv} [2, 9]. The improved mobility was attributed to the screening of the soft optical phonons in the HfO_2 gate dielectric by the metal electrode [2] and/or to an optimum thickness of the HfO_2 dielectric which minimized bulk trapping under electrical stress thereby resulting in high electron mobility [9]. It was previously shown with HfO_2/CVD W gates that even with low N_{it} (Figure 5.1, stack VII), **low T processing** resulted in low electron mobilities. Increasing the thermal budget resulted in significantly improved mobilities albeit, at the expense of T_{inv}. [5]. To understand the effect of nitrogen on different HfO_2/Metal Stacks, we have compared nFET-like low workfunction HfO_2/CVD TaSiN with HfO_2/CVD W gate stacks, with both nitrided and nitrogen free interface layers, and it was shown that the presence of N at the interface, while it can provide T_{inv} scaling, *clearly degrades the electron mobility* even after high T processing [10]. Indeed, nitrogen free gate stacks have been shown to exhibit high mobilities in aggressively scaled replacement gate devices [11]. Here, we report a self-aligned high temperature process flow for various metal electrodes on HfO_2 gate dielectrics using non-nitrogen starting interfaces and show that high electron mobilities in aggressively scaled High k/Metal Gated Stacks can exceed those of scaled poly-Si/SiON stacks [12]. This is achieved by using high thermal budgets without allowing for interfacial layer (IL) regrowth.

HfO_2 gate dielectrics were deposited by CVD processes on non-nitrogen starting interfaces (unless otherwise specified). TiN and TaSiN gates were then deposited by reactive sputtering (PVD). W and some TaSiN films were deposited by CVD [5, 10, 12]. TaN films were deposited by ALD. Self aligned nFETs were fabricated by capping thin

metal layers with poly-Si. Most stacks were subjected to a 1000°C (5s) S/D activation RTA (unless otherwise specified). Final passivation anneals in FGA were optimized to minimize N_{it} and N_{ox}. Solid Phase Epitaxial Regrowth (SPER) [13] - using high energy Arsenic for S/D amorphization followed by a 600°C (10 min.) crystallization and various activation anneals (none, 800°C, 5s & 1000°C, 5s RTA) - was used to evaluate the impact of low-T activation on mobility (see Chapter 1, section 1.7). For the SPER process we used PVD TaSiN gates and NiSi contacts for S/D and gate. For mobility measurements, the split-CV technique (100 kHz) was used on 10x20 μm^2 large area FETs. Charge Pumping (CP, amplitude sweep mode at 1 MHz/10 kHz) and conductance measurements were used to evaluate the interface state density (N_{it}). N_{ox} was derived from V_{fb}/V_t shifts. N_{it} corrections to the mobility were made using Mathiessen's rule using an independent calibration with thermally generated Si dangling bonds on SiO_2 as described in Chapter 4. As pointed out in the preceding Chapter, since the Si dangling bond is the dominant interface defect for $Si/SiO_2/HfO_2$ gate stacks, N_{it} corrections as derived from SiO_2 experiments and using Mathiessen's rule, can be directly applied to correct for the N_{it} induced mobility loss.

Figure 5.2 shows inversion and accumulation side split-CV curves for self-aligned nFETs (with CVD W, PVD TiN, ALD TaN, CVD TaSiN and PVD TaSiN electrodes and HfO_2 gate dielectrics after S/D activation anneals 1000 ^{0}C, 5s). The large T_{Ox} and T_{inv} with W gates are attributed to IL regrowth from oxygen dissolved within the W, also potentially contributing to mobility improvement [5]. The regrowth could not be prevented even with appropriate barrier like Poly-Si capping layers. TiN, TaN and TaSiN show

significantly smaller T_{inv} values even after high temperature processing, clearly showing that with Poly-Si capping, regrowth can be suppressed and low T_{inv} can be obtained with some metal gates. The electron mobility curves corresponding to the devices in Figure 5.2 are shown in Figure 5.3 demonstrating that record mobilities with TiN gates (Fig. 5.3, $\mu_{1MV/cm}$ = 215 cm^2/Vs; 1MV/cm corresponds to $N_{inv} = 1.1x10^{13}cm^{-2}$ for $N_a = 1x10^{17}$ cm^3, which are better than those reported previously at a T_{inv} = 1.4 nm [2,9]) were achieved.

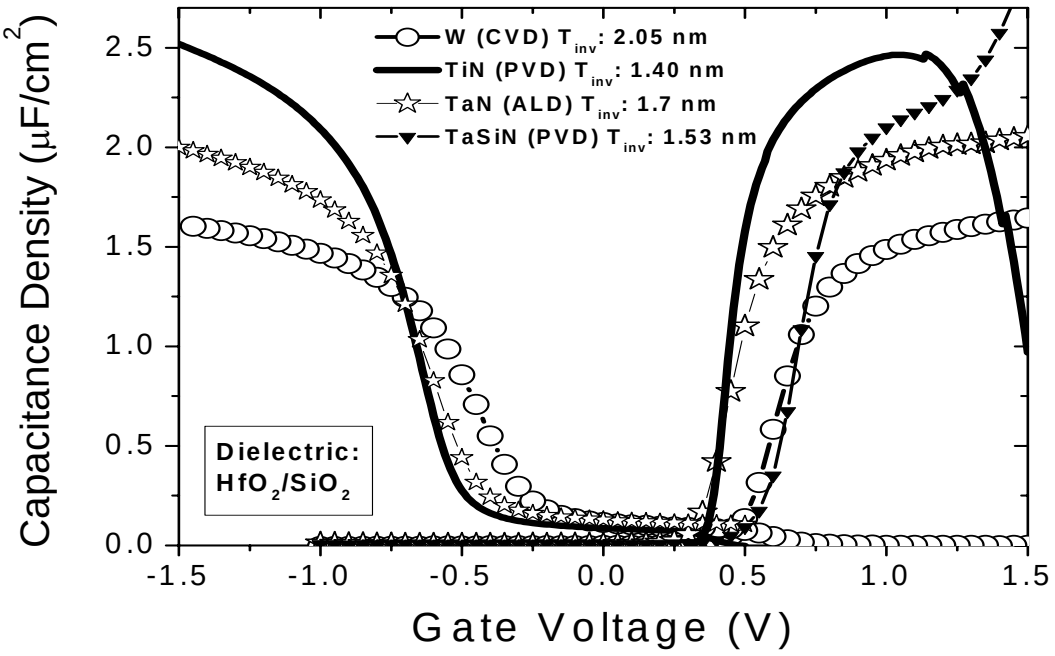


Figure 5.2. Inversion and accumulation side split-CV characteristics of self-aligned nFETs with W, TiN and TaN & TaSiN gates. (1000^0C, 5 s)

It is quite clear that the deposition technique can have a profound impact on the mobility as observed when comparing CVD vs. PVD TaSiN (33.3 % Ta, Si and N each) gates (see Figure 5.3). For similar T_{inv} and interface layers the CVD TaSiN process

exhibit higher N_{it} ($1.5x10^{11}$ cm^{-2}) and bulk traps (Figure 5.4) compared to any of the other aggressively scaled metal gate/high k stacks (N_{it} $<3x10^{10}$ cm^{-2}). However, for the CVD TaSiN/HfO_2 with $N_{it} \sim 1.5x10^{11}$ cm^2, the mobility correction is very small (inset in Figure 5.5). It is also shown that bulk charge in the high-k layer has a smaller impact on mobility than N_{it} (as pointed out earlier in Chapter 3). Additionally, the 10 KHz CP curve for CVD TaSiN clearly shows an uptake at an earlier gate voltage (unlike TiN with similar T_{inv}) suggesting that these charges are close to the Si surface, possibly at the SiO_2/HfO_2 interface and not in the bulk of the HfO_2 films. These charges may be attributed to the SiH_4-based electrode process which damages the HfO_2 gate dielectrics similar to HfO_2/Poly gate stacks and/or additionally to incorporation of N into the dielectric stack [10]). In this context, it is important to emphasize that for very low N_{it} samples, it is difficult to carry out charge pump measurements to estimate the interface state density. As an alternative, a calibrated conductance technique, described in [14] was used to get a good first guess of the N_{it}, for instance in the aggressively scaled TiN stacks (as pointed out towards the end of Chapter 2).

To understand the effect of high temperature processing on the mobility of aggressively scaled devices, we have designed an experiment wherein IL regrowth is minimized during thermal processing and thereby IL thickness is decoupled from the observed mobility improvement. The experiment entails using PVD TaSiN to minimize interfacial regrowth during high temperature annealing, SPER processes for low T activation, and SiON, SiON/HfO_2 gate dielectrics to compare oxynitride and high k gate stacks. Figure 5.3 shows that substantial mobility improvement (25%, peak) for ***both***

SiON/TaSiN and SiON/HfO_2/TaSiN stacks is achieved ***only after 1000°C anneals*** with little to no change in T_{inv} or V_t (Chapter 4, Figure 4.1). This conclusion is not tainted by N_{it} variations, as the mobility curves are corrected for N_{it} (as has already been described in detail in Chapter 4). It is to be noted that annealing of sputter damage at high temperatures could also contribute to some of the observed improvement. The importance of high temperature processing is further corroborated with SiO_2/ HfO_2/TiN stacks (not shown) where, by changing the S/D anneal from a 900 °C (5s) RTA to a 1085 °C spike anneal, a 10 % improvement in high field mobility (@ 1 MV/cm) was observed. This improvement in mobility occurred without a change in N_{it}, V_t or T_{inv}. Also we have conclusively shown in Chapter 4, that improvement in mobility after high temperature treatment cannot be explained by role of the interfacial rearrangement of N alone.

All these results suggest unequivocally that the *high thermal budget process* modifies the dielectric stack *without interfacial regrowth in the presence or absence of high k* to enhance the mobility. It has been shown previously that for Si/SiO(N) /Poly gate stacks wherein the dielectric is grown at high temperatures > 850ºC, it is still necessary for the stacks to see a T > 950 °C after gate stack deposition to allow for the formation of a relaxed Si/SiO(N) interface which is required for optimal performance [15].

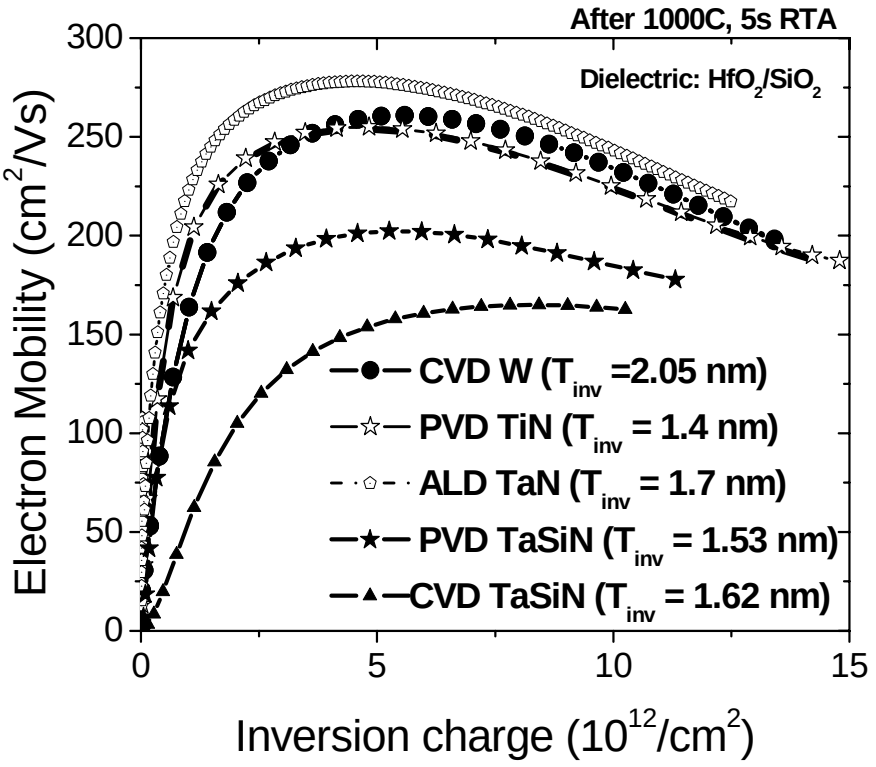


Figure 5.3. Electron mobility for stacks shown in Figure 5.2. Please note channel doping for all stacks except PVD TaSiN is $N_a = 1x10^{17}cm^{-3}$. The PVD TaSiN films had a channel doping density of $4x10^{17}cm^{-3}$. (The PVD TaSiN has been used in the context of justifying the N_{it} calibration technique in Chapter 4)

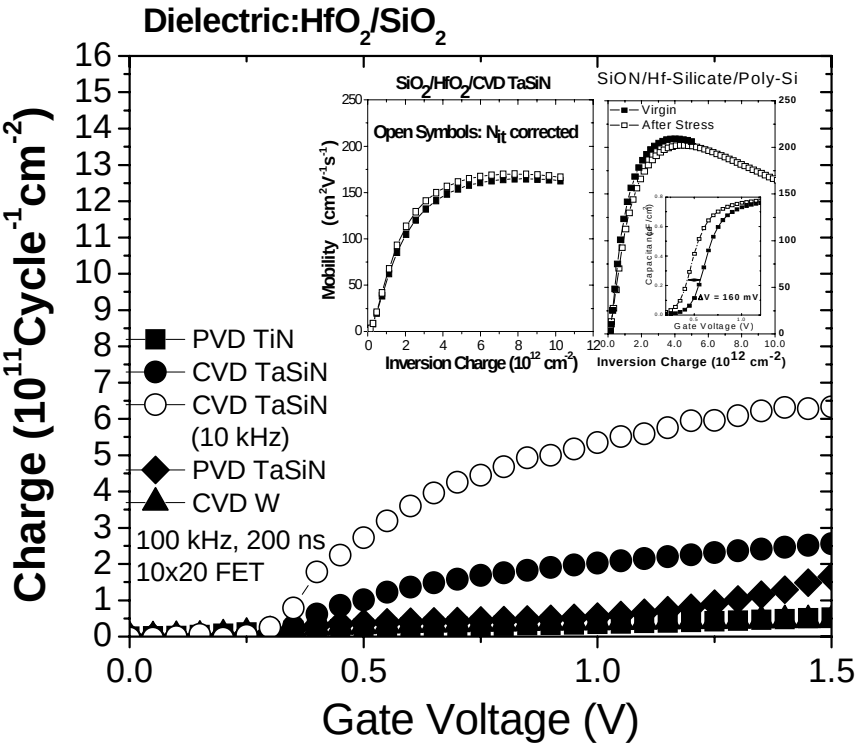


Figure 5.4. : Charge pumping (amplitude sweeps) for nFETs with W, TiN and CVD/PVD TaSiN gates (after 1000 °C (5s) RTA). The left inset shows mobility for CVD TaSiN after correcting for $N_{it} = 1.5x10^{11}$ cm^{-2}. The right insert shows the impact of 1.02 x 10^{12} cm^{-2} bulk charges (located in center of silicate layer) on mobility in a SiON/Hf-silicate/poly-Si gate stack, showing little impact of bulk charge (This has already been discussed in detail in Chapter 3, and is given here for easy reference)

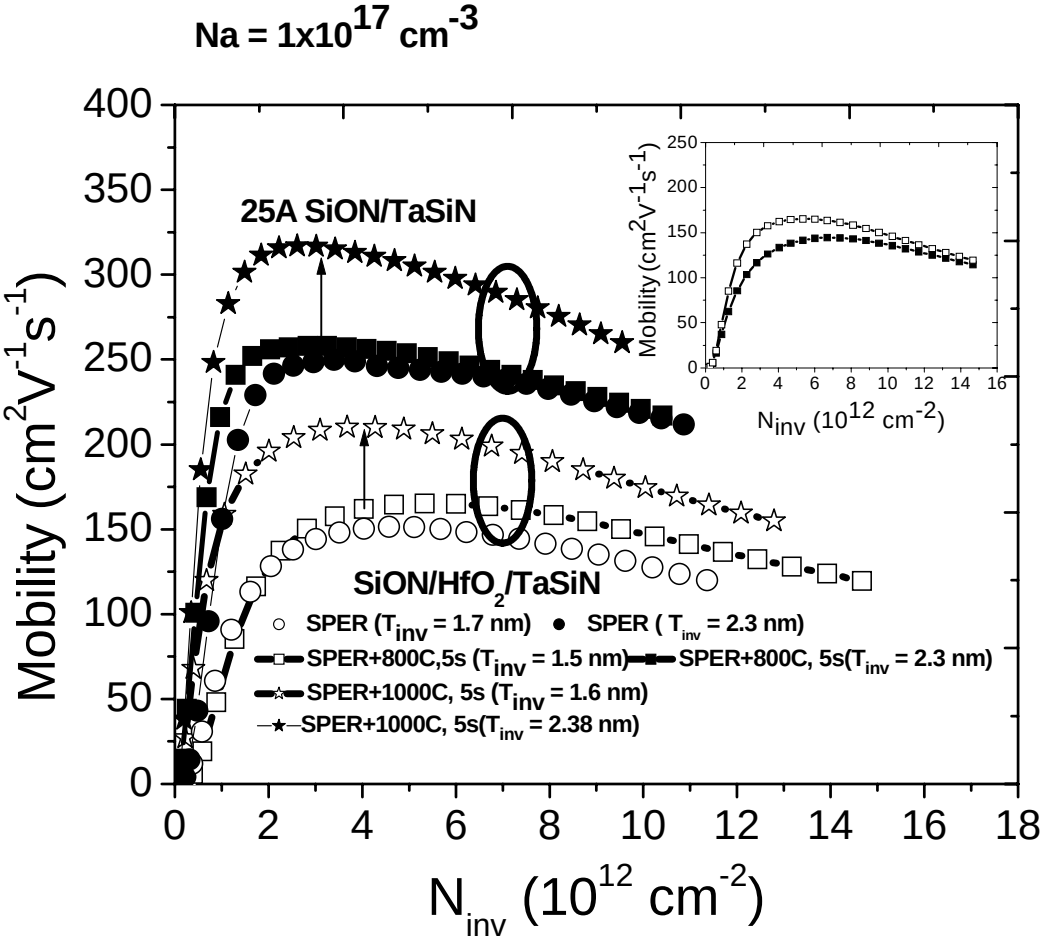


Figure 5.5. Electron mobility measurements after N_{it} corrections for $SiON/HfO_2/TaSiN$ and 2.5 nm SiON/TaSiN. Note the substantial improvement in mobility from 800°C to 1000°C which cannot be attributed to N_{it} or T_{inv} increase. The inset shows an example of the N_{it} corrected mobility (open symbols) for a $SiON/HfO_2/TaSiN$ stack with $N_{it} = 5x10^{11}$ cm^{-2}.

Now, we take a closer look at the physical nature of the structural change possibly initiated by the high temperature process step. Previously, it had been suggested that some Hf from the HfO_2 intermixes with a *non-nitrogen* interfacial layer (suggested mostly by comparison of TEM and electrical measurements) to form a higher κ Hf

silicate interface layer [5] which results in higher mobility than HfO_2 since it has a weaker coupling of the SO phonons compared to HfO_2 [1,16]. However, it is well known that Hf silicate films tend to phase-separate rather than mix together at high temperatures [17]. Additionally, most analytical techniques do not show evidence that the SiO_2 IL converts to a Hf-silicate film at the interface upon annealing [18, 19]. Thus, the true nature of the stack modification is not yet well understood, though it appears to be unrelated to IL regrowth or Hf-silicate formation.

Therefore, by careful process optimization such as the use of non-nitrogen interface layers, high temperature processing (to modify the IL and the IL/HfO_2 interface), low N_{it} ($<3x10^{10}$ cm^{-2}), and appropriate electrode and electrode structures to prevent IL regrowth, all the undesirable sources of Coulomb scattering are largely minimized yielding the high electron mobility, which is competitive with poly-Si/SiON stacks at thicker T_{inv} as illustrated in Figure 5.6. It is also clear that by this optimization [12] a different high-κ/metal gate scaling trend has been achieved , similar to Poly/SiON and significantly above the old metal/high κ scaling trend [5, 7, 8].

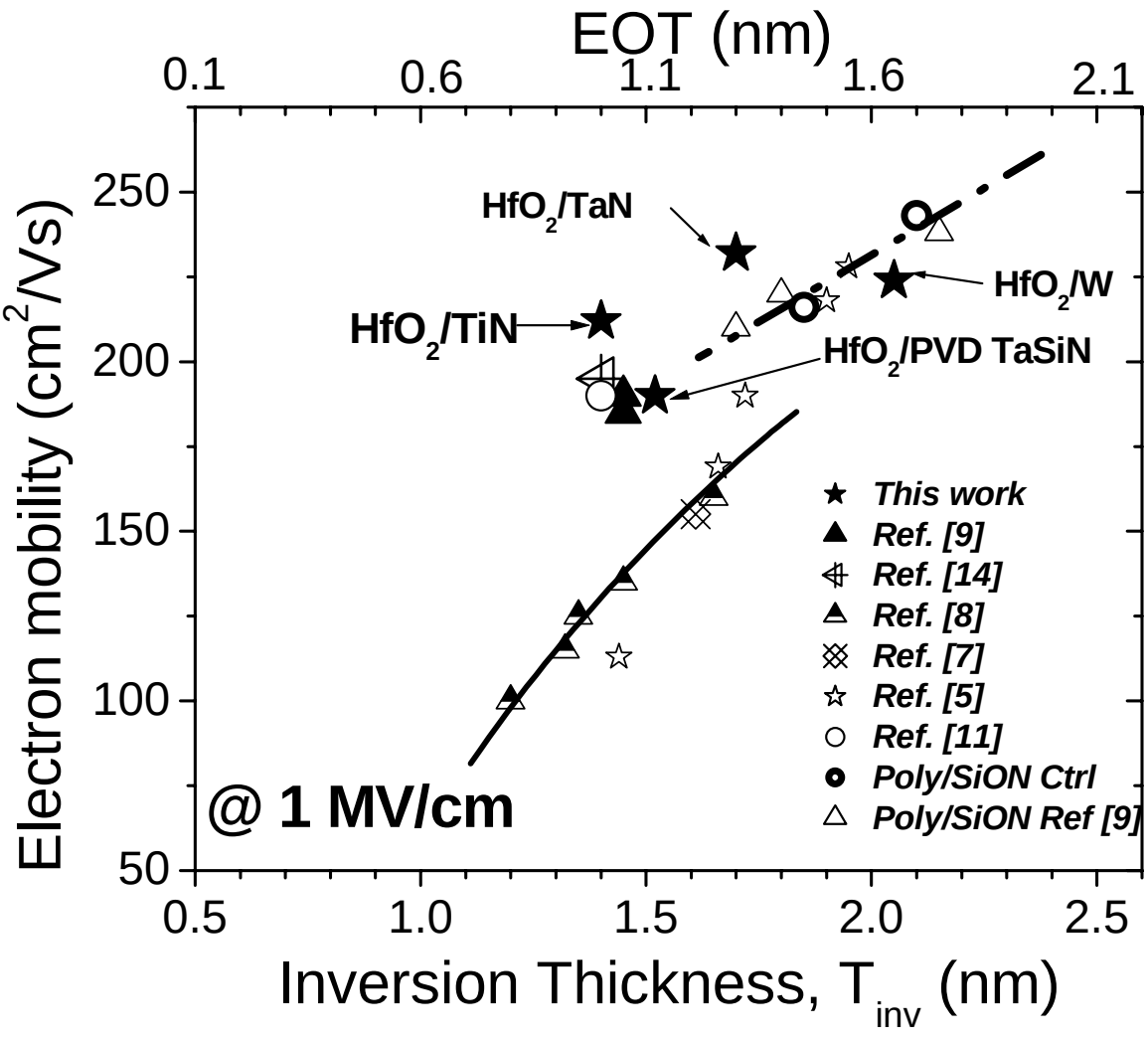


Figure 5.6. Comparison of electron mobility at 1 MV/cm versus T_{inv} for different metal-gated HfO_2 gate stacks. Mobility from this work is competitive or better than Poly/SiON and lies on a scaling trend significantly higher than was observed previously for different metal/high κ stacks.

5.4 Choice of Metal gates-Bandedge or Midgap?

In the foregoing analysis, we discussed a number of metal gates most of which has midgap workfunction (W, TiN, TaN) and an n+ metal gate (TaSiN). From a transport perspective, as pointed out in [20], bandedge metal gates are expected to yield worse mobility compared to midgap metal gates, particularly at aggressively scaled EOTs/T_{inv}'s. The former, because of the alignment of the metal gate Fermi level with the conduction band edge of Si, has higher vertical field at aggressively scaled T_{inv}'s compared to the latter. This higher vertical field leads to additional surface scattering mechanisms in the channel thereby degrading mobility compared to midgap metal gates. But the choice metal gate workfunction is not guided by mobility alone. A more pressing concern is setting the right threshold voltage (V_t). In this context, we carry out the following simulation study to evaluate the feasibility of midgap metal gates on ultra short channel length (25 nm) CMOS transistors.

The replacement of poly-Si gates with metal gates to eliminate the adverse effect of poly depletion in near future has been suggested in the International Technology Roadmap for semiconductors [21]. The commonly accepted belief is to use bandedge metal gates (or dual metal gates) for the best performance-around 4.15 eV for NMOSFETs and 5.25 eV for PMOSFETs [22]. This presents immense process complexity (because of the use of separate gate electrode materials for NMOS and PMOS devices), and actual incorporation into a regular commercial process flow is difficult. In this context, using a common work-function gate both for NMOSFETs and PMOSFETs with a value around the midgap (around 4.7 eV) seems very attractive. Several studies (both simulation and

experimental) have been reported over the past few years which predicted that a midgap metal gate is unlikely to match the performance of bandedge metal gates [22]. Studies also indicate that in order to set the correct threshold voltage of a transistor with midgap metal gates, a buried channel structure is necessary [22, 23]. At the same time, some studies predict that significant performance improvement over poly-Si gates is possible with W/TiN midgap gate stacks [24, 25]. Indranil De *et al.* [21] tried to capture the advantage of higher work function and hence V_t of midgap metal gates by using a super steep retrograde channel (SSR) profile, the reduced channel doping was expected to give higher drive current (I_{on}) by way of mobility improvement due to reduced doping. The inherent problem with this scheme requires a very low doped (nearly intrinsic) channel region and a very heavily doped (nearly degenerate) substrate immediately below it to capture the full advantage of the proposed scheme. It may be difficult to realize such a device geometry practically. Both Josse *et al.* [24] and Matsuda *et al.* [26] have studied the impact of using midgap metal gates in buried channel MOSFETs with pocket halo implants. However, buried channel structure poses process complexity, viz, use of disposable poly gate to introduce the channel counterdoping [26]. Also most of the prior literature [23, 25, 27] deals with fairly long channel lengths, dimensions at which metal gates may not come into the picture [21]. Here, we consider the conventional bulk MOSFET structure, with lightly doped drain (LDD) and pocket halo implants [28] at ultra short channel lengths (metallurgical channel length, also identified earlier as the limit of bulk CMOS scaling in [29]). We compare the performance of midgap metal gates against their bandedge metal gate and polysilicon gate counterparts, and also comment on their relative merits and demerits. The strength of this scheme lies in the fact that pocket

halos can be suitably engineered to set the desired level of off state leakage (I_{off}) . It is worth mentioning that in order to set the right I_{off} , the channel doping in midgap metal gate transistors needs to be considerably lower than the bandedge metal gate ones, which in turn implies that short channel behavior of the midgap devices will be considerably degraded. This worsening of short channel behavior would manifest itself in the form of higher DIBL. It is also instructive to note that halo engineering is a common practice in the VLSI industry today, and hence we do not need to deviate a lot from the conventional bulk CMOS process flow to fabricate these structures.

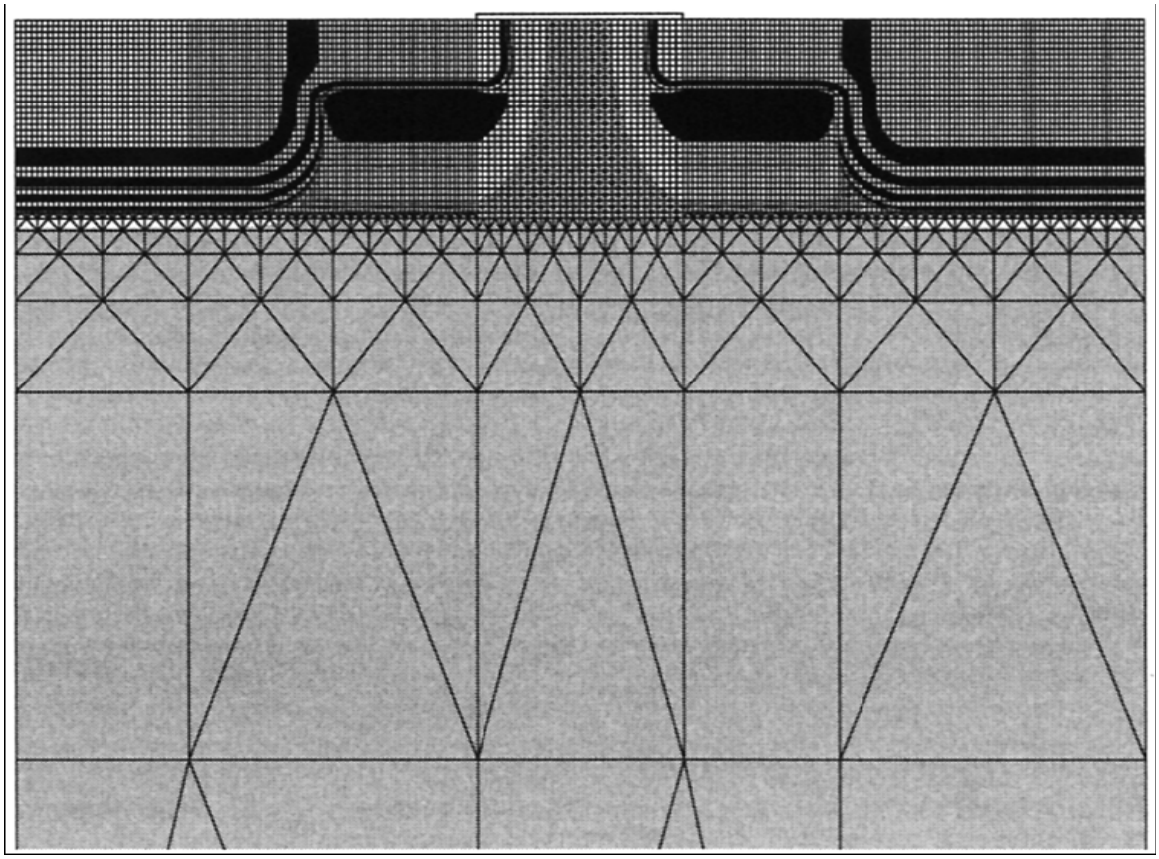


Figure 5.7. Schematic showing MOSFET structure used in device simulation.

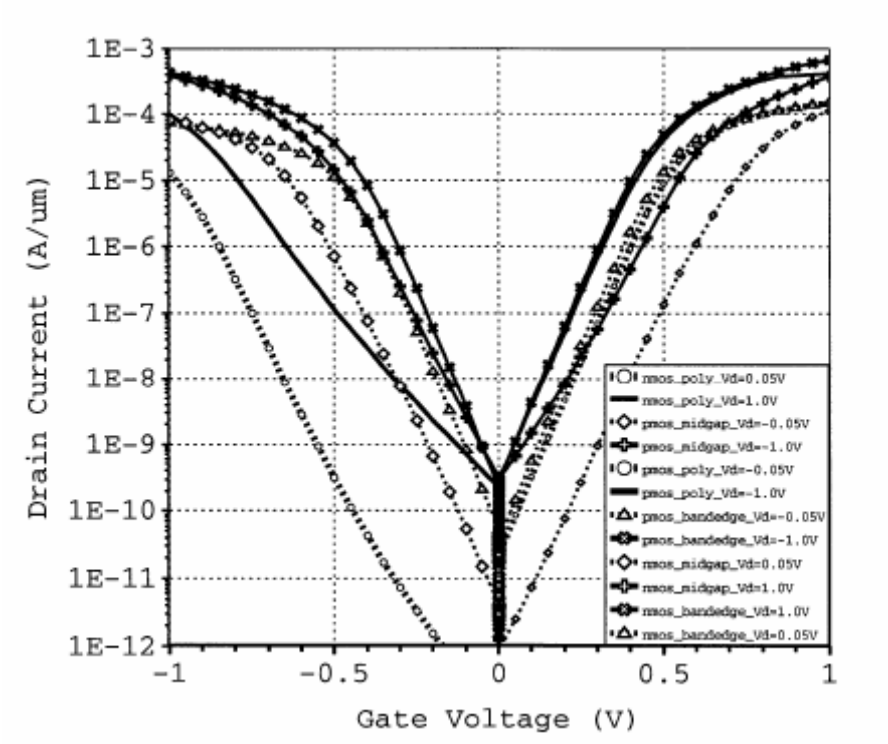


Figure 5.8. I_d-V_g Characteristics for shorter channel devices (L_-).

A commercial device simulator DESSISe from ISE [30] has been used for this work. Classical drift diffusion model with Van Dort quantum correction [31] has been used for the relative performance evaluation of the device geometries. The relative simplicity of these models makes them very handy for ultrashort channel device simulations [22]. Three devices with different L_{met} have been taken into consideration, *viz.*, 22.5 nm, 25 nm, and 27.5 nm (identified as L_- , L and L_+ respectively) keeping in mind around 0.1 L_g lithographic tolerance on the nominal physical gate length (L_g) [21]. Pocket halo implants are simulated by Gaussian profiles (Figure 5.7). Analytical profiles have also been used to make the deep source/drain and source drain extension (SDE) regions. Very abrupt and highly nonuniform doping profiles have been adopted for the halos to suppress short channel effects. Similar 25 nm device designs with superhalo engineering had been

reported earlier [29, 32]. Fairly shallow junctions (deep source drain of 30 nm and shallow source drain of 15 nm) have been used to suppress short channel effects. The constant background substrate doping and the halo doping were adjusted such that the off-state leakage for L_- devices is 0.3 nA/μm, the low operating power leakage specification in [21]. It is worth emphasizing that the L_- devices are expected to have the worst short channel effects, and the L_+ devices are expected to have the worst on-state current [22]. Further, full device simulations were performed on a five stage linear chain of inverters for both bandedge and midgap metal gate transistors to understand the circuit behavior of midgap metal gate transistors compared to bandedge metal gate transistors. Quantum corrections were considered for the full device simulations of the inverter circuit. A low leakage dielectric with effective oxide thickness of 1.0 nm has been assumed for all the devices. The supply voltage V_{dd} value of 1.0 V has been chosen.

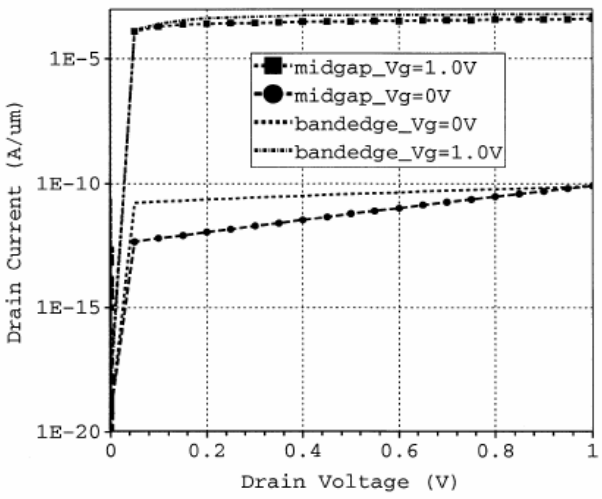


Figure 5.9. I_d-V_d characteristics for the longer channel devices

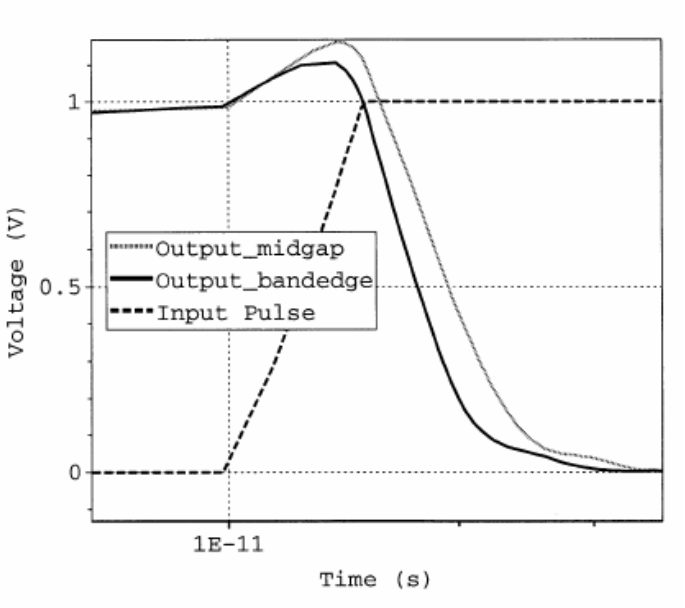


Figure 5.10. Falling edge waveforms of a five stage linear chain of inverters

Figure 5.7 illustrates the MOSFET structure used in our device simulations, The pocket halos are depicted by dark shades just below the source drain extensions (SDE). Figure 5.8 depicts the I_d-V_g plots of the L. devices, it is worth noting that the expected mobility improvement because of lower doping in midgap metal gate NMOSFETs fails to offset the negative impact of higher V_t in NMOSFETs, hence the I_{on} is lower compared to the bandedge metal gate devices. An I_{on} degradation of about 36% is observed in midgap metal gate NMOSFETs over bandedge metal gate NMOSFETs. However, the midgap metal gate NMOSFETs are fairly close in performance to the n+ poly-Si NMOSFETs (an 11% degradation in I_{on} is observed). Severe short channel degradation is noticed in the midgap metal gate NMOSFETs. It has been observed that midgap metal gate NMOSFETs have a DIBL (drain induced barrier lowering) value twice as large as bandedge NMOSFETs (around 80 mV/V for bandedge metal gate NMOSFETs, and 160 mV/V for midgap metal gate NMOSFETs). It is instructive to note that the short channel

behavior of poly-Si NMOSFETs (n+ poly-Si doping of 1 x 10^{20} cm^{-3}) is very similar to that of bandedge NMOSFETs (DIBL value around 90 mV/V). However, an interesting phenomenon is observed in case of PMOSFETs. Due to higher poly depletion of the p+ poly gates, midgap metal gate PMOSFETs performance is significantly better than the poly-Si devices (p+ poly-Si doping level of 1 x 10^{20} cm^{-3}). It was observed that the value in case of midgap PMOSFETs is thrice as large as that of poly-Si PMOSFETs. However, the performance is lower compared to bandedge metal gate PMOSFETs. Figure 5.9 depicts the I_d-V_d characteristics of the longer channel (L_+) NMOS devices , the slope observed in the low gate voltage I_d-V_d characteristics of midgap devices is attributed to higher DIBL for these devices as pointed out earlier. Figure 5.10 gives a comparison between the circuit behavior of the bandedge metal gate and the midgap metal gate transistors. The devices considered are the L_+ transistors, keeping in mind that these devices are expected to have the worst case circuit delays because of lower I_{on} values. A five stage linear chain of inverters is simulated at the device level, wherein the circuit equations and the transport equations in the devices have been simultaneously and self-consistently solved. The output is recorded at the first stage and the delay of the falling edge is measured for both the midgap and bandedge linear chains. The delay of the bandedge metal gate linear chain is found to be 5.16 ps whereas the delay of the midgap metal gate linear chain is 6.91 ps at V_{dd}=1.0 V , clearly a degradation of about 34% is observed in case of midgap metal gate devices over bandedge metal gate devices.

Thus, we have shown that from setting the right threshold voltage perspective, midgap metal gates may not be a suitable choice. Tradeoffs do exist between process simplicity

associated with midgap metal gates over bandedge metal gates, expected higher mobility at aggressively scaled T_{inv}'s with midgap metal gates, and setting the bandedge V_t for the devices. All these factors need to be simultaneously taken into account before choosing the right workfunction for transistors. In the next section, we discuss the beneficial impact an ideal metal gate is expected to have on mobility in the channel by possible screening of the soft optical phonon modes in the high-κ insulator based NMOSFETs. First, in the following section, the theoretical formalism leading to metal gate screening is discussed. And in the section immediately afterwards, novel ultra low temperature measurements are performed on the aggressively scaled TiN stack to experimentally study the validity of the metal gate screening hypothesis.

5.5 Role of Soft optical phonon screening in high-κ gate stacks by ideal metal gates

Based on the seminal work by R. Kotlyar et. al in IEDM 2004 [33], the essential steps of the metal gate screening hypothesis are given below. It should be noted here that though the original work on metal gate screening hypothesis has been published earlier, the present theoretical work has been performed independently of Ref. [33]. Also the present theoretical scheme can be extended to include any non-ideal metal gate unlike [33]. A direct comparison between poly-Si and ideal metal gate mobility was not carried out in [33].

A three layer structure (Figure 5.11) with ideal metal or non-ideal poly-Si gate (doping ~ 1 X 10^{19} cm^{-3}) ($z < 0$), high-κ/HfO_2 gate stack ($0 \leq z < t$) with physical thickness of t nm,

and bulk-Si channel ($z \geq t$) is shown. The phonon potentials, as described in [1], are given by the following equations:

(i) For Poly-Si gate,

$$\Phi_{Q,\omega_Q^{(i)}} = \begin{cases} a_{Q,\omega_Q^{(i)}} e^{Qz} & z < 0 \\ b_{Q,\omega_Q^{(i)}} e^{-Qz} + c_{Q,\omega_Q^{(i)}} e^{Qz} & 0 \leq z < t \\ d_{Q,\omega_Q^{(i)}} e^{-Qz} & z \geq t \end{cases} \quad \text{……(5.1)}$$

(ii) Whereas for an ideal metal gate, because of high carrier concentration in the gate, the field lines do not penetrate. Naturally,

$$\Phi_{Q,\omega_Q^{(i)}} = \begin{cases} 0 & z < 0 \\ b_{Q,\omega_Q^{(i)}} (e^{-Qz} - e^{Qz}) & 0 \leq z < t \\ d_{Q,\omega_Q^{(i)}} e^{-Qz} & z \geq t \end{cases} \quad \text{……..(5.2)}$$

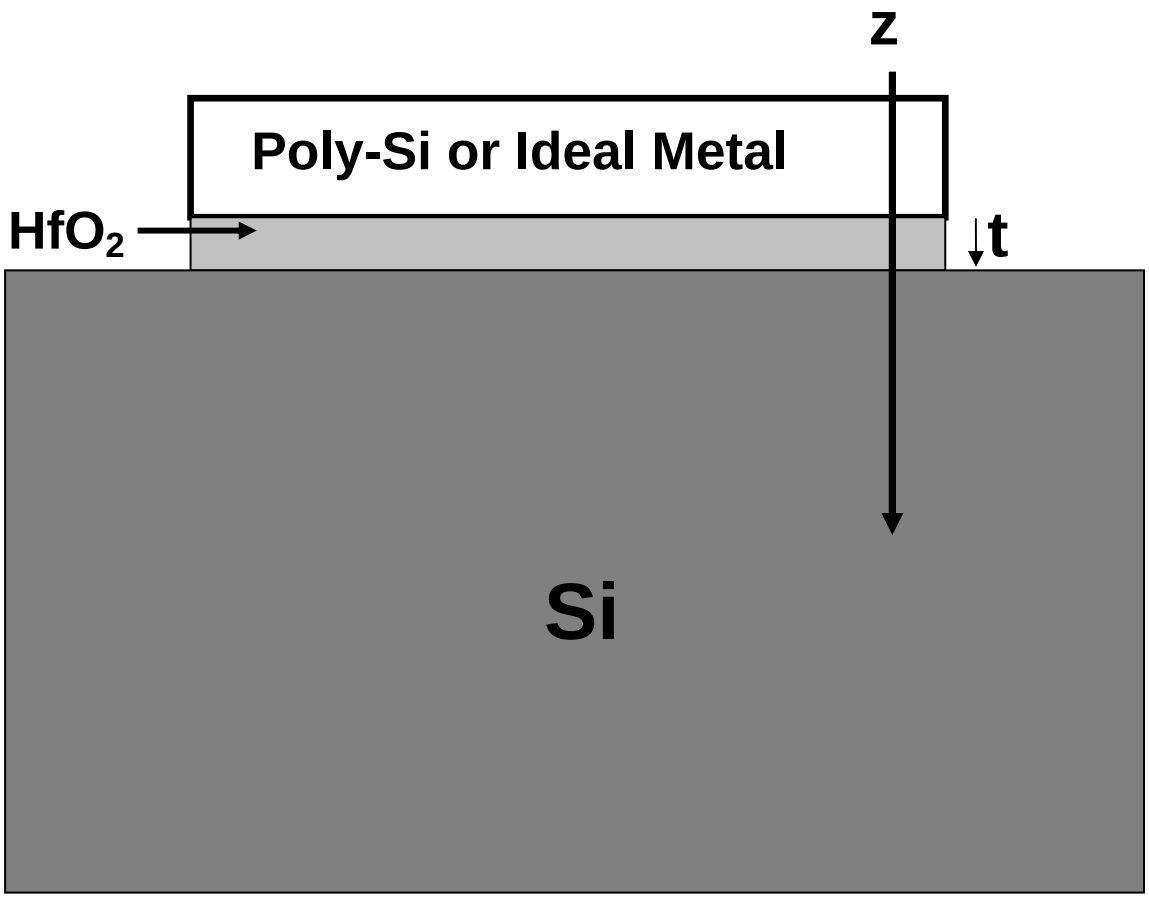


Figure 5.11. The tri-layer structure showing the direction of the electric field as illustrated in equation 5.1 and 5.2. (z =0 is @ the gate electrode HfO_2 interface)

Applying continuity of electric field at the two different interfaces, we get the phonon dispersion relations for two different systems, depicted in Figure 5.12.

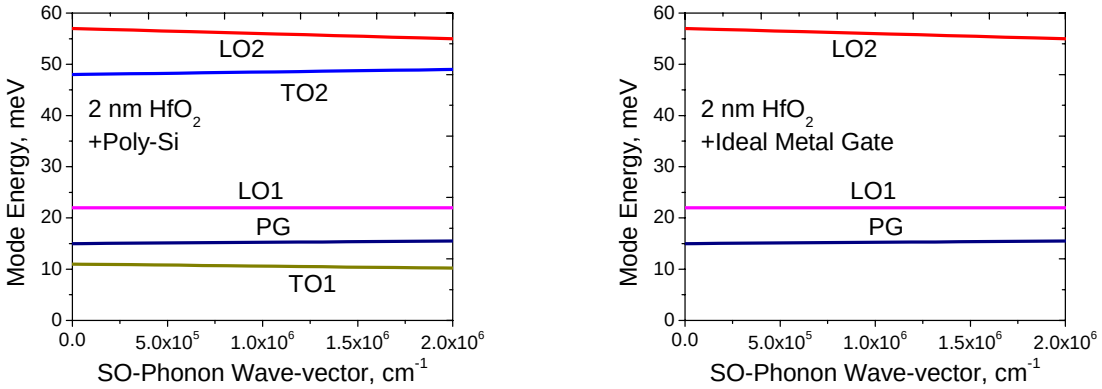


Figure 5.12. Phonon dispersion relations for the high-κ insulator in presence of ideal metal gate and poly-Si gate

As evident from Figure 5.12, with an ideal metal gate, two of the phonon modes (namely TO1 and TO2) phonon modes are absent which are present in the poly-Si/high-κ gate stack. We would like to ascertain the impact disappearance of phonon modes has on mobility. Once the phonon dispersion relations are known, the phonon content of each mode is estimated using simple Bose Einstein statistics. With the phonon content of each mode evaluated, the phonon content is plugged into the Kubo Greenwood transport integral to estimate the phonon limited mobility. It is illustrated in Figure 5.13, that with ideal metal gate, there is approximately about 14% improvement in mobility at low fields. Expectedly, the curves tend to merge together at high fields, because the length scale of relevance here is guided by the Fermi wave vector which implies that in order for the metal to screen at higher fields, the metal needs to be placed closer to the channel

electrons. This will be significant when we discuss the experimental verification of metal gate screening hypothesis in the next section.

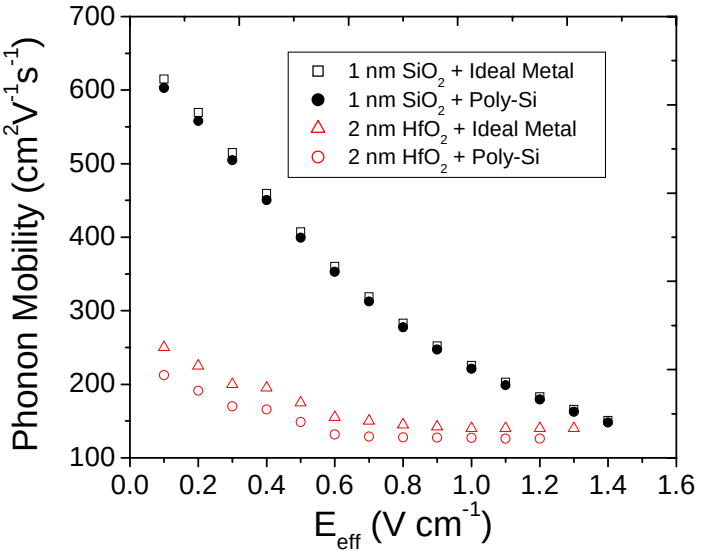


Figure 5.13. Phonon limited mobility for high-κ insulator based NMOSFETs with Poly-Si and ideal metal gate. The SiO_2 case is also given here from [34] for reference, as expected, the difference between the ideal metal gate and the poly-Si gate is minimal. (The mobility values are corrected for bulk phonon and surface roughness using the methodology described in Chapter 4)

5.6 Experimental Verification of Metal Gate Screening-Ultra low temperature measurements

In order to experimentally validate the hypothesis of metal gate screening, we adopt an approach used by Z. Ren et. al. [16] to evaluate the temperature exponent of the mobility in the intermediate field range. We consider the high mobility, aggressively scaled (EOT~ 1.0 nm) TiN and HfO_2 stack described earlier. A 300 mm wafer compatible industrial prototype prober from Karl Suss Microtech capable of going down to liquid He

temperatures was used for the low temperature measurements [34]. The mobility curves at different temperatures are shown in Figure 5.14.

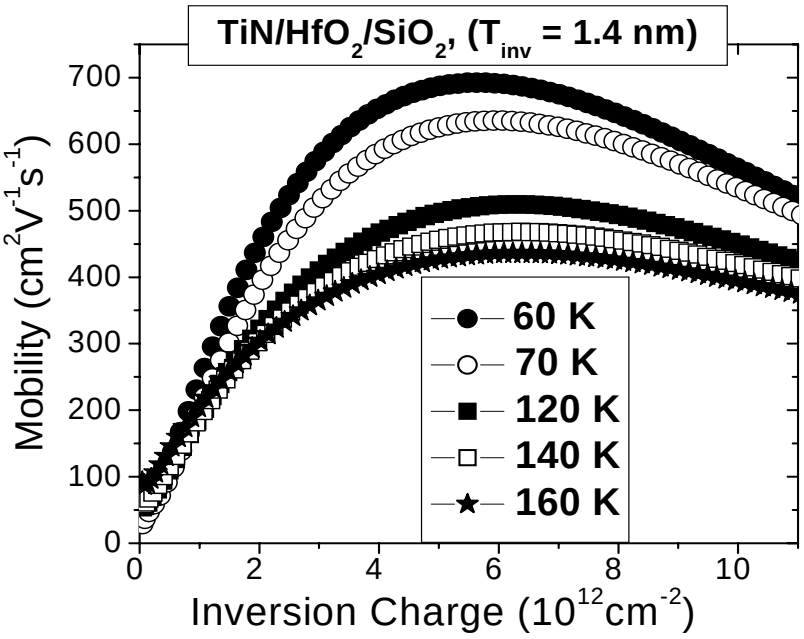


Figure 5.14. Family of mobility curves for the aggressively scaled TiN + HfO_2 gate stacks at different temperatures

Plotting the temperature exponent for these stacks, we find that there is a weak temperature dependence like poly-Si/HfO_2 gate stacks as depicted in Figure 5.15. It needs to be emphasized here that though the temperature dependence is weak, it may not be conclusively stated that metal gate screening does not occur for the reasons stated as follows:

(i) The metal gates used in this experiment are not necessarily ideal

(ii) The impact of metal gate screening is expected to be stronger for thinner T_{inv}'s for reasons stated in section 5.5.

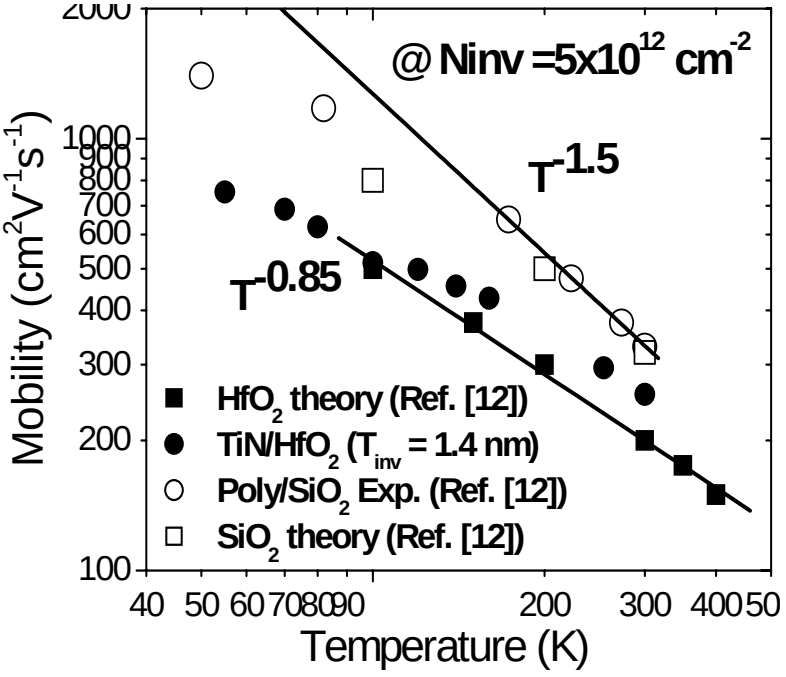


Figure 5.15. Temperature exponent for high-κ gate stack with poly-Si and metal gate stack. Comparison with poly-Si/SiO_2 conventional NMOSFETs is also made.

5.6 Some comments about low temperature measurements

In this section, we would highlight briefly an aspect which is expected to be important in any low temperature measurements involving MOS transistors:

The impact of series resistance, as mentioned briefly in Chapter 2 assumes added significance in light of low temperature measurements primarily because at low temperatures, the channel conductance shoots up significantly because of increase in channel mobility, which in turn implies that series resistance of the source drain junctions

starts influencing the device performance more than it does at room temperature. So, it is imperative that this factor should be kept in mind while conducting low temperature measurements on MOS transistors. Using shift and ratio technique described in [35], we determine series resistance for the TiN based HfO_2 MOSFETs and find that the transistors have an extremely high series resistance of 164 ohm (Figure 5.16) at the temperature of T=40 K . It is instructive to note at this point that this factor will be more pre-dominant at shorter channel transistors than long channel ones. Or in other words, for ultra short channel devices, mobility is probably less pressing a concern than series resistance of the source and drain junctions, because it is the latter which will set the I_{on} of the transistors.

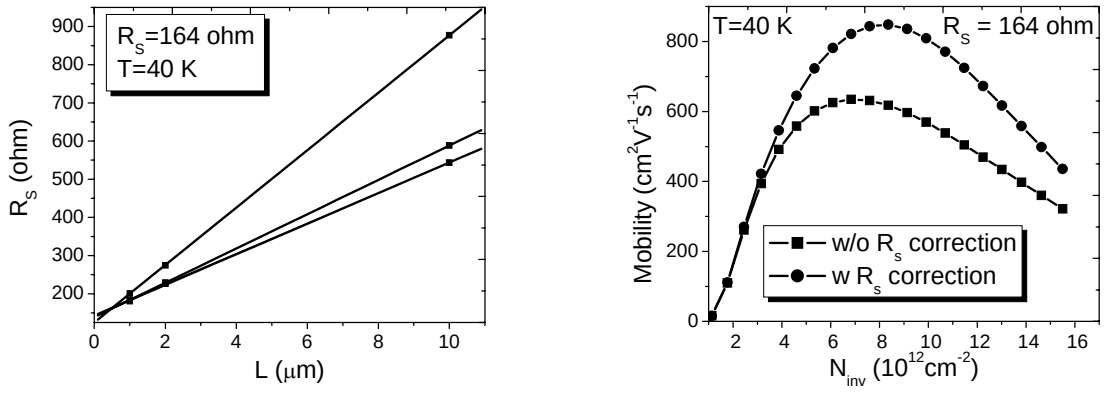


Figure 5.16. Determination of series resistance for the aggressively scaled, high mobility TiN+HfO_2 gate stack (EOT~ 1.0 nm). The series resistance (V_{dd}/I_d) are evaluated for three different gate overdrives [(V_{GS}-V_{Th})=0.5, 0.6 and 0.7 V]. Evidently, the mobility correction is substantial

5.7 Conclusions

In the foregoing discussion, we have described the process details that lead to high mobility in aggressively scaled metal gate/high-κ NMOSFETs. Particularly, the necessity of the high temperature process step had been given maximum attention. The physical origin of the high mobility post high temperature process step is thoroughly investigated. The role of metal gate screening in high mobility aggressively scaled metal gate high-κ gate stacks is investigated by novel ultra low temperature measurements.

So far we have completely ignored the role of interfacial oxide layer thickness. But as described in [1], the interfacial oxide layer can have a profound impact on the phonon limited mobility in high-κ based gate stacks. The interfacial oxide layer, be it oxynitride or SiO_2 shields the soft optical phonon modes from coupling with the channel electrons very strongly. This is illustrated in Figure 5.17. The theoretically computed phonon limited mobility is plotted as a function of interfacial oxide layer thicknesses (solid line) [1] and the theoretical curve is corrected for ionized impurity scattering in the channel by using the parameterized formulae described in Chapter 2. The theoretical trend curve shown is compared with mobility at intermediate inversion charge densities where phonon scattering is expected to dominate. Interestingly, the experimental curves tend to follow the theoretical trend line qualitatively thereby suggesting that process optimization has helped us to achieve the best mobility possible on a bulk-Si NMOSFET with high-κ insulator and the high-κ mobility are approaching the theoretically predicted upper limits for the high-κ gate stack.

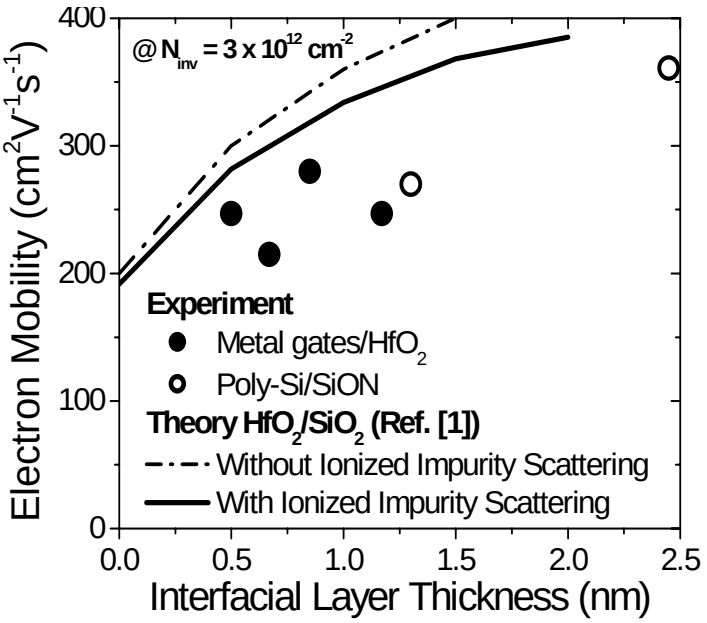


Figure 5.17. Dependence of phonon limited mobility on interfacial oxide layer thicknesses

References

[1]. M.V Fischetti, D.A Neumayer, and E. A Cartier, "Effective Electron Mobility in Si Inversion Layers in Metal-Oxide-Semiconductor Systems with a High-K Insulator: The Role of Remote Phonon Scattering", 2001, Journal of Applied physics, Vol. 90 , No. 9, pp. 4587-4608.

[2]. R. Chau, S. Dutta, M. Doczy, B. Doyle, J. Kavalieros, and M. Metz , "High-κ/Metal Gate Stack and its MOSFET Characteristics", IEEE Electron Device Letters, 2004, Vol. 25, No. 6, pp. 408-410.

[3]. W. Zhu, J. P Han, and T. P Maa, "Mobility Measurements and Degradation Mechanism of MOSFETs made with Ultrathin High-K Dielectrics", IEEE Transactions on Electron Devices, 2004, Vol. 51, No. 1, pp. 98 – 105.

[4]. E. Gousev et. al., "Advanced Gate Stacks with Fully Silicided Gates and High-κ Dielectrics: Enhanced Performance at Reduced Gate Leakage", in IEDM Technology Digest 2004.

[5]. S. Callegari et. al., " Interface engineering for enhanced electron mobilities in W/HfO_2 gate stacks", in IEDM Technology Digest 2004, pp. 825-828.

[6] E. A Cartier et. al., "Systematic study of pFET V_t with Hf-based gate stacks with poly-Si and FUSI gates", 2004, VLSI Technology Symposium, pp. 44-45.

[7] S. B. Samavedam *et al,* "Dual-metal gate CMOS with HfO_2 gate dielectric" IEDM Tech. Dig. **,** 2002, pp. 433-436

[8] W. Tsai et al, "Performance comparison of sub 1 nm sputtered TiN/HfO_2 nMOS and pMOSFETs" IEDM Tech. Dig.**,** 2003, pp. 311-314.

[9] P.D. Kirsch et al., "Mobility Enhancement of High-k Gate Stacks Through Reduced Transient Charging", in Proceedings of the 35th European Solid-State Device Research Conference, Grenoble, France, 2005, pp. 367- 368.

[10] V. Narayanan *et al*, "Dual work function metal gate CMOS using CVD metal electrodes", VLSI Tech. Digest**,** 2004, pp. 192-193

[11] Y. Akasaka *et al*, " Impact of Electrode-side Chemical Structures on Electron Mobility in Metal/ HfO_2 MISFETs with Sub-1nm EOT," VLSI Tech. Digest**,** 2005, pp. 228-229

[12] V. Narayanan, E. Cartier, K. Maitra et. al.., "Process Optimization for High Electron Mobility in Aggressively Scaled Metal Gate High-κ Gate Stacks", (Submitted to IEEE Electron Device Letters)

[13] L.-Å. Ragnarsson *et al.*, "High Performing 8Å EOT HfO2 / TaN Low Thermal-Budget n-channel FETs with Solid-Phase Epitaxially Regrown (SPER) Junctions", VLSI Technology Digest**,** 2005, pp. 234-235.

[14] R. J Carter, E. Cartier, A. Kerber, L. Pantisano, and T. Schram, "Passivation and Interface state density of High-κ Gate Stacks", Appl. Phys. Lett., 2003, Vol. 83, No. 3, pp. 533-535.

[15] G. Lucovsky and J. C. Philips, "Bond Strain and defects in high-k gate stacks" , Microelectronics Reliability, 2005, Vol. 45**,** pp 770-778.

[16] Z. Ren *et al.*, "Inversion channel mobility in high-k high performance MOSFETs", IEDM Tech. Dig.**,** 2003, pp. 793-796.

[17] S. Stemmer, Z. Chen, C. G. Levi, P. S. Lysaght, B. Foran, J. A. Gisby and J. R. Taylor, "Application of Metastable Phase Diagrams to Silicate Thin Films for Alternative Gate Dielectrics", Jpn. J. Appl. Phys. Part 1, 2003, Vol. 42, pp. 3593-3597.

[18] H. S. Baik, M. Kim, G.-Su Park, S. A. Song, M. Varela, A. Franceschetti, S. T. Pantelides, and S. J. Pennycook, "Interface structure and non-stoichiometry in HfO_2 dielectrics", 2004, Appl. Phys. Lett., Vol 85, pp. 672- 674

[19] M. P. Agustin, L. R. C. Fonseca, J. C. Hooker, and S. Stemmer, "Scanning transmission electron microscopy of gate stacks with HfO_2 dielectrics and TiN electrodes", Appl. Phys. Lett., 2005, Vol. 87, 12909-12911.

[20] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the Universality of Inversion Layer Mobility in Si MOSFETs: Part I-Effects of substrate Impurity Concentration", IEEE Transactions on Electron Devices, 1994, Vol. 41, No. 12, pp. 2357-62.

[21] The International Technology Roadmap for Semiconductors (2004), available online at http://public.itrs.net.

[22] I. De, D. Johri, A. Srivasatava, and C. S Osburn, "Impact of Gate Workfunction on Device Performance at the 50 nm Technology node", Solid State Electronics, 2000, Vol. 44, pp. 1077-1080.

[23] Y. Abi, T. Oishi, K. Shiozawa, Y. Tokuda, and S. Satoh, "Simulation Study on Comparison between Metal gate and poly-Si gate for sub-quarter μm MOSFETs", IEEE Electron Device Letters, 1999, Vol. 20, pp. 632-634.

[24] E. Josse, and T. Skotnic, "Poly-Si gate with Depletion or Metallic Gate with Buried Channel-What evil worse?", in IEDM Tech. Dig., 1999, pp. 661-664.

[25] R. Li, and Q. Xu, “Damascene W/TiN Gate MOSFETs with improved performance for 0.1 μm regime”, IEEE Trans. Elect. Dev.(To be published)

[26] S. Matsuda, H. Yamakawa, A. Azuma, and Y. Toyoshima, “Performance Improvement of Metal Gate CMOS Technologies”, in VLSI Tech. Dig., 2001, pp. 63-64.

[27] B. Cheng, B. Maity, S. Samavedam, J. Grant, B. Taylor, P. Tobin, and J. Mogab, “Metal Gates for Advanced sub-80 nm SOI CMOS Technology”, IEEE Int. SOI Conf., 2001, pp. 91-92.

[28] C. H Wann, K. Noda, T. Tanaka, M. Yoshida, and C. Hu, “A Comparative Study of Advanced MOSFET Concepts”, IEEE Trans. Elect. Dev., 1996, Vol. 43, pp. 1742-1753.

[29] ISE TCAD 10.0, ISE (Now Synopsis)

[30] DESSISe Manuals, ISE TCAD (Now Synopsis)

[31] Y. Taur, C.H Wann, and D.G Frank, “25 nm CMOS Design Considerations”, in IEDM Tech. Dig., 1998, pp. 789-792.

[32] Y. Taur, and E. J Nowak, “CMOS devices below 0.1 μm: How high will performance go?”, in IEDM Tech. Dig., 1997, pp. 215-218.

[33] R. Kotlyar et. al., “Inversion Mobility and Gate Leakage in High-κ/Metal gate MOSFETs”, in IEDM Tech. Dig., 2004

[34] Karl Suss Microtech, Conti Cryoprober Cryostat User Manuals.

[35] Yuan Taur, and T. H Ning, “Fundamentals of Modern VLSI Devices”, Cambridge University Press, 1998.

Chapter 6

Conclusions and Scope of Future Work

6.1 Contributions

We summarize here the contributions made so far in this thesis.

(i) A novel variant to the inversion charge pump based mobility estimation technique is developed (Chapter 2). It is shown that the one and two contact measurement scheme described in literature [1] does not necessarily yield the correct inversion charge for accurate estimation of mobility in absence of charge trapping. Instead, we used two symmetric devices (namely 50/50 µm/µm and 100/100 µm/µm) and carried out inversion charge pump measurements to estimate inversion charge in absence of trapping. Assuming inversion charge loss to occur from the edges of the inverted channel, we developed a correction technique given by equation (2.3) to accurately compute the inversion charge. Using the technique so developed, we benchmarked the different frequency inversion split-CV's to choose the most accurate split-CV/I_d-V_g technique for mobility estimation.

(ii) In Chapter 3, a set of Hf-silicate based samples are chosen to ascertain semi quantitatively the location and magnitude of trapped charge. The samples are so designed that varying cycles of ALD Al_2O_3 capping layers are applied between the Hf-silicate layer and the poly-Si gate thereby ensuring that the ground plane where the charge in high-κ is screened is pushed farther and

farther out. By using a simple model, we determined the first moment of charge centroid uniquely, which suggests that the stable charge is located in the Hf-silicate gate stack rather than in the Al_2O_3 capping layer. This is further corroborated recently by ESR measurements by a French group on these stacks [2], who confirmed appearance of a response from the X-center purported to be located in the Hf-silicate stack, rather than in the Al_2O_3 cap. The possibility of a connection between the two warrants further investigation. To conclude, if Al_2O_3 caps are used for V_t adjustment with high-κ gate stacks, charge trapping from these stacks would not be a serious limiting factor.

(iii) A novel N_{it} calibration experiment was performed to understand and quantify the response of interface states on transport in the channel. It is conclusively shown for the first time that electrically the interface states at the high-κ/SiO_2/(100)-bulk Si interface and the SiO_2/(100)-bulk Si interface are identical. A novel electrical measurement technique was also devised to conclusively demonstrate that "charges far away from the interface" have minimal impact on transport of electrons in the channel.

(iv) The role of metal gates on the soft optical phonon modes in the high-κ gate stack is investigated by an independent theoretical calculation already reported in literature [3]. Novel ultra low temperature measurements were performed on gate stacks with aggressively scaled EOT's and record high

electron mobilities to objectively investigate the beneficial impact metal gates are expected to have on soft optical phonon limited electron mobility in high-κ insulator based NMOSFETs. Tell-tale signatures of mode stiffening is observed but the effect is considered insignificant when compared with the role of other limiting factors, for instance the impact of interfacial oxide layer thickness. It is conjectured that if the high-κ layer is scaled further, the role of soft optical phonon screening will become more discernible, given the proximity of the ideal metal gate to the surface phonon modes in the underlying high-κ insulator layer in the scaled high-κ gate stack. A possible direct experimental technique to understand the full extent of metal gate screening will be discussed in the next section. However, it is conclusively shown that with HfO_2 based gate stacks, even at very thin EOT's (EOT~1.0 nm), mobilities comparable to much thicker (~2.3-2.5 nm) SiON stacks may be achieved. It is also demonstrated that for ultra short channel transistors, mobility is as important as the role of the junction series resistance when considering the on state drive current as the performance metric.

(v) Recent developments [4] lead to high mobility high-κ based metal gate NMOSFETs. These studies stressed on the need for a high temperature process step which brought about a structural change in the high-κ gate stack itself, leading to the attainment of high mobility despite minimal change of EOT. However, the physical origin of high mobility and the actual chemistry associated with the high temperature process step is not clearly understood. It

had been hinted in [4] that the high process temperature induced structural change brings about a change in the phonon response of the stack with respect to the channel electrons. For the first time, we provide experimental evidence which indicate that indeed the phonon spectra of the stacks pre and post high temperature anneal may be completely different (Chapter 4)

6.2 Scope of Future Work

It is instructive to note here, that mobility measurements are an extremely indirect way to ascertain the phonon response of the high-κ gate stacks on channel electrons. A method of directly observing surface phonon modes in high-κ gate stacks in presence of metal gates may be attempted on similar lines as Martin Frank et. al. [4]. It is well known that one of the popular methods of measuring the phonon spectra in high-κ gate stacks is through FTIR (Fourier Transform Infra Red) spectroscopy, wherein the vibrational phonon modes are excited and their traces thereby measured and quantified. However, accurately tracing the disappearance of phonon modes in presence of pure metallic layers, is a very difficult prospect. Particularly, getting the transmission response out of the metallic layer is a very difficult task. To get around this problem, we suggest two possible alternatives:

(i) Use of metallic clusters as suggested in Martin Frank's work [5], on the high-κ substrates is a possibility, analyzing the local phonon response to investigate

the missing modes in presence of metallic layer, and then extrapolating it to reconstruct the global phonon response in a tri-layer system (namely: metal gate, HfO_2 and bulk-Si). This is very difficult again because of the uncertainty involved in connecting the local phonon response with the global phonon response.

(ii) Alternatively, instead of using a transmission FTIR setup, one may use a reflection FTIR setup, a costly modification to the well known FTIR equipments. However, this is a very credible way to measure the actual phonon response in the tri-layer system mentioned above. This may yield very objective and direct insight about the metal gate screening problem.

So far, we have demonstrated that the transport properties of the channel electrons namely the different scattering mechanisms which affect I_{on} in high-κ insulator based NMOSFETs, can be studied in a similar fashion as the conventional SiO_2 NMOSFETs. Apart from the soft optical phonon scattering, no other scattering mechanism in the channel is unique to the high-κ insulator. We have conclusively shown that even in presence of soft optical phonon scattering, high-κ insulator based NMOSFETs can exhibit mobilities competitive with SiON at higher effective oxide thicknesses. So from purely a transport perspective, mobility degradation should not be an impediment to integrating high-κ based gate stacks in advanced generation CMOSFETs.

References

[1] A. Kerber, “Methodology for Electrical Characterization of MOS devices with Alternative Gate Dielectrics”, PhD Thesis, T.U Darmstadt, and IMEC Belgium, 2003.

[2] H. J Von Bardeleben et. al., Presented at the NATO Advanced Research Work shop, Moscow 2005.

[3] R. Kotlyar et. al., “Inversion Mobility and Gate Leakage in High-κ/Metal Gate MOSFETs”, in IEDM Tech. Dig., 2004.

[4] S. Callegari et. al., “ Interface engineering for enhanced electron mobilities in W/HfO_2 gate stacks”, in IEDM Technology Digest 2004, pp. 825-828.

[5] Martin Frank et. al., “Phonons of Clean and Metal Modified Oxide Films”, Surface Science, 2001, Vol. 492, pp. 270-284.