Caractérisation de transport des électrons dans les transistors MOS à canal court

Narasimhamoorthy Subramanian

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Nano électronique et nano technologies**

Présentée par

Narasimhamoorthy SUBRAMANIAN

Thèse dirigée par : **Mireille MOUIS**
Codirigée par : **Gérard GHIBAUDO**

Préparée au sein du, **IMEP-LAHC, MINATEC, INP-Grenoble**
dans, **l'École Doctorale « Electronique, Electrotechnique, Automatisme et Traitement du Signal »**

« Caractérisation de transport des électrons dans les transistors MOS à canal court »

Thèse soutenue publiquement le **« 29 Nov 2011 »** devant le jury composé de :

M. Abdelkader SOUIFI
Professeur – INSA de Lyon, Président (proposé)

M. Arnaud BOURNEL
Professeur – Université de Paris Sud, Rapporteur

M. Philippe LORENZINI
Professeur – Université de Nice, Rapporteur

M. Stéphane MONFRAY
Ingénieur – ST Microelectronics Crolles, Membre

MME. Mireille MOUIS
Directeur de Recherche CNRS, Membre

M. Gérard, GHIBAUDO
Directeur de Recherche CNRS, Membre

Résumé

La qualité du transport électronique est l'une des clés permettant de soutenir la progression des performances pour les futures générations de composants. De très nombreux facteurs, comme le choix de l'isolant et du métal de grille, le matériau de canal ou la présence de contraintes mécaniques, affectent de façon négative ou positive ces propriétés de transport. L'épaisseur du canal, qui atteint des dimensions nanométriques joue également un rôle : interactions avec les interfaces, fluctuations d'épaisseurs, effets de couplage électrostatique ou quantique entre ces interfaces. Il est probable que des mécanismes d'interaction associés à la proximité des zones surdopées de source et de drain puissent également intervenir. A ces dimensions, on s'attend à observer des phénomènes de transport hors d'équilibre, voire balistique, qui peuvent remettre en question la validité des paramètres utilisés pour caractériser le transport. Donc avec l'avancement de la technologie, il devient nécessaire de faire évoluer les modèles de transport et les paramètres afin de mieux expliquer le fonctionnement du MOSFET. Cette thèse se concentre sur la compréhension des modèles de transport existants et des méthodes d'extraction pour les nœuds technologiques actuels et futures. Les modèles de transport et les méthodes d'extraction de paramètres en régime linéaire et de saturation ont été explorés au cours de cette thèse. L'impact de la résistance série, qui est une fonction de la tension de grille, dans les MOSFET avancés est pris en compte et une nouvelle méthode d'extraction améliorée a été développée dans le régime linéaire. Des mesures à basse température ont été utilisées en régime linéaire pour l'extraction des mécanismes de diffusion en utilisant le modèle de mobilité. Une nouvelle méthode de correction pour le courant de drain dans le régime de saturation pour les MOSFET canal court est développée en utilisant les mesures à basse température. Cela permet de corriger du DIBL ainsi que des effets de « self heating ». Le modèle de saturation de vitesse et la méthode d'extraction associée sont explorés dans le régime de saturation et sont étudiés en fonction de la température et de la longueur de canal. Les modèles balistique et quasi-balistique avec le concept de la « kT layer » en régime de saturation sont également étudiés pour les nœuds sub 32 nm. Mesurer la magnétorésistance offre des perspectives prometteuses pour les dispositifs à canal court et permettant d'extraire directement la mobilité, sans la nécessité de la connaissance des dimensions du canal. Un modèle analytique pour la magnétorésistance est développé dans le cadre des nœuds technologiques sub 32 nm pour les modèles de transport balistique et quasi-balistique. La mesure de la magnétorésistance est explorée dans la région de saturation pour la première fois jusqu'à 50 nm sur les MOSFET « bulk » afin de comprendre l'applicabilité de cette méthode d'extraction à ce régime. Enfin les dispositifs bulk+ FDSON, FinFET, et GAA sont caractérisés en fonction de la température et les mécanismes de transport dans ces nouveaux dispositifs sont étudiés jusqu'à 35 nm (FinFET). En outre, le paramètre de champ effectif η est extrait pour les dispositifs sSOI. On trouve qu'il est différent du cas « bulk » comme c'était le cas pour les résultats obtenues sur bulk contraint et FDSOI. Cela est interprété par la rugosité de surface et la diffusion des phonons en raison de l'occupation préférentielle de la sous la bande fondamentale dans ces dispositifs avancés.

Characterization of Electron Transport in Short channel MOS Transistors

Abstract

Electron transport is one of the key properties that need to be improved in order to sustain performance improvement for the next technological nodes. Many factors, such as the choice of gate stack materials, channel material or the presence of mechanical strain contribute to degrade or improve transport properties. Body thickness, which reaches dimensions of a few nanometers, is playing a role as well, through interface scattering, thickness fluctuations or electrostatic and quantum coupling effects between front and back interfaces. In addition, it is strongly suspected that additional scattering mechanisms are associated with the proximity of the highly doped source and drain regions. For the sake of sub 32nm technology nodes development, it is of fundamental importance that these various mechanisms be identified and studied. In this range of dimensions, electron transport is governed by out of equilibrium, or even ballistic, phenomena. Therefore along with the advancement in the technology nodes, it becomes necessary to evolve the transport models and parameters to better explain the MOSFET operation. This thesis focuses on understanding the existing transport models and extraction methods and improving the same under the context of current and future technology nodes mainly sub 32nm. The MOSFET transport models and static parameter extraction methods in linear and saturation regime have been explored during the course of this thesis. The impact of gate voltage dependent series resistance in the advanced MOSFETs is taken into account and a new improved extraction method has being developed in the linear regime. Low temperature measurement is used in linear regime for the extraction of scattering mechanisms using mobility model. A new saturation drain current correction for short channel MOSFETs is developed for taking into account both DIBL and self-heating using low temperature measurement. Velocity saturation v_{sats} model and extraction method is explored in the saturation regime and v_{sats} is studied against temperature and channel lengths. Ballistic and quasi ballistic model with concept of kT layer in saturation regime is also studied for the sake of sub 32nm nodes. Channel magnetoresistance measurement offers promising prospects for short channel devices as we can directly extract the channel mobility without the need for the knowledge of channel dimensions. An analytical magnetoresistance model is developed in the context of sub 32nm technology nodes for full ballistic and quasi ballistic transport models. Magnetoresistance measurement is explored in the saturation region for the first time down to 50nm on bulk MOSFETs in order to understand the applicability of this extraction method in this regime. Finally Bulk+ FDSON, FinFET, and GAA devices are characterized with temperature and studied the transport mechanism in these novel devices down to 35nm (FinFET). Also effective field parameter η is extracted for sSOI devices and found that this is significantly different from bulk value as in the case of previous results in strained bulk and FDSOI devices and this is interpreted as increased surface roughness and phonon scattering due to preferential sub band occupation in these advanced devices.

Acknowledgement

I express my wholehearted gratitude to IMEP-LAHC, MINATEC - INPG, Grenoble, France for accepting my candidature for this PhD and provide its support and resources for the research work. I express my gratitude to ST Microelectronics, CEA LETI and IMEC for providing the samples for the characterization which enabled me to complete this thesis. I express my gratitude to LNCMI CNRS for allowing us to use its world class high magnetic field measurement facility to do the magnetoresistance measurement.

I would like to express my wholehearted gratitude to my research Director Dr. Mireille Mouis, IMEP-LAHC, for accepting my candidature and taking me as a PhD student under her supervision. I would like to thank my Co-Director Dr. Gérard Ghibaudo, IMEP-LAHC, for his kind guidance and teaching. I would like to thank reviewers and other jury members for sparing their valuable time for this thesis and for their valuable comments.

I would like to thank Dr. Florent Rochette (post Doc IMEP LAHC) for his guidance and support on FDSOI Measurements and magnetoresistance measurements.
I would like to thank Dr. Stéphane Monfray, ST Microelectronics for his support. I would like to thank Dr. Perrine Batude, CEA LETI as well as Mme. Xu Cuiqin, PhD student LETI for providing their wafer for characterization and their support. I would like to thank Dr. Gregory Bidal and Dr. Jean-Luc Huguenin for their support during this thesis.

I would like to thank Mr Xavier Mescot, IMEP LAHC for his guidance and training for the electrical characterization tools. I would like to thank Mme Martine Gri, IMEP-LAHC for her support for electrical characterization (wafer preparation and packaging). I would also like to thank Mr Olivier Drouin for his assistance in the lab for the characterization. I would like to thank Mr Duncan Maude, LNCMI - CNRS for his guidance and support for the magnetoresistance measurement.

I also thank other researchers and PhD students in IMEP for extending their kind support and friendship during my stay in Grenoble.

Table of Contents

1 Introduction 8
Reference 11
2 Electron Transport in MOS – Linear Regime 12
2.1 Drift Diffusion model – Effective Mobility 13
2.2 Scattering - Coulomb 19
2.3 Scattering – Surface Roughness 23
2.4 Scattering - Phonon 26
2.5 Ballistic/quasi ballistic transport 30
2.6 Conclusion 33
References 34
3 Electron Transport in MOS – Saturation Regime 37
3.1 High Field effect - Velocity Saturation 37
3.2 Velocity Saturation in MOSFETs 41
3.3 Velocity overshoot 45
3.4 Ballistic/quasi ballistic transport model 49
3.5 Conclusion 51
Reference 52
4 Experimental Extraction in Linear Regime 54
4.1 Extraction based on mobility model in linear regime 59
4.1.1 Y function and McLarty Method 59
4.1.2 R_{SD} for advanced CMOS devices – gate voltage dependence 63
4.1.3 New extraction method - Modified Y function for Advanced CMOS using $R_{SD}(V_g)$ 65
4.1.4 Impact of Θ_2 dependence with mobility and L_{eff} dependence with gate voltage 70
4.2 C-V method – Q_{inv}, L_{eff}, t_{ox} 74
4.2.1 Equivalent Oxide thickness t_{ox} 74
4.2.2 Effective Channel length L_{eff} 76
4.2.3 Effect of parasitic capacitance on C_{gc} for short channel lengths (~ 200nm and below) 78
4.3 Effective mobility from Split C-V and η parameter extraction 80
4.3.1 η - Extraction result on sSOI 83
4.4 Low temperature analysis of mobility 85
4.5 Magnetoresistance 87
4.6 MR in ballistic Transport 89
4.6.1 Quasi ballistic case 95
4.6.2 Limitations of ballistic MR model 97
4.6.3 Comparison with previous results 101
Conclusion and future prospective 104
Appendix A 105
Appendix B 109
Reference 113

5 Experimental Extraction in Saturation Regime 116
5.1 Drain Induced Barrier Lowering - DIBL 116
5.2 Self-Heating effect 119
5.3 Combining Self Heating and DIBL-new extraction and correction method for sub-µm MOSFETs ... 120
5.4 Saturation Velocity– v_{sats} 122
5.5 Magnetoresistance in Saturation Regime 124
Conclusion and future prospective 129
Appendix A 130
Reference 131
6 Results and Discussions 132
6.1 FDSON –Extraction results 133
6.2 Extraction on FinFet 136
6.3 Extraction on planar GAA 140
Conclusion 143
Reference 144
7 Conclusion and future prospective 146
8 Résumé en Français 148
8.1 Introduction 149
8.2 Transport électronique dans le transistor MOS – Régime linéaire 150
Références 152
8.3 Transport électronique dans le transistor MOS – Régime de saturation 153
Références 155
8.4 Extraction de paramètres en régime linéaire 156
8.4.1 Nouvelle méthodologie d'extraction - Fonction Y modifiée incorporant une dépendance $R_{SD}(V_g)$ pour les CMOS avancés 157
8.4.2 Paramètre η - Extraction pour des transistors sur SOI contraint (sSOI) 158
8.4.3 Etude de la mobilité à basse température 160
8.4.4 Magnetorésistance en régime balistique et quasi-balistique 161
Références 163
8.5 Extraction en régime de saturation 165
8.5.1 Effets combines de DIBL et d'auto-échauffement 165
8.5.2 Magnétorésistance en régime de saturation 166
Références 168
8.6 Résultats et discussion 169
8.6.1 FDSON – Résultats d'extraction 169
8.6.2 Extraction de paramètres pour des transistors FinFET 171
8.6.3 Extraction pour des transistors GAA planaires 172
Références 173

Reference Books

[Sze] S M Sze "Physics of Semiconductor Devices" 2nd Edition Wiley

[Bal] F.Balestra "Nanoscale CMOS Innovative Materials, Modeling and Characterization" Wiley

[Cri] S Cristoloveanu, "Silicon On Insulator Technology and Devices XI", ECS Proceedings Volume 2003-05

[Lund] M.Lundstrom, "Fundamentals of Carrier Transport", 2nd Edition, Cambridge University press

[Esseni] D.Esseni, P Palestri, L Selmi, "Nanoscale MOS Transistors Semi-Classical Transport and Applications", Cambridge University press

1 Introduction

As per ITRS (2009) road map [1], continuous device scaling would bring gate lengths down to 22nm around 2015. Main drive factors are faster and smaller devices with lower power. This means high I_{on} with low I_{off} and other leakage currents in general. Speed and low power requirements are however subject to trade off. There are several difficulties in using planar bulk/SOI MOSFETs for sub 32nm regime due to its issues related to I_{off} (DIBL, sub threshold slope etc) i.e. insufficient electrostatic control of channel by gate. Meanwhile the performance limitations, in terms of its drive current I_{on} and speed, lies on the other side of the spectrum of issues that has to be addressed.

In the case of planar bulk MOSFETs, issues related to I_{off}, SCE/DIBL etc were addressed by pocket implants, ultra shallow S/D, and LDD and worked well down to deep sub µm technology nodes. The mobility degradation caused by these modifications was addressed by applying strain, by using stressors, Ge implants, SiGe etc. Gate oxide is changed to High K/Metal gate to achieve 1nm range equivalent oxide thickness of SiO_2 needed for sub 100nm nodes while keeping the physical thickness of gate oxide to ~ 3nm and thereby controlling the problem of gate leakage and scalability. FD SOI devices were an alternative for the bulk devices owing to advantages of using thin silicon on SCE and sub-threshold and avoid using pocket implants thereby improving mobility. Also SOI gives proper isolation among the devices and thereby avoiding the latch up problems. The presence of BOX reduces substrate leakage currents and parasitic capacitance. However sub 32 nm node needs further improvements. These are partially addressed in devices such as UT2B or UTB^2 (ultra-thin body and Box) in the case of FDSOI and FDSON. Thin silicon film multiple gate devices like Double Gate, Triple gate, Pi Gate, Gate All Around GAA, FinFETs etc, that has more than one gate to control the channel region there by immunizing devices against SCE are alternative structures for sub 32nm nodes. Surrounding gate nano-wire is another solution in terms of low leakage and SCE/DIBL immunity [2-6].

Electron transport is one of the key properties that need to be improved in order to sustain performance improvement for every technological node. Many factors, such as the choice of gate stack materials, channel material or the presence of mechanical strain contribute to degrade or improve transport properties. Body thickness of thin Si devices reaches dimensions of a few nanometers, is playing a role as well, through interface scattering, thickness fluctuations or electrostatic and quantum coupling effects between front and back interfaces. In addition, it is strongly suspected that additional scattering mechanisms are associated with the proximity of the highly doped source and drain regions.

For the sake of future technology nodes, it is of fundamental importance that these various mechanisms be identified and studied. Electron transport in MOSFETs in nm scales is governed by out of equilibrium, or even ballistic, phenomena. It questions the validity of the physical parameters currently used to characterize transport as well as the validity of the experimental extraction methodologies. Therefore along with the advancement in the technology nodes, it becomes necessary to evolve the transport models and parameters to better explain the MOSFET operation. For example: ultra-thin film Si will have thickness fluctuation that contributes to scattering and introduces carrier confinement, introduction of High K, ballistic transport etc. This means that various parameter extraction methods also have to be evolved with new technology nodes. For example, extraction methods have to be improved in order to facilitate accurate parameter extraction by avoiding various parasitic effects (for example: gate voltage dependence of series resistance, fringing capacitances) whose impact becomes more and more significant as technology node advances.

This thesis focuses on understanding the existing transport models and extraction methods and improving the same under this context. The MOSFET transport models and static parameter extraction methods in linear and saturation regime have been explored during the course of this thesis. Bulk+ FDSON, FinFET, and GAA devices were characterized with temperature and studied the transport mechanism in these novel devices down to 35nm (FinFET). Channel magnetoresistance measurement offers promising prospects for short channel devices as we can directly extract the channel mobility without the need for the knowledge of channel dimensions. An analytical magnetoresistance mobility model is developed for sub 32nm technology nodes using full ballistic and quasi ballistic transport models. Magnetoresistance measurement was explored in the saturation region for the first time down to 50nm on bulk MOSFETs and studied the applicability of MR measurement in this regime.

The subsequent chapters in this thesis are organized as follows. Chapter 2 is dedicated to linear regime MOSFET operation. Different scattering mechanisms that govern the mobility identified so far in advanced MOSFETS are discussed in this chapter. Linear regime ballistic and quasi ballistic models that govern the transport especially in sub 32nm MOSFETs are also been discussed in this chapter.

Chapter 3 is dedicated to saturation regime MOSFET operation and models. The on current of the MOSFET is one of the main performance indicators. It is therefore important to study this region in its actual context apart from studying the linear regime parameters. This chapter covers high field effects in the saturation regime for the sub μm MOSFETs like v_{sat}, and velocity overshoot effect for deca-nanometer regime MOSFETs and its physical model. This chapter also discusses the ballistic and quasi ballistic model in saturation regime for advanced MOSFETs and the concept of kT layer in explaining the quasi ballistic transport in saturation.

Chapter 4 discusses different state of the art mobility empirical models and static extraction methods based on empirical models in the linear regime. The impact of gate voltage dependent series resistance in the advanced MOSFETs and a new improved static extraction method by taking into account this effect is also discussed in this chapter. Impact of higher order dependences on the new extraction method is also studied and discussed in this chapter. State of the art linear regime extraction method, split CV method, for the extraction of μ_{eff}, t_{ox}, L_{eff} and 'η' parameter for effective field is discussed in the chapter. η parameter, that determines the inversion charge contribution to effective field, is taken as 0.5 for bulk nMOSFETs. However in the case of strained bulk MOSFET and FDSOI devices, this is reported to be different. η extraction for sSOI devices is done for the first time and is discussed in this chapter. It is found that this is significantly different from bulk and even higher than sSi and FDSOI cases reported before. State of the art magnetoresistance measurement for extraction of magnetoresistance mobility μ_{MR} is discussed in this chapter. The magnetoresistance model needs to be improved for sub 32nm regime MOSFETs using ballistic/quasi ballistic transport models. Therefore an analytical model for magnetoresistance at low magnetic field for full ballistic and quasi ballistic regime is derived and compared the results with previous simulation results. This is discussed in this chapter. A new parameter extraction method using gain factor vs temperature based on state of the art low field mobility temperature model is also discussed in this chapter.

Chapter 5 is dedicated for state of the art empirical model and extraction methods in the saturation regime namely saturation velocity v_{sats}, and DIBL. Some of the state of the art extraction methods for extraction of v_{sat} and DIBL used in this thesis is discussed in this chapter. The extraction method in this regime suffers from both DIBL as well as self-heating effect in the case of deep sub µm devices. New method is developed for the extraction of DIBL parameter and thermal resistance using temperature measurements and is discussed in detail in this chapter. Magnetoresistance measurement is made on sub µm (down to 50nm) bulk nMOSFETS in this regime for the first time and is discussed in detail in this chapter.

Finally chapter 6 is dedicated to extraction results on advanced devices like bulk+ FDSON, FinFET and planar GAA devices. FDSON device is studied in both linear and saturation regime under different temperatures. FinFET and GAA are studied in linear regime and since their physical dimensions were not fully known, the technique using gain factor β vs temperature discussed in chapter 4 is used to obtain insight into the scattering mechanisms. The devices characterized so far down to 35nm (FinFET side wall) showed that transport is still governed by scattering and on current can be modeled by high field velocity saturation and velocity overshoot effects.

Reference

[1] www.itrs.net

[2] J P Colinge “Multi-gate SOI MOSFETs” Microelectronic Engineering, vol 84, p2071, 2007

[3] F Andrieu et al “Low Leakage and Low Variability Ultra-Thin Body and Burried Oxide UT2B SOI technology for 20nm Low Power CMOS and Beyond” IEEE Symposium on VLSI Technology Digest of Technical Papers, p57, 2010

[4] C Fenouillet-Beranger et al “Requirements for ultra-thin-film devices and new materials for the CMOS roadmaps” Solid State Electronics, vol48 p961 2004

[5] T C Chen, “Challenges for silicon technology scaling in the Nanoscale Era” IEEE ESSCIRC, p1-7, 2009

[6] Y Nara, “Scaling Challenges of MOSFET for 32nm Node and Beyond” IEEE VLSI, p72, 2009

2 Electron Transport in MOS – Linear Regime

This chapter is dedicated to understanding the mobility μ_{eff} that appears in the linear regime MOSFET drain current expression. The main advantage of using linear regime operation for MOSFET characterization is that it is a well understood regime in terms of its physics of operation and that extraction in this regime is simpler. Also the linear regime transport parameter, mobility, is correlated to the on current/saturation velocity and so MOSFET performance can be studied indirectly by understanding this regime. Different scattering mechanisms that govern the mobility identified so far in advanced MOSFETs are discussed in this chapter. Linear regime ballistic and quasi ballistic models that better models the transport in sub 32nm MOSFETs are also discussed in this chapter.

INTRODUCTION:

The electron transport in linear regime MOSFET operation is modeled by the term effective mobility μ_{eff}. This term is generally used as an indicator for the MOSFET performances, owing to better understanding of physics in this regime, convenience in terms of extraction. The linear regime MOSFET drain current is expressed as shown in Eq 2.1.

$$I_{DS} = \frac{W}{L} C_{ox} \mu_{eff}(V_g) \left(V_g - V_t - \frac{V_d}{2} \right) V_d \tag{2.1}$$

W, L and C_{ox} are channel width, channel length, and gate oxide capacitance per unit area respectively. V_t is the threshold voltage and V_d, V_g are the drain voltage and gate voltage respectively. The scaling driven technology generally refers to reducing L (which increases I_{on}, increases transistor density, as well as reduces load capacitance (gate capacitance of transistors of succeeding stages of the circuit)), W (counterproductive in terms of I_{on} but necessary for increasing the transistor density and also helps in reducing load capacitance) and t_{ox} (for sufficient gate control of the channel as well as increasing I_{on} but increased capacitance). As we see, reducing the channel length only have advantages in terms of transistor performance and density while understanding the impact of scaling of W and t_{ox} are not that straightforward. However sub µm lengths (especially ~ 100nm and below) require more than just scaling to push the transistor performance. The oxide cannot be scaled as very thin oxide thickness would mean large gate leakage current, while thicker oxide has lesser electrostatic gate control of channel and so short channel effects SCE/DIBL would cause increased off state current I_{off}. I_{on} is no

longer dependent on the channel length due to high field effects namely v_{sat} where physics of transport is different from the physics of low field mobility (discussed in section 3). However it is found that there is a correlation between v_{sat}/I_{on} and linear regime mobility μ_{eff} [1, 2] and engineering the channel to increase mobility (for example, by applying tensile strain) is a promising way to increase I_{on} without compromising the leakage issues.

Therefore it is important to study μ_{eff} and understand the underlying mechanism that determines this term. This chapter is dedicated to understanding the effective mobility in the linear regime of MOSFET operation.

2.1 Drift Diffusion model – Effective Mobility

As electrons travel in a semiconductor they get deflected due to several mechanisms and their path becomes random (known as scattering). Therefore transport can happen in two ways (we do not discuss the effect of temperature difference in the material in this thesis. See [Lund] and others for the same). Firstly, due to applied electric field electrons acquire a net velocity after scattering known as drift velocity v_d and drift in the direction opposite to electric field. Secondly, due to the electron concentration gradient due to which the carriers spread until a uniform distribution is reached known as diffusion. These can be described by Boltzmann transport equation through charge, momentum and energy balance equations (also known as hydrodynamic model) [Lund]. In this section however we talk about solution using most commonly used simplification of Boltzmann transport equation known as relaxation time approximation. The complete solution for momentum/energy balance equation is necessary at high fields and is discussed in section 3.

As electrons travel in the device, they encounter scattering due to various mechanisms. This is generally modeled, known as Drude model, as damping force acting against the electron flow and this damping force is proportional to the electron average velocity. The expression for Drude model is shown in eq 2.2.

$$\frac{d<v_d>_x}{dt} = -q\frac{E_x}{m_{eff}} - \frac{<v_d>}{\tau}$$
$$\therefore <v_d>_x = -\frac{q\tau}{m_{eff}}E_x \qquad (2.2)$$

The term $<v_d>_x$ is the average drift velocity of electrons in the x direction, $-qE_x$ is the force acting on the electrons, where q is the electron charge absolute value, m_{eff} is the effective mass of the electrons in the transport plane and τ is the characteristic mean free time between collisions. The term $1/\tau$ is the scattering rate (number of collisions per sec). The term $q\tau/m_{eff}$ is called the mobility μ_{eff}. This is a simplified version of the momentum balance equation by using relaxation time approximation or the

mean free time (which has different values depending on the scattering entity) for Boltzmann transport equation [Lund] (but neglecting the diffusion term). The mean free time between the collisions does not really reflect the mobility all the times (unless the scattering is isotropic i.e. equal probability for all electron scattering directions) but mean free time for momentum relaxation (randomization), known as momentum relaxation time τ_m and energy relaxation time τ_E (electron loses energy to lattice) have to be used. These two relaxation times are larger than the mean free time between collisions in the order $\tau_E > \tau_m \geq \tau$. This means that even through there was a collision it does not change the electron trajectory significantly and such scattering is anisotropic. The momentum relaxation time does not involve change in the electron energies but only in trajectory that would affect the current. In the case of scattering due to lattice vibrations also known as phonon scattering, electron loses/gains some energy. In case of small exchanges, a fraction of kT_L, the scattering is still considered as elastic scattering. The acoustic phonon scattering is considered elastic as the energy exchange is very small except at low temperatures. The mobility term is only valid at low lateral electric fields (so low V_d). The mobility expression then uses τ_m instead of τ. The impact of energy relaxation (τ_E) is discussed in section 3 and for low field condition, the energy equation is not taken into account as energy gained by carriers from low electric field is very small.

In the case of diffusion transport, the current is proportional to the carrier concentration gradient along the device (known as Fick's Law) and is characterized by a parameter known as diffusion coefficient, D_n and D_p (for electrons and holes respectively). This is shown in Eq 2.3. Einstein showed that this diffusion coefficient can be related to mobility, as shown in eq 2.3, also known as Einstein relation.

$$J_{nx} = qD_n \frac{dn(x)}{dx};\ D_n = \frac{kT}{q}\mu_n \tag{2.3}$$

Ficks law is a simplified expression for the component "momentum flux gradient" in momentum balance equation under the conditions that the electron temperature gradient negligible (it is taken to be same as lattice temperature for small fields) and electron drift velocity is constant under small uniform field (or gradient of drift velocity along channel is negligible). So momentum flux gradient is just due to the carrier density gradient. Complete momentum balance equation from the Boltzmann transport equation includes both the drift model as well as Fick's law. This gives total current as sum of drift current and diffusion current as seen in literatures (as well as thermal in case of temperature gradient) [Sze, Lund]. The electric field in the drift part appears as the momentum generation term in the momentum balance equation. The mobility value of electrons in intrinsic or low doped ($<10^{16}cm^{-3}$) Si is 1450 $cm^2/V.s$. and for holes it is 450 $cm^2/V.s$ at 300K [Sze]. This mobility rolls off with increasing doping concentration due to Coulomb scattering and with increasing temperatures due to phonon scattering.

MOBILITY IN MOSFET CONTEXT:

The mobility of carriers in MOSFETs is different from the bulk Si mobility. This is because the transport in MOSFETs is due to 2D surface carriers at Si/SiO_2 interface (generally few nm deep) unlike in bulk silicon and therefore the mobility is surface carrier mobility. The surface mobility ((100) surface) is normally smaller at room temperature than bulk value due to additional scattering. Also in the case of MOSFETs the effective mobility is an average value, averaged over the channel area as well as the inversion depth. Assuming uniform charge distribution along the channel area (which is true only for very low V_d) and uniform mobility in the plane of surface carriers, we can average the mobility along the inversion layer depth as shown in Eq 2.4.

$$\mu_{eff} = \frac{\int \mu(z) n(z) dz}{\int n(z) dz} \tag{2.4}$$

The term $\mu(z)$ and $n(z)$ are mobility and inversion charge density as a function of depth z assuming they are uniform in x-y plane. The denominator would give the inversion charge per area Q_{inv} in cm^{-2} and in the case of low V_d this is given by C_{ox} $(V_g - V_t)$. In the case where the inversion layer is not uniform then we have to average also in the plane of inversion layer (especially along the channel length). This is the case of higher V_d (but low field) values as well as scattering due to neutral defects, pocket implants etc which are dominant near S/D region. The mobility can be expressed in terms of mobility due to different scattering mechanisms as shown in Eq 2.5a. The subscript 'dd' is mentioned to stress the drift diffusion transport.

$$\frac{1}{\mu_{effdd}} = \frac{1}{\mu_C} + \frac{1}{\mu_{SR}} + \frac{1}{\mu_{ph}} + \frac{1}{\mu_N} \tag{2.5a}$$

This is known as Mathiessen rule named after Augustus Mathiessen who has extensively worked on conductivity of metals and alloys [eg: ref 48]. The terms appearing in these expressions (Eq 2.5) are different mobility terms identified so far for semiconductors namely, μ_C Coulomb scattering mobility, μ_{ph} phonon scattering, μ_{SR} is surface roughness (SR) scattering and μ_N is neutral defect scattering. Even though this may not hold for all the cases at temperatures above 0K, this method provides a simple analytical tool to separate different scattering mechanisms. This is because the mean free time is energy dependent and it is averaged over energy to calculate the mobility. Each of the scattering mechanisms has different energy dependence and can be approximated by a power law [Lund]. This is shown in Eq 2.5b. The term f_0 is the Fermi-Dirac function, k is the Boltzmann constant and T_L is the lattice temperature (for low field). When several scattering mechanisms are present, it is the scattering rates $(1/\tau)$ that add up. Mathiessen rule generalizes this to energy averaged mean free time. But at 0K the energy averaging would simply yield $\tau(E_F)$ and so the Mathiessen rule applies whatever be the energy dependence. Multi sub band occupation and anisotropic scattering further degrades the accuracy of applying this rule. [11, 39, Lund, Esseni]

$$\frac{1}{\tau_{total}(E)} = \frac{1}{\tau_C(E)} + \frac{1}{\tau_{ph}(E)} + \frac{1}{\tau_{SR}(E)} .....$$

$$\tau_{eff} = \frac{< E\tau_{total}(E) >}{< E >} = \frac{\int E\tau_{total}(E)\left(-\frac{df_0(E)}{dE}\right)dE}{\int_C E\left(-\frac{df_0(E)}{dE}\right)dE} \qquad (2.5b)$$

$$\tau\ (E) = \tau_0\left(\frac{E}{kT_L}\right)^s \ ; f_0(E) = \frac{1}{1+e^{\frac{E-E_F}{kT}}} \approx e^{-\frac{E-E_F}{kT}} \ (non\ \deg enerate\ Boltzmann)$$

Several extraction methods for mobility μ_{eff} are discussed in the section 4. The effective mobility is then studied against inversion charge Q_{inv}/ transverse electric field, temperature, doping concentration, oxide charges/ interface charges and by making use of Eq 2.5 different scattering mechanisms are separated (for example see section 6.1). Figure 2.1 shows how effective mobility varies with effective field E_{eff} (defined as average transverse electric field in the inversion layer [7]), temperature and doping concentration [6]. The surface roughness SR becomes dominating mechanism at high electric fields where all the mobility curves for different doping merge to a universal trend against electric field. The mobility curve at low doping at 300K is fully due to phonon and SR. The doping modifies this universal curve due to Coulomb scattering which is pronounced at lower electric field (Coulomb mobility decreases with increase of doping concentration). Also doping increases E_{eff} for same Q_{inv} and cause mobility to degrade. At low temperatures the phonon scattering is reduced and scattering is dominated by Coulomb and SR. Different mobility components therefore have different behavior with inversion charge/ Electric field and temperature. The use of the term E_{eff} using triangular potential well approximation works very well for bulk MOSFETs. This is however not the case for FDSOI, DG, FinFets etc. This concept of universal mobility has been questioned when referring to these advanced MOSFET structures [49].

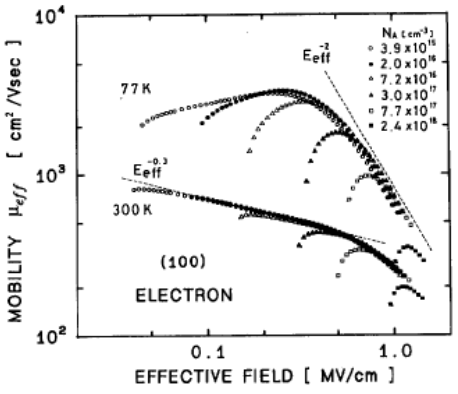


Figure 2.1

Effective mobility vs Effective Field at 300K and 77K and different doping concentration Takagi et al [6].

SHORT CHANNEL MOSFET MOBILITY:

In the case of short channel MOSFETs new scattering mechanisms like neutral defects are identified and found to be significant contributor for the channel mobility [3, 4]. For MOSFET channel length especially ~ 100nm and below, the mobility is highly degraded. This phenomenon is common for all types of transistors, bulk, FDSOI, FinFet, GAA, DG and has been reported by many as well as results presented in this thesis in later chapters. This was at first blamed on the pocket implants or dopant pile up causing an increase in the electric field near Source/Drain (S/D) region as well as increased Coulomb scattering (near S/D) thereby reducing the mobility. As channel length reduces, the region of pocket implants becomes more significant and therefore mobility roles off. But studies in ref [4] have reported additional scattering mechanism which is independent of temperature. This mechanism has negligible impact on long channel devices (>100nm), while considerably reducing the channel mobility for short channel devices. The presence of crystalline defects near S/D has been considered to be the cause and it has been reported in [4] that this can be somewhat improved by spike anneal at 1080^{o}C. This local 'neutral' defect mobility $\mu_N(x)$ is therefore channel length dependent and dominant near S/D. The overall neutral defect mobility μ_N is the average $<\mu_N(x)>_x$ along the channel length and therefore, as channel length reduces, the contribution of the defects near S/D increases reducing the overall mobility. The presence of interface state defects around the S/D extension could introduce a channel length dependent Coulomb mobility μ_C however this will have temperature dependence and also can introduce a reverse short channel effect RSCE (increase of threshold voltage with reducing channel lengths) [4]. For advanced FDSOI MOSFETs with ultra-thin Si film, channel mobility is found to be affected by thickness fluctuation. Remote Coulomb scattering due to S/D implants, High K as well as remote phonon scattering due to high K further degrades mobility (discussed in the coming sections). For very short channel lengths, especially sub 32nm it is expected that part of the transport is not governed by scattering and said to be in the quasi ballistic regime. This also is modeled as a virtual mobility term μ_{bal} (Shur model [5]) and added to the drift diffusion mobility using the Mathiessen rule [4, 5] as shown in Eq 2.5c. μ_{eff} is the apparent mobility and generally stands for the mobility that is measured.

$$\frac{1}{\mu_{eff}} = \frac{1}{\mu_{effdd}} + \frac{1}{\mu_{bal}} \qquad (2.5c)$$

For devices 10nm and below it is expected that the transport will be full ballistic and therefore the I_{ON} is no longer governed by channel length. It is therefore important to study different scattering mechanisms in order to identify possibilities to improve MOSFET performances. For example the presence of strain in the channel improves the on current by increasing the channel mobility without compromising the leakage problems.

SUB-BANDS AND SUB-BAND ENGINEERING:

Electrons at the interface are confined in the direction normal to the surface (here shown as z-direction) and conduction band splits into several sub bands. Six conduction band valleys of bulk Si in 3D has to be translated to 2D surface. Referring to (100) surface, electrons with a heavy longitudinal effective mass m_l in the z-direction have the lowest confinement energy. There are two such valleys one in +z and one in –z direction and therefore called as twofold valleys in general. The other 4 valleys (referred to as fourfold valleys in literature) in x-y transport plane have smaller transverse effective mass m_t in the z direction and have their sub band energies higher than the sub bands of the two fold valleys. These sub bands are sometimes referred to as unprimed (two fold) and primed (four fold) and the sub band levels are numbered as 0, 1, 2 etc for unprimed twofold valleys and 0', 1' etc for primed fourfold valleys. This is shown in fig 2.2.

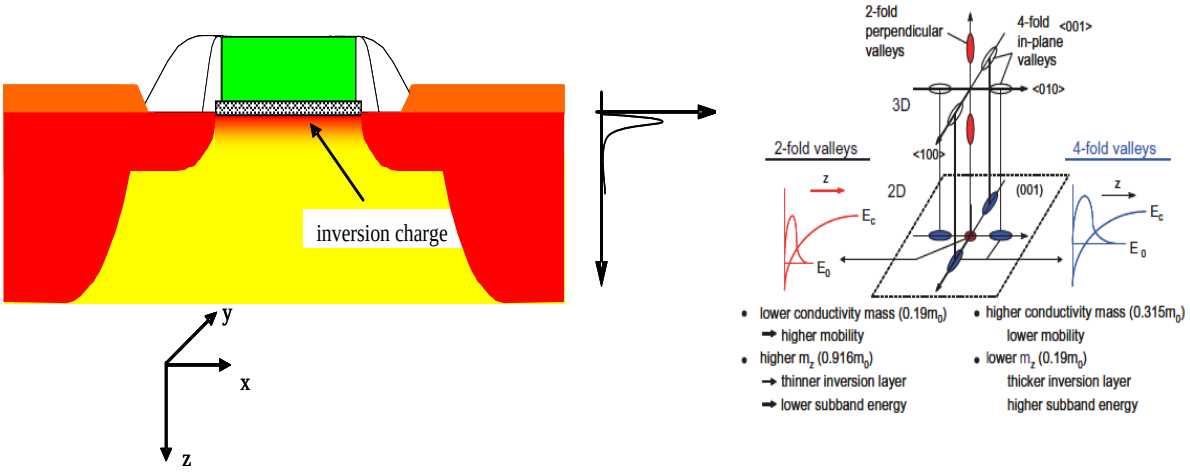


Figure 2.2

The MOSFET structure with inversion charge is shown on the left. On the right: Image taken from Takagi et al [25] showing the valleys in bulk and corresponding valleys for 2D inversion layers. The figure shows the electron wave function extension in two and four fold valleys in the inversion layer. In many models of inversion channel, the electron wave function is approximated by Airy functions [7]. The twofold valley electrons are more confined to the interface than the fourfold valley electrons.

The sub band occupation of electrons is very important as it determines the effective mass in the mobility term as well as scattering rates due to screening [15]. Scattering of electrons between sub bands known as inter sub band scattering is said to reduce the mobility [15]. This is discussed in brief in sections below for different scattering mechanisms. Fig 2.2 on the right side shows the 6 valleys of the silicon in 3D translated to two fold valley (unprimed) and four fold valley (primed) in the inversion layer plane. The electron wave function in the inversion layer is approximated to first order by Airy's functions [8] in the z-direction. In the transport plane the electron wave function is taken as traveling

waves for planar MOSFETs with 2D density of states. For 2 fold valleys (red) the confinement effective mass is m_l which is $0.91m_0$ and is the highest and therefore has the lowest energy and at the same time transport mass is m_t, which is $0.19m_0$, which is the lowest and therefore has higher mobility. This is made use in so called sub band engineering by using strain in the channel or in case of FDSOI, by using ultrathin silicon to increase the carrier confinement (Si thickness ~ 5nm and less).

Takagi et al. [25] and others in references therein have showed that tensile strain in the silicon increases the energy difference between the sub bands and valleys and forces more electrons to occupy the lowest sub band of the twofold valleys that has smallest effective mass m_t in the transport plane for (100) Si. This will improve the electron mobility for the strain engineered devices. Also higher energy difference between valleys/sub bands would suppress inter valley/sub band scattering. It is shown to be effective in improving I_{on} even in the case of full ballistic transport [25]. Therefore in the case of advanced MOSFETs (sub 100nm), channel strain (tensile or compressive for pMOS) is one of the ways of overcoming the transistor I_{on} limitations without compromising I_{off}.

For the calculation methods of scattering rates, refer to articles like [Lund, Esseni], [8] and others. I have explored the experimental aspects and therefore discussions in this thesis are more focused on empirical formulas and analytical/qualitative methods that facilitate experimental extraction of mobility and inference of transport using the same. Important scattering mechanisms like Coulomb, surface roughness and phonon are discussed in more detail in the below sections and ballistic limit and quasi ballistic transport for short channel MOSFET are also discussed briefly in section 2.5.

2.2 Scattering - Coulomb

Coulomb scattering in bulk Si is attributed to dopants. This is dependent on dopant concentration and temperature as given in the Eq 2.6 [Sze].

$$\mu_{Cb} \propto N_I^{-1} T^{3/2} \tag{2.6}$$

In the case of MOSFETs, Coulomb scattering arise from several factors. The most important being the interface charges and oxide charges. This is because these Coulomb centers are very close to the 2D inversion layer. The impact of channel doping is shown in the fig 2.1 above. Normally doping concentration of ~ 10^{16} cm^{-3} and below can be neglected and Coulomb scattering can be attributed to oxide charges [8-15]. Apart from this, remote charge/Coulomb scattering RCS due to poly depletion is also being blamed for reduced mobility in short channel devices with thin oxides that experiences high electric fields. For very short channel devices the depletion charges from S/D is also suspected to take

part in RCS [3]. Advanced MOSFETs use high K gate stack in order to circumvent the issue of gate leakage in thin oxides ~ 2nm EOT (effective oxide thickness) and below. Thinning of oxide for advanced MOSFETs is necessary to boost I_{on} as well as to ensure good channel control (reducing DIBL/SCE and therefore reducing I_{off}) by gate. Mobility of devices with High K gate stack is consistently smaller than those with SiO_2 gate, and are attributed to Coulomb scattering (as well as soft optical phonon scattering which is reduced by TiN metal gate deposition discussed in section 2.4 [53]). Therefore Coulomb scattering mechanism needs to be studied properly especially for short channel, High K gate devices.

In the case of classical long channel bulk MOSFETs with thick SiO_2 / Poly Si gate, the main Coulomb centers held responsible are oxide charges and interface state charges. Low channel doping is neglected. In the absence of carrier screening at high temperatures, the Coulomb limited mobility is given by Sah [16] as shown in Eq 2.7 assuming that N_{ox} lies at the interface. Carriers with large kinetic energy are less deflected by the charge centers and therefore with increase of temperature the Coulomb scattering should reduce. Also at higher electric fields where electrons attain higher kinetic energy the Coulomb scattering would reduce. It is important to note that Coulomb scattering only involves momentum relaxation (randomization) as there is no exchange of electron energies with scattering centers. Therefore Coulomb scattering is elastic.

$$\mu_C \propto T / N_{ox} \tag{2.7}$$

The influence of screening effect is studied by Stern [11, 18], Shiue and Sah [17] and measured by Hartstien [9], Takagi [6] and others. Screening effect is generally observed for low temperatures <100K and low inversion [11, 9]. And it has been shown that in this region the temperature dependence is different. This has been attributed to the temperature dependence of screening effect at low inversion (screening is dependent on wave vector [11]), and this effect dominates the reduced Coulomb scattering rate of high energy carriers. However, above 100K the expression of eq 2.7 could be used with confidence [8]. The low temperature <100K, low inversion Coulomb scattering is empirically modeled as eq 2.8 [8-13]. Formulation by Stern [11] gives Coulomb scattering at low temperature as $N_{inv}^{1.9}$.

$$\mu_C \propto N_{inv}^{\alpha} / TN_{ox}; \alpha \approx 1.6 - 2 \tag{2.8}$$

In the case of sub μm bulk MOSFETs channel doping is higher or halo implants are made near S/D in order to tackle the problem of DIBL, punch through etc and ensure good gate control. Therefore scattering is also influenced by the channel doping due Coulomb and higher electric fields as discussed previously. This therefore is a tradeoff between DIBL/SCE (I_{off}) and I_{on}/mobility in the case of bulk

MOSFETs. In the case of FDSOI, DIBL/SCE control is achieved by thinning the Si and therefore channel is un-doped, increasing the mobility. Therefore, in FDSOI MOSFETs, Coulomb scattering comes mainly from the N_{ox}/ interface states, HK etc and so is the case of DG, FinFets where the Si is un-doped. In general we still use eq2.7 for temperatures above 100K in order to de-correlate the Coulomb scattering mobility from overall mobility value using temperature dependence.

REMOTE COULOMB SCATTERING (RCS) AND HIGH K/METAL GATE:

In the case of thin gate oxide, poly depletion is also said to take part in remote Coulomb scattering as mentioned previously. For gate dielectric less than 1.5nm, the poly depletion is said to reduce mobility by 20-30% [19]. This effect is screened at high inversion. Its impact is somewhat smaller for holes. However due to high gate leakage, physical oxide thickness scaling is limited to ~ 3nm. For MOSFETs requiring ~ 2nm or less, the gate stack is changed to high K (HfO_2, HfSiON) in order to avoid the scaling of the physical thickness of overall dielectric. Also poly depletion is no longer a problem as metal gate electrodes are used (TiN, TaN). Therefore this type of poly depletion RCS is no longer present. However high K gate stack has consistently showed reduced mobility and one of the reasons for this is shown to be RCS. The study of the effect of HfO_2 thickness and SiO_x IL thickness [22] has shown that the charges contributing to RCS lie at HfO_2/SiO_2 interface. Measurement using magnetoresistance by L Thevenod et al [20] showed excess Coulomb scattering and is attributed to RCS by charges at high K/SiO_2 interface and charges in the high K itself. The mobility is shown to decrease with reducing the thickness of SiO_x interface layer IL due to the proximity of high K/SiO_2 interface to the channel [20, 21, 22]. Apart from this, metal gate also has been shown to have impact on the mobility by itself [22]. It has been shown that the high field mobility of metal gate is improved and this has been attributed to Coulomb “shielding” of the surface roughness by the so called border traps close to the Si/SiO_2 interface (with in 3nm inside the oxide or rather within the range where these traps could communicate with Si) [22, 23, 24]. This has the impact of reducing the transconductance gm at low inversion due to charge trapping that causes reduction of inversion charge and increasing gm at high inversion due to roughness “smoothening”. Mobility measurement reported by W Zhu et al [27] for ultra-thin HfO_2 metal gate stack ~ 2nm EOT shows Coulomb scattering due to interface state traps in these devices. Therefore interface trap on High K devices also has to be taken into account at low inversion region. (The report shown in ref [22] however neglects the interface state trap mobility as it was small). Another reason for mobility degradation from High K is soft optical phonon scattering and it is discussed in section 2.4.

MULTI SUB-BAND AND HIGHER ORDER EFFECTS:

In the case of multi sub band occupation, which is more pronounced for inversion charges ~ 10^{13} cm^{-2} and above, the mobility is shown to be degraded due to inter sub band scattering compared to single sub band assumption [15]. Since Coulomb scattering is studied at low inversion cases, this effect is not considered.

The higher order effects like many body effects, occupation 0' sub band due to its low energy tails due to level broadening, failure of linear screening approximation (for calculating the low temperature behavior of Coulomb scattering) like bound states due to attractive potentials etc change this simple picture. For example, many body effects is said to change the energy difference and therefore electron occupation of the sub bands. In this case this is advantageous as this has the effect of increasing the band separation. But it also has the effect of reducing the electron mobility [7, 15]. In the case of bound states, at liquid He range of temperatures, we will have carrier freeze out effect and account for strong temperature dependence at low inversion i.e. conductance is thermally activated in this region. Also the mobility extracted against inversion charge will show a steep increase with increasing inversion carriers if this carrier freezing effect is not taken into account in the extraction [10]. Carrier freeze out is taken as non-significant in the temperature range used for the study in this thesis (>77K and especially at room temperatures). These higher order effects are out of the scope of this thesis and so not discussed further. Refer to [7] and others for more details.

EFFECT OF STRAIN:

In the case of advanced MOSFETs for sub 100nm nodes, strain engineering is one of the main solutions for improving I_{on} [25] without compromising low I_{off} requirement. Studies made by Weber and Takagi [26] explain the impact of strain on the Coulomb scattering due to substrate impurity N_{sub} and N_{it} or oxide charges. They have found that strain has different effects on the scattering by these different scattering centers. The biaxial tensile strain has the effect of increasing the mobility due to the substrate doping (channel doping and halo implants) while it degrades the mobility due to the interface states. This study is contradiction to the previous studies [26 and ref there in] that had reported that strain does not improve the Coulomb mobility due to dopants. This would mean that for sub 100nm devices with halo implants, whose mobility is mainly degraded by channel doping and neutral defects, there will be no significant advantage for strain engineering. This finding is therefore important for advanced MOSFETs and show that strain engineering is indeed effective. The interface state Coulomb mobility degradation and dopant Coulomb mobility enhancement due to biaxial tensile strain in the channel is being attributed to the valley occupation, where strain causes the energy gap between the 4 fold valleys and 2 fold valleys to increase. This would mean higher electron occupation

in the 2 fold valleys. The 2 fold valley has lowest effective mass and therefore higher mobility. However in the case of interface states, the 2 fold valley occupation also increases the scattering rate as electrons are confined nearer to the interface compared to the 4 fold valley and therefore negate the effective mass advantage obtained for mobility [25].

2.3 Scattering – Surface Roughness

While scattering at low transverse field is dominated by impurities and oxide/interface charges, the scattering at very high fields is dominated by the Si/SiO_2 interface non uniformity or surface roughness. As 2D electron gas is confined closer to the interface, they come under the influence of the interface roughness. See fig 2.3. The effective mobility will then roll off rapidly with the gate voltage or electric field. This is seen in the fig 2.1 where at high electric fields, mobility rolls off as E_{eff}^{-2}. The mobility curves for all doping levels converge to same trend at high electric field showing that the scattering mechanism is common and universal in this region. Hartstien et.al [9, 14] studied the temperature and oxide charge dependence of mobility and extracted SR and Impurity mobility component by varying the oxide charges at 4K where the phonon scattering is suppressed. They have showed that the surface roughness mobility dependence can be better fitted as in Eq 2.9 unlike with the weighting factor ½ for Q_{inv}.

$$\mu_{SR} \propto \left(Q_D + Q_{inv}\right)^{-2} \tag{2.9}$$

This mobility term is also shown to be independent of temperature to first order. (Higher order effects like inter sub band scattering could increase as temperature rises and so apparent temperature dependence has to be expected for SR too, normally reducing the SR mobility with increasing temperature). Theoretical derivation of the SR mobility can be found from [7, 28, 29]. The SR mobility calculation shows that the scattering rate can be expressed in terms of power spectral density of the SR (by using Fourier transform for the scattering potential). If the random displacement of the surface is modeled with a known auto correlation function like Gaussian function as in Eq 2.10, then we can relate the scattering rate with two physical quantities like rms value of the interface displacement Δ and correlation length Λ. In the transition/scattering rate expression [Lund, 7 etc], vector **q** will be the difference between wave vector before and after the scattering denoted generally in literature as **k** and **k**' respectively. In the case of exponential auto correlation function given by Goodnick et al [30], (in which they have measured the SR by TEM analysis and generated the autocorrelation function that was best fitted by exponential rather than Gaussian) the expression for the power spectral density gets modified as Eq 2.11. This type of measurement using TEM etc may or may not represent the actual SR in a MOSFET as sample preparation can alter the surface asperities.

Also these measurements only yield the surface morphology and not the actual potential fluctuation necessary to calculate the mobility.

$$< \Delta(\vec{r})\Delta(\vec{r}\,') > = \Delta^2 e^{-\left(\frac{\vec{r}-\vec{r}'}{\Lambda}\right)^2}$$

$$|\Delta(\vec{q})|^2 = \pi\Delta^2\Lambda^2 e^{-\frac{q^2\Lambda^2}{4}} \qquad (2.10)$$

$$< \Delta(\vec{r})\Delta(\vec{r}\,') > = \Delta^2 e^{-\left(\frac{\sqrt{2}|\vec{r}-\vec{r}'|}{\Lambda}\right)}$$

$$|\Delta(\vec{q})|^2 = \frac{\pi\Delta^2\Lambda^2}{\left[1+\frac{q^2\Lambda^2}{2}\right]^{3/2}} \qquad (2.11)$$

Important thing to notice here is that in either case the scattering rate is proportional to $\Delta^2\Lambda^2$. These two parameters can be fitted or estimated from the TEM as in ref [30, 31]. The typical values are $\Delta \sim 0.4$nm and $\Lambda \sim 1.5$nm.

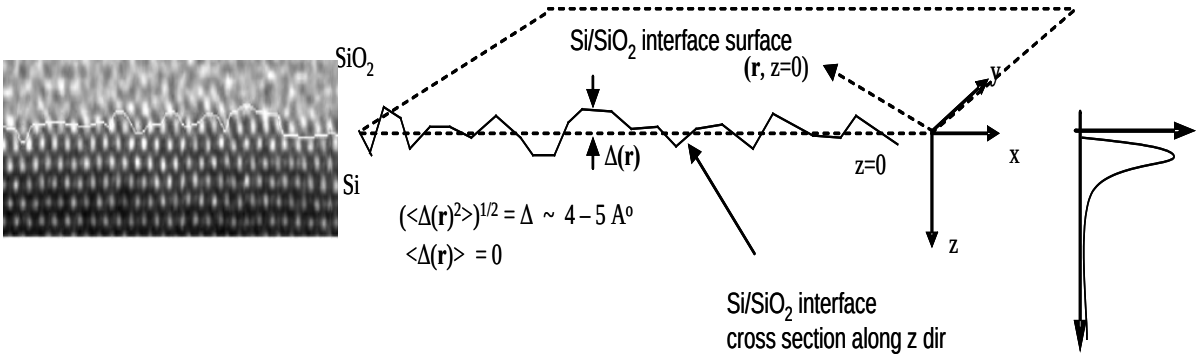


Figure 2.3

TEM image (left) and representation of surface roughness (middle). TEM image is taken from Zhao et al [31]. x-y plane is parallel to the interface and roughness is represented as the displacement of the interface in the z-direction. RMS value Δ lies typically in the range of 4-5A^0. Mean displacement of interface is taken as z=0. Inversion charge distribution is also shown in the right end.

The scattering rate is then given by Eq 2.12 which is supported by the measurement results in [14, 9, 7]. All these derivations given in literatures assume triangular potential well and therefore constant electric field called effective field E_{eff} which is proportional to (Q_D+0.5Q_{inv}) [7]. The factor 0.5 for the Q_{inv} does not appear in the findings of ref [14, 7] for surface roughness dependence. The effective mobility plotted against the E_{eff} is however universal for bulk MOSFETs and varies as E_{eff}^{-2} at high

fields. Takagi et al (referring to fig 2.1) have fitted the field dependence of μ_{SR} as $E_{eff}^{-2.6}$. Several literatures also talk about the SR in terms of E_{eff}^{-2}. However, considering the measurement results of Hartstien et al [14] eq 2.12/2.9 is justified. Also at high inversion E_{eff} α Q_{inv} and so the SR mobility dependence with Q_{inv} (>> Q_D at high inversion and moderate doping ~ 10^{11} cm^{-2}) can be approximated to vary as E_{eff}^{-2}.

$$\frac{1}{\tau_{SR}} \propto \Delta^2 \Lambda^2 (Q_D + Q_{inv})^2 \qquad (2.12)$$

MULTI SUB-BAND AND STRAIN:

The expressions shown previously are based on single sub band occupation assumption. By taking into account higher sub bands, because of inter sub band scattering, μ_{SR} like Coulomb is shown to get degraded [15]. However, higher sub band occupation as mentioned before will become significant when inversion charge is close to $10^{13}cm^{-2}$ and at high temperatures. The single sub band model discussed above is therefore more and more accurate as we lower the temperature and confine our analysis to inversion charges up to high $10^{12}cm^{-2}$. Nonetheless this model is very useful in the experimental mobility analysis. The presence of biaxially tensile strain has been shown to improve the SR mobility for electrons as reported by Zhao et al and Bonno et al [31, 32] and others. They have shown that SR mobility of electrons is improved and both of them have attributed this effect to the reduction of the SR amplitude Δ. The strain effect for the holes SR mobility has been reported by Zhao et al [31] and has been found to have a reduction of mobility for low strain and then for strain corresponding to 40% Ge on SiGe substrate and above, the strain has no influence on the hole-mobility. The reason for this has been given as reduction in both Δ and Λ or due to non-Gaussian correlation function. The impact of the reduction of Λ is different for electrons and holes due to their k_F differences [see ref 14, 33]. Either of these interpretations is however not fully validated.

HOLE SR MOBILITY AND SR AUTO CORRELATION FUNCTION:

The hole-mobility has a different E_{eff} dependence and it has been shown in literatures to vary as E_{eff}^{-1} [6, 33]. The effective field for the holes is also different from electrons and is given as (Q_D+0.33Q_{inv}). Pirovano et al. [33] have shown that for holes, the field dependency can be explained by changing the spectrum shape without compromising on the electron SR theory. Since the SR is same for n and p MOSFETs, the spectral density is modified to explain the field dependence of both electron and hole-motilities using the same function. They proposed a new spectral density function as Eq 2.13 and were able to explain the field dependence of both electrons and holes as E_{eff}^{-2} (αQ_{inv}^{-2}) and E_{eff}^{-1} respectively without changing the parameters Δ and Λ.

$$\left|\Delta(\vec{q})\right|^2 = \pi\Delta^2\Lambda^2 e^{-\frac{q^4\Lambda^4}{4}} \tag{2.13}$$

A gate voltage dependent mobility taking into account SR is discussed in section 5. In this case however the carrier distribution is taken to be exponential and the mobility is derived in terms of the constant field E_{eff} and Δ. This derivation is simpler and fully analytical and tries to explain the second order mobility attenuation parameter needed for explaining the negative transconductance seen at high gate voltages [34].

In the case of advanced MOSFET structures like FDSOI, FDSON, GAA, DG, FinFets etc the silicon has more than one oxide interfaces and the SR in this case manifests itself as thickness fluctuation. This becomes significant for t_{Si} ~ 5nm and less and for FinFets, fin width ~ 5nm and less (as the rms Δ ~ 0.5nm) [25, 57]. In this region the mobility is shown to increase as T_{Si}^6. Above thickness of around 5nm this effect vanishes and the SR mobility applies. In the case of DG or ultra-thin FDSOI (t_{Si} ~ 3nm) in DG operation, carriers are located within the Si (volume inversion) and this is said to improve the mobility [57].

2.4 Scattering - Phonon

The phonon limited mobility is the dominant mobility limiting mechanism at room temperature at low and medium fields. As electron travel in the material, they get disturbed by the lattice vibration and get scattered and even loses/gains energy during this interaction. Elastic waves propagate through the lattice due to oscillation of atoms about its equilibrium site. As these waves propagate similar to Bloch waves through the crystal lattice, like electrons, their energy gets quantized and these quanta of energy are perceived as phonons. There are 6 types of modes of lattice vibration or phonons. There are three acoustic phonons or modes of atomic vibrations where adjacent atoms displace in the same direction (similar to sound waves) but at different magnitudes (The dispersion relation for small wave vector is related to sound velocity v_s as $\omega(\beta) = v_s\beta$ i.e. group velocity of acoustic phonon for low β is same as sound velocity) and three optical phonons where the adjacent atoms vibrated in opposite direction (said to be able to interact with light. Here $\omega(\beta) \approx \omega_0$ (simplified) or no dispersion and has non zero frequency ω_0 at Brillouin zone center or at $\beta = 0$). There is one longitudinal acoustic phonon in which atoms displace in the same direction as wave propagation and two transverse mode of acoustic phonon where atom displacement is transverse to the wave propagation abbreviated as LA and TA respectively (similarly for optical its LO and TO). The number of phonons N_β is related to the temperature through Bose-Einstein Factor as shown in the Eq 2.14. The term $\omega(\beta)$ is the dispersion relation, T_L is the lattice temperature, k is Boltzmann constant and $\hbar$ is the reduced Plank's constant. [Lund]

$$N_\beta = \frac{1}{e^{\frac{\hbar\omega(\beta)}{kT_L}} - 1} \approx \frac{kT_L}{\hbar\omega(\beta)};\ \hbar\omega(\beta) << kT_L \tag{2.14}$$

The carrier energy change due to acoustic phonon scattering is very small ~ 1 meV (<< kT ~ 26meV at room temperature) [Lund] and generally modeled as elastic scattering with momentum relaxation time. Also we need to consider the conservation of energy and conservation of momentum of the interacting electron-phonon system since there is no external force involved. This is described in Eq 2.15.

$$\vec{p}_e' = \vec{p}_e \pm \hbar\vec{\beta}$$
$$E_e' = E_e \pm \hbar\omega(\vec{\beta}) \tag{2.15}$$

Unlike collision of solids, here for the term 'elastic collision' we talk only about the energy change of the electrons. The scattering due to Coulomb centers, SR, neutral defects as well as acoustic phonon (due to very small energy change of the electrons) are considered elastic. In the case of an optical phonon, except for very high kinetic energy carriers where the OP absorption/emission energy change is negligible compared to the carrier initial energy, we have to consider energy relaxation time also considering it as inelastic scattering.

In the case of compound materials like GaAs in which bonds are partially ionic, the atoms get slightly polarized with As atom having slight positive charge and Ga atom a negative charge. Therefore in this case we have perturbation of electric dipole moments due to lattice vibrations which results in electric fields and causes scattering. The acoustic phonons causes therefore a polar acoustic phonon scattering also known as piezoelectric scattering and it is said to be dominant at low temperatures while optical phonons causes polar optical phonon scattering and said to be dominant at room temperatures [Lund].

We focus mainly on the acoustic phonon scattering for Si material for linear regime, while inelastic optical phonon LO scattering comes into picture for high fields where electrons gain high kinetic energy. The optical phonon scattering rate remains small for low carrier energies where the scattering is only by absorption of OP. Above a carrier energy threshold, which is the energy of the OP $\hbar\omega_0$, the scattering happens by both emission and absorption and the scattering rate increases drastically and it is dominated by emission. In this case we need to evaluate also the energy relaxation time τ_E. This is more important in the case of high field transport where carrier energy is high enough to emit optical phonons. In this case we need to solve energy and momentum balance equations [Lund] rather than

using the relaxation time approximation (that lead to eq 2.5b). As we include the energy balance equation, the momentum relaxation time gets modified as Eq 2.16. [Lund].

$$< \tau_m >= \frac{\sqrt{\hbar \omega_0 m^*}}{qE_x} \tag{2.16}$$

This leads to carrier velocity saturation with electric field E_x. This is discussed in the section 3. The scattering caused by AP and OP can be intra valley scattering or inter valley scattering. The intra valley scattering scatters the electrons such that they remain in the same valley after scattering or the momentum change after scattering is not large and therefore involves phonons with wave vector near the center of the Brillouin zone. The discussions in previous paragraphs are for intra valley scattering. The inter valley scattering causes electrons to scatter to other valleys and is classified as f-type and g-type. The g-type scattering causes the electron to scatter to exactly opposite valley. For example in Si the electron in valley in <100> direction scatters to <-100> valley. The f-type causes electrons to scatter to other 4 valleys. Since momentum change is large, we need large wave vector AP and OP for this and they are also termed as inter valley phonons.

The expressions in [[Lund] and other literatures] shows that at room temperature, acoustic phonon scattering is considered as isotropic, i.e. the mean free time and momentum relaxation time are same, while at low temperatures it is anisotropic. The scattering rate is proportional to the density of states for both AP and OP scattering and in the case of inter valley scattering it is proportional to number of valleys and density of states of final valley. Inter valley phonon scattering is isotropic. We can express the scattering rates as a simple power law relating to carrier energy and temperature T_L. [Lund]

PHONON MOBILITY IN MOSFET CONTEXT:

In the case of MOSFETs we need to consider 2D confined carriers which lie in different sub bands. Here we have surface acoustic and optical phonon scattering [36 37] which can be intra sub band or inter sub band / valley scattering. At low temperature Sah et al [16] and Kawaji [36] has shown that μ_{ph} is given by Eq 2.17a considering acoustic phonons.

$$\mu_{ph} \propto \frac{z_{av}(E_{eff})}{T} \propto \frac{1}{T(\eta Q_{inv} + Q_D)^{1/3}} \tag{2.17a}$$

The presence of the term $\eta\, Q_{inv} + Q_D$ shows the dependence of phonon mobility to effective electric field in the constant field or triangular potential well approximation for the inversion charge

distribution [8]. This is normally 0.5 for electrons and the wave function in the confinement direction is given by Airy function in terms of this effective field and also the sub band energies. Kawaji/Sah derivation models a 2D acoustic longitudinal phonon in the x-y plane in the inversion region of thickness z_{av}. Note that if we use Fang and Howard approximation for the wave function, the term η will be 11/32 [7]. But this is only an approximate solution intended for simpler analytical expressions without using Airy functions. Generally the analytical solution using the triangular potential well and Airy function is used widely and the value of 0.5 for η is also experimentally more accurate (see section 4 for experimental η extraction).

MULTI SUB BAND AND HIGH K EFFECTS:

The phonon mobility of this type is a very simple picture with only one sub band. By including multi sub band occupation and inter sub band scattering as well as scattering by optical phonons, this relationship has to be more generalized as Eq 2.17b. [8, 16, 37]

$$\mu_{ph} \propto \frac{1}{T^{n}(\eta Q_{inv} + Q_D)^{1/\gamma}}; n \Rightarrow 1-1.5, \gamma \Rightarrow 3-6 \qquad (2.17b)$$

Additional electron mobility degradation is consistently seen in the devices using High K material gate dielectric as discussed in the previous section. Apart from the remote Coulomb scattering due to charges in the interface between High K and SiO_2 inter layer and interface states, mobility degradation is said to arise from remote phonon scattering [22, 53]. This is due to the surface optical (SO) phonon modes at the Si/insulator interface due to the soft LO phonons in the High K dielectric [38]. The soft LO phonons are inherent features of High K materials due to the ionic bond between the metal (Hf for example) and Oxygen (and so easily prone to polarization). They have lower energy and therefore easily couple to the electrons at room temperature. In the case of SiO_2 the bond between Si and O are said to be hard (not easily polarized) and the optical phonons in this material is of very high energy. So electrons in the inversion layers cannot emit phonons at this energy and also at room temperature number of phonons of this energy is low. Therefore this remote phonon scattering is insignificant in the case of SiO_2 (although the non-polarizability of the SiO_2 bond results in a reduced dielectric constant due to which we can no longer use the SiO_2 for MOSFETs in the deca-nanometric regime). It has been experimentally shown that the use of metal gate like TiN reduces the SO phonon from coupling to the channel thereby improving the channel mobility. [53]

In the case of ultra-thin FDSOI devices, the silicon thickness will be smaller than the normal inversion layer extension z_{av}. This causes confinement which [25] has the advantage of preferential occupation of lowest sub band and so enhancement of mobility. However from eq 2.17 we see that the phonon

mobility is dependent on the average inversion layer thickness. And so the phonon scattering mobility will be degraded in these devices. It is shown in [25] that phonon mobility degradation become significant at t_{Si} < 3.5nm. Ultra-thin FDSOI t_{Si} = 3nm in double gate mode pushes the carriers to the bulk and this volume inversion is shown to improve the phonon limited mobility compared to single gate mode. [57]

Inelastic phonon scattering in the quasi ballistic regime MOSFETS (L ~ 10nm) is said to improve I_{on} and reach ballistic currents even in the presence of scattering [50]. This is because the carriers back scattered from the channel should have retained its kinetic energy in order to reach the source. This is not the case for inelastic scattering. Also this scattering happens by emitting LO phonon and therefore happens only when electrons have acquired enough energy. So the carriers that are back scattered to source are those scattered due to elastic scattering near the source region (modeled as kT layer where scattering is predominantly elastic see section 3 also). Those carriers scattered due to LO phonons deep in the channel get absorbed in the drain and so the inelastic scattering helps reach the ballistic currents even for quasi ballistic MOSFETs. This effect is shown to be absent in longer devices in 20nm range and above due to channel potential modification due to carrier pile up (carriers that had lost energy and couldn't return to source pile up in the channel raising the potential and reducing the current).

2.5 Ballistic/quasi ballistic transport

At gate lengths sub 32nm, the channel length will become comparable and even smaller than mean free path and the scattering mechanism becomes less frequent. The electron transport will then reach a limit termed as ballistic limit and in this case the current will reach its maximum limit. This case was already expected in the case of high field effect where the channel electron velocity would reach a saturation velocity. However it was found that there is still advantage in scaling and also v_{sat} model (section 3) shows that the term v_{sat} increases with reducing channel length due to velocity overshoot effect. The channel strain seems to improve I_{on} for short channel devices even with v_{sat} and therefore scaling is still advantageous. However this is not the case for ballistic limit. The ballistic MOSFET drain current expression for one sub band occupation is given in Eq 2.18. The terms $\Im_{1/2}, \Im_0$ are Fermi-Dirac integrals of order ½ and 0 [43, 54] respectively. These integrals take into account the carrier degeneracy. E_F is the Fermi level and E_0 is the bottom of the sub band at virtual source. The ballistic current expression is derived at the top of the source to channel potential energy barrier which is referred to as virtual source where the electric field is 0 and slowly increasing towards the channel. A more general expression is given by Natori [40-42] by solving the effective mass equation and

taking into account multi sub band occupation along with inversion charge approximation for the integrals, expressing the injection velocity and drain current as a function of inversion charge.

$$I_{DS} = WC_{ox}\left(V_g - V_t\right)v_{th}\frac{\Im_{1/2}(\eta_F)}{\Im_0(\eta_F)}\left[\frac{1-\dfrac{\Im_{1/2}\left(\eta_F - \dfrac{qV_d}{kT}\right)}{\Im_{1/2}(\eta_F)}}{1+\dfrac{\Im_0\left(\eta_F - \dfrac{qV_d}{kT}\right)}{\Im_0(\eta_F)}}\right]$$

$$v_{th} = \sqrt{\frac{2kT}{\pi m^*}};\eta_F = E_F - E_0 \qquad (2.18a)$$

$$\tilde{v}_{th} \equiv v_{th}\frac{\Im_{1/2}(\eta_F)}{\Im_0(\eta_F)}$$

We can rewrite the ballistic drain current expression in terms of an apparent mobility term μ_{balapp} and express it in the classical linear regime MOSFET form shown in Eq 2.18b.

$$\mu_{balapp} \equiv \frac{L}{V_d}v_{th}\frac{\Im_{1/2}(\eta_F)}{\Im_0(\eta_F)}\left[\frac{1-\dfrac{\Im_{1/2}\left(\eta_F - \dfrac{qV_d}{kT}\right)}{\Im_{1/2}(\eta_F)}}{1+\dfrac{\Im_0\left(\eta_F - \dfrac{qV_d}{kT}\right)}{\Im_0(\eta_F)}}\right] \qquad (2.18b)$$

$$\therefore I_{DS} = \frac{W}{L}C_{ox}\left(V_g - V_t\right)\mu_{balapp}V_d$$

In the case of small V_d, ($qV_d << kT$) the expression in Eq 2.18a and b can be reduced to Eq 2.19a and non-degenerate limit case is given in Eq 2.19b respectively using the properties of the Fermi-Dirac integrals [54].

$$I_{DS} = WC_{ox}\left(V_g - V_t\right)\tilde{v}_{th}\left[\frac{\Im_{-1/2}(\eta_F)}{\Im_{1/2}(\eta_F)}\right]\frac{qV_d}{2kT} \qquad (2.19a)$$

$$I_{DS} = \frac{W}{L_{eff}}C_{ox}\left(V_g - V_t\right)\mu_{bal}V_d$$

$$\mu_{balapp} = \mu_{bal} = v_{th}\frac{q}{2kT}L_{eff} \qquad (2.19b)$$

The term μ_{bal} was introduced by Shur [5, 55] by introducing L_{eff} and re-arranging the expression for the drain current to look like classical MOSFET form. The expression in the ref [5] and in Eq 2.19b

for μ_{bal} is the same except for the definition of v_{th}. For degenerate case (Eq 2.19a) we can still express μ_{bal} by replacing v_{th} with $\tilde{v}_{th}\left[\frac{\Im_{-1/2}(\eta_F)}{\Im_{1/2}(\eta_F)}\right]$ in which case μ_{bal} will be a function of inversion charge. In the case of quasi ballistic regime devices, scattering still play a role but is less frequent. This is taken into account by a back scattering coefficient "r" which is the probability that the injected carrier get back scattered and collected at source reducing the current. The back scattering/reflection coefficient is related to the mean free path λ_m as $L_{eff}/(L_{eff} + \lambda_m)$ [46]. Assuming the reflection coefficient is same for source injected and drain injected carriers, the quasi ballistic drain current expression is obtained as shown as in Eq 2.20a.

$$I_{DS} = WC_{ox}\left(V_g - V_t\right)\left[\frac{1-r}{1+r}\right]\tilde{v}_{th}\left[\frac{1-\dfrac{\Im_{1/2}\left(\eta_F - \dfrac{qV_d}{kT}\right)}{\Im_{1/2}(\eta_F)}}{1+\left[\dfrac{1-r}{1+r}\right]\dfrac{\Im_0\left(\eta_F - \dfrac{qV_d}{kT}\right)}{\Im_0(\eta_F)}}\right] \tag{2.20a}$$

In the linear regime and non-degenerate limit we can reduce the expression as Eq 2.20b. The term (1-r) $= \lambda_m/(L_{eff} + \lambda_m) \approx \lambda_m/L_{eff}$ (for large L_{eff}) is the ballistic ratio BR in the linear regime.

$$I_{DS} = WC_{ox}\left(V_g - V_t\right)(1-r)\tilde{v}_{th}\left[\frac{\Im_{-1/2}(\eta_F)}{\Im_{1/2}(\eta_F)}\right]\frac{qV_d}{2kT}$$
$$I_{DS} = WC_{ox}\left(V_g - V_t\right)(1-r)v_{th}\frac{qV_d}{2kT};(non\ degenerate) \tag{2.20b}$$

If we go back to Shur model and classical MOSFET equation, then we can see that the product of ballistic ratio and ballistic mobility gives the effective mobility as shown in Eq 2.21.

$$\mu_{eff} = (1-r)\mu_{bal} = \left[\frac{1}{\mu_{bal}} + \frac{1}{\mu_{effdd}}\right]^{-1}$$
$$\therefore \frac{1}{\mu_{effdd}} = \frac{kT/q}{D_n} = \frac{1}{\mu_{bal}}\left[\frac{r}{1-r}\right] = \frac{1}{\mu_{bal}}\left[\frac{L_{eff}}{\lambda_m}\right] \tag{2.21}$$
$$\therefore D_n = \frac{1}{2}v_{th}\lambda_m\ (non\ degenerate)$$

The drift diffusion mobility μ_{dd} can be written in terms of μ_{bal} and mean free path as shown. D_n is the diffusion coefficient using the Einstien relation for μ_{effdd}. The case involving multi sub band occupation is simplified by introducing a correction factor of 2.5 assuming that 80% of carriers occupy lowest sub band in Natori model [40]. V. Barral et al take care of multi sub band by following similar approach and generalizing the Eq 2.20a [46-47] by adding the contribution of carriers in each of the

sub bands and valleys. Something to remember here is that in the case of multi sub band occupation, we have twofold and fourfold valleys and therefore depending on sub band we have to take appropriate valley degeneracy (2 or 4) and appropriate effective masses. Final drain current expression is therefore somewhat more complex as given in ref [46]. Multi sub band occupation however has the impact of smaller ballistic current compared to single sub band case due to various factors (for example the effective mass in higher sub bands are larger and so overall v_{inj} becomes smaller than that of the single sub band case). This therefore gives a scope for improvement in the I_{on} even in the ballistic case. With the help of sub band engineering we can force preferential occupation of lowest sub band there by 'restoring' v_{inj} [25, 51]. In this case ultra-thin Si FDSOI is more effective in preferential occupation of lowest sub band than strain. Once all the electrons occupy the lowest sub band there will be no further improvement possible unless significant band deformation (tensile uni-axial strain in addition to ultra-thin Si) could be induced to reduce the effective mass smaller than m_t [58]. In our case, in results shown in chapter 6 we have analyzed the low field mobility and applied the non-degenerate limit ballistic mobility for the correction using Shur approach.

Some of the important results for drain current based on ballistic and quasi ballistic theories from the literature have been discussed in this chapter. Detailed derivation can be obtained from ref [40- 47]. An analytical expression derived by similar approach for ballistic magnetoresistance is discussed in the chapter 4.

2.6 Conclusion

Different scattering mechanisms that are known so far to govern the linear regime MOSFET operation focusing on advanced MOSFETs has been covered in this chapter. Extraction in the linear regime is much more convenient as extraction methods are simpler as well as governing physics is known and improved from time to time. As linear regime mobility and I_{on} are correlated, understanding and improvement of linear regime models is still important. Some of the state of the art methods explored in this thesis for the extraction in linear regime are discussed in the section 4. We have also come up with an improved extraction method for more accurate extraction in this regime by taking into account gate dependent series resistance for deep sub µm MOSFETs. Section 4 also gives a temperature model for low field mobility used for analyzing different scattering mechanisms based on the temperature dependence of different mechanisms discussed in this chapter. Ballistic MOSFET and quasi ballistic MOSFET expressions become more important as technology node advances below 32nm. We have also studied its impact in magnetoresistance and a new analytical expression for ballistic and quasi ballistic MR is developed and they are given in section 4.

References

[1] M Saitoh, Ken Uchida “Universal relationship between Low Field Mobility and High Field Carrier velocity in high K and SiO_2 gate dielectric MOSFETs” IEDM, 10.1109/IEDM.2006.346757, 2006

[2] A. Lochtefield, D. A. Antoniadis “Investigating the relationship between electron mobility and velocity in deeply scaled NMOS via mechanical stress”. IEEE EDL, vol 22, n 12, p591, 2001

[3] G.Ghibaudo et.al, "Electrical Transport Characterization of nano CMOS devices with ultra thin silicon film", p58, IEEE 9th intl workshop on Junction Technology 2009.

[4] A Cros et.al , "Unexpected Mobility degradation for very short devices: A new challenge for CMOS Scaling", IEDM Tech. Digest, DOI:10.1109/IEDM. 2006.346872 (2006)

[5] M S Shur, “Low Ballistic Mobility in Submicron HEMTs” IEEE EDL vol23 n9 p511 2002

[6] Shin-ichi Takagi et.al “On the universality of Inversion layer Mobility in Si MOSFET’s: Part I – Effects of Substrate Impurity Concentration”, IEEE TED vol41, n12, p2357, 1994

[7] T.Ando, A B Fowler, F Stern, “Electronic Properties of 2D systems”, review of modern physics vol54, n2, p437, 1982.

[8] D.S.Jeon, D.E.Burk, “MOSFET Electron Inversion Layer Mobilities-A Physically based Semi-Empirical Model for a Wide Temperature Range” IEEE TED vol 36, n8, p1456, 1989

[9] A Hartstein, A B Fowler and M Albert, “Temperature dependence of scattering in the inversion layer”, Surface Science p181-190, 1980

[10] F.Stern, W.E Howard, “Properties of Semiconductor Surface Inversion Layers in the Electric Quantum Limit” Physical review, vol163, n3, p816, 1967

[11] F Stern, “Calculated Temperature Dependence of Mobility in Si Inversion Layers”, PRL vol 44 n 22 p1469 1980

[12] K M Cham, R G Wheeler, “Temperature-Dependent Resistivities in Silicon Inversion Layers at Low Temperatures” PRL vol 44 n 22 p1472 1980

[13] Y Kawaguchi, T Suzuki, S Kawaji “Carrier Concentration Dependence and Temperature Dependence of Mobility in Silicon (100) N-Channel Inversion Layers at Low Temperatures” Solid State Comm, vol 36 p257 1980.

[14] A Hartstein, T H Ning and A B Fowler, “Electron scattering in Silicon inversion layers by oxide and surface roughness” Surf Sci 58 p178-181 1976

[15] S.Mori, T Ando, “Intersubband scattering effect on the mobility of a Si (100) inversion layer at low temperature” Phy rev B, vol19, n12, p6433, 1979

[16] C T Sah, T H Ning, and L L Tschopp, “ The Scattering of Electrons by Surface Oxide Charges and by Lattice Vibrations at the Silicon-Silicon dioxide Interface”, Surface Science vol 32 p561-575 1972

[17] C C Shiue, C T Sah, “Studies of electron screening effects on electron mobility in silicon surface inversion layers”, Surface Science vol58 p153 1976

[18] F Stern, “Screening and Level broadening in inversion layers with random fixed charges”, Surface Science vol 58, p 162, 1976

[19] M S Krishnan et.al "Remote Charge Scattering in MOSFETs with ultra-thin gate dielectrics" , IEDM p571 1998

[20] L.Thevenod et.at "Magnetoresistance mobility extraction on $TiN/HfO_2/SiO_2$ metal-oxide-semoconductor field effect transistors" APL, vol 90 -152111, 2007

[21] D Heh et al, "Mobility Reduction and Reliability Assesment of high-K/Metal Gate Stacks in deep Sub-Nanometer EOT region", IEEE ESSDERC, p257, 2009

[22] M Casse et al, "Carrier Transport in HfO_2/Metal Gate MOSFETs: Physical Insight Into Critical Parameters", IEEE TED vol 53 n4 p759 2006

[23] M.Khare et al, "Transconductance in Nitride-Gate or Oxinitride-Gate Transistors", IEEE EDL vol 20 n1 p57 1999

[24] D M Fleetwood, " "Border Traps" in MOS devices" IEEE Trans on Nuclear Science, vol 39, n2, p269 1992

[25] S Takagi et al, "Sub-band structure engineering for advanced CMOS channels", Solid State Electronics vol 49, p684, 2005

[26] O Weber and S Takagi "Experimental Examination and Physical Understanding of the Coulomb Scattering Mobility in Strained-Si n MOSFETs" IEEE TED, vol 55, n9, p2386, 2008

[27] W Zhu, J P Han, T P Ma, "Mobility Measurement and Degradation Mechanisms of MOSFETs Made with Ultra thin High-K Dielectrics", IEEE TED vol 51, n1, p98, 2004

[28] T Ando, "Screening Effect and Quantum Transport in a Silicon Inversion Layer in Strong Magnetic Fields", Journal of the Phys Soc of Jap, vol 43, n5, p1616 1977

[29] Y Matsumoto, Y Uemura, "Scattering Mechanism and Low Temperature Mobility of MOS Inversion layers", Proc 2nd intl Conf on Solid Surf, JJAP suppl, 2 pt 2 p367 1974

[30] S M Goodnick et al, "Surface Roughness at Si (100)-SiO_2 interface" Physical Rev B vol 32 n12 p8171 1985

[31] Yi Zhao, M Takenaka, S Takagi, "On Surface Roughness Scattering – Limited Mobilities of Electrons and Holes in Biaxially Tensile-Strained Si MOSFETs" IEEE EDL, vol 30, n9, p987, 2009

[32] O Bonno et al, "High Field Electron Mobility in Biaxially-Tensile Strained SOI: Low temperature Measurement and Correlation with the Surface Morphology", Symposium on VLSI Tech Digest p134 2007

[33] A Pirovano et al, "Explaining the dependences of Hole and Electron Mobilities in Si Inversion Layers" IEEE TED vol 47, n4 p718 2000

[34] R Versari, B Ricco " MOSFET's Negative Transconductance at Room Temperature" IEEE TED, vol 46, n6, p1189, 1999

[35] H Tsuchiya, S Takagi, "Influence of Elastic and Inelastic Phonon Scattering on the Drive Current of quasi ballistic MOSFETs" IEEE TED, vol 55, n 9, p2397, 2008

[36] S Kawaji "The Two-Dimensional Lattice Scattering Mobility in a Semiconductor Inversion Layer" Journal of Physical Society of Japan vol 27, n4, p906, 1969

[37] H Ezawa, S Kawaji, K Nakamura "Surfons and the Electron Mobility in the Silicon Inversion Layers" Japanese Journal of Applied Physics, vol 13 n1 p126 1974

[38] M V Fischetti et al, "Effective electron mobility in Si inversion layers in metal-oxide-semiconductor systems with a high-k insulator: The role of remote phonon scattering" Journal of Applied Physics vol90, n9 p4587 2001

[39] D Esseni, F Driussi, "A Quantitative Error Analysis of the Mobility Extraction According to the Matthiessen Rule in Advanced MOS Transistors" IEEE TED vol58 n8 p2415 2011

[40] K Natori, "Ballistic metal-oxide-semiconductor field effect transistor", J. appl Phy, vol 76, n8 p4879 1994

[41] K Natori, " Ballistic MOSFET Reproduces Current – Voltage Characteristics of an Experimental Device" IEEE EDL, vol23, n11, p 655, 2002

[42] K Natori, "Ballistic/quasi ballistic transport in nanoscale transistor" Applied Surface Science vol 254 p6194 2008.

[43] A Rehman and M S Lundstrom, "A Compact Scattering Model for the Nanoscale Double-Gate MOSFET" IEEE TED, vol49, n3, p481, 2002

[44] M S Shur, L F Eastman, "Near Ballistic Electron Transport in GaAs Devices at 77K", SSE, vol24, p11, 1981

[45] J Lusakowski et al "Electron mobility in quasi ballistic Si MOSFETs", SSE vol50 p632 2006

[46] V Barral et al "Experimental Investigation on the quasi ballistic Transport: Part I – Determination of a New Backscattering Coefficient Extraction Methodology" IEEE TED vol56 n3 p408 2009

[47] V Barral et al "Experimental Investigation on the quasi ballistic Transport: Part II – Backscattering Coefficient Extraction and Link with the Mobility" IEEE TED vol56 n3 p420 2009

[48] A.Mathiessen and C Vogt, "On the influence of Temperature on Electric Conducting-Power of Thallium and Iron", Phil.Trans.R.Soc.London, vol153, p369-383, 1863

[49] Sorin Cristoloveanu, Noel Rodriguez and Francisco Gamiz, "Why the Universal Mobility Is Not" IEEE TED, vol57, n6, p1327, 2010

[50] H Tsuchiya, S Takagi "Influence of Elastic and Inelastic Phonon Scattering on the Drive Current of quasi ballistic MOSFETs" IEEE TED, vol55, n9, p2397, 2008

[51] M Ferrier et .al "Saturation Drain Current analytical modeling of Single Gate Fully Depleted SOI or SON MOSFETs in the quasi ballistic Regime of Transport" IEEE Intl SOI Conf Proceedings, p107, 2006

[52] M J Chen et.al "Temperature Dependent Channel Backscattering Coefficient in Nanoscale MOSFETs", IEDM, p39, 2002

[53] R Chau et.al "High-k/Metal-Gate Stack and Its MOSFET Characteristics" IEEE EDL vol25 n6 p408 2004

[54] J S Blakemore "Approximations for Fermi-Dirac integrals, especially the function $F_{1/2}(\eta)$ used to describe electron density in a semiconductor" SSE vol25 n11 p1067 1982

[55] K Huet "Monte Carlo Study of Apparent Mobility Reduction in Nano-MOSFETs", ESSDERC, p382, 2007

[56] J P McKelvey, J C Balogh, "Flux Method for the analysis of Transport Problems in Semiconductors in the presence of Electric Fields" Physical Rev, vol137, n 5A, p A1555, 1965

[57] T Ernst et.al "Ultimately thin Double-Gate SOI MOSFETs" IEEE TED vol 50 n3 p830 2003

[58] F.Rochette et.al "Piezoresistance effect of strained and unstrained fully-depleted silicon-on-insulator MOSFETs integrating a HfO2/TiN gate stack", Solid-State Electronics, vol53, n3, p392, 2009

3 Electron Transport in MOS – Saturation Regime

This chapter is dedicated to the saturation region operation of MOSFETs. The transport in this regime for sub μm MOSFETs is governed by high field effects which is different from that in the linear regime even though there is a correlation between the transport model parameters of the two regimes. It is therefore important to study this region apart from studying the linear regime parameters. This chapter covers the high field transport seen in the saturation regime for sub μm MOSFETs and discusses saturation velocity v_{sat}, and velocity overshoot effect (for deca nanometer regime MOSFETs). This chapter also discusses the full ballistic and quasi ballistic models in saturation regime more suitable for sub 32nm MOSFETs.

INTRODUCTION:

The classical MOSFET saturation region transport is different from its linear region in that the channel inversion layer is pinched off near the drain. The drain current expression still uses the effective mobility term in the case of long channel MOSFETs with channel length few μm and above. The expression is shown in the Eq 3.1[Sze]. The channel pinch off happens when the drain voltage exceeds V_g-V_t near the drain side and the current reaches a maximum value as given in Eq 3.1.

$$I_{DS} = \frac{W}{2L} C_{ox} \mu_{eff} (V_g)(V_g - V_t)^2 \tag{3.1}$$

This model is no longer used in the advanced devices starting from 1 or so μm channel length and below. This is because at these channel lengths the tangential electric field exceeds the so called critical electric field and at this high field the electron drift velocity would saturate and the drain current tends to saturate even before the pinch off point. We focus our attention on the high field effects and related saturation velocity v_{sat} model for sub μm MOSFET devices.

3.1 High Field effect - Velocity Saturation

In the case of low field transport, the relaxation time approximation (using mean free time) was used to solve the Boltzmann transport equation [Lund, 20]. This would enable us to use energy averaged mean free time we discussed in the chapter 2 for various scattering mechanisms. We are mainly

concerned with the momentum relaxation and momentum balance equation (Drude model and Fick's law) in the linear regime. We use Mathiessen rule to add the rates due to individual scattering leading to the final mobility term that appears in the linear regime drain current expressions. In the case of high field transport, we need to use the complete momentum and energy balance equations, derived from Boltzmann transport equation. The relaxation time approximation or assumption of small perturbation of distribution function (Boltzmann or Fermi-Dirac) due to lateral field is no longer valid at high fields. These balance equations are similar to the carrier density balance/continuity equations seen in [Sze and other literatures] that keep an account of number of carriers in terms of carrier flux and generation/recombination rates. In the case of momentum/energy, we describe average momentum/energy of carriers per unit volume, which is the product of average electron momentum/energy per carrier and number of carriers per unit volume. This is then expressed in terms of energy/momentum "flux" and "generation/recombination" rates. In the case of momentum/energy generation rate, we have increase/decrease of momentum/energy per unit volume due to carrier generation/recombination as well as due to the electric field (or by absorbing LO phonon) as it accelerates carriers to higher momentum/energy (thereby "generating" momentum/energy per unit volume). The recombination rate in the context of momentum/energy is loss of momentum/energy due to scattering (momentum and energy relaxation times) as well as loss due to carrier recombination. Comparing this to continuity equation of the carriers, we have rate change of carrier per unit volume accounted for by carrier influx and carrier generation and recombination. The lateral field and collision does not create or destroy carriers, they only affect the momentum/energy and so they appear only in the momentum and energy balance equations [Lund page 216, 20].

The full energy and momentum balance equation uses ensemble average relaxation rates as given in Eq 3.2.

$$\left\langle\left\langle \frac{1}{\tau_m(E(p))} \right\rangle\right\rangle = \frac{\left\langle \frac{E(p)}{\tau_m(E(p))} \right\rangle}{\langle E(p) \rangle}; \mu = \frac{q}{m^* \left\langle\left\langle \frac{1}{\tau_m(E(p))} \right\rangle\right\rangle}; \qquad (3.2)$$

$$\left\langle\left\langle \frac{1}{\tau_E(E(p))} \right\rangle\right\rangle = \frac{\left\langle \frac{E(p)}{\tau_E(E(p))} \right\rangle}{\langle E(p) \rangle}$$

This expression is based on assumption that the carrier distribution can be approximated by displaced Maxwellian distribution function as shown in fig 3.1 due to heavy scattering. Fig 3.1 we see three distributions, one at equilibrium with lattice temperature T_L, second one is shifted Maxwellian in low field, which is also in thermal equilibrium with lattice but with a net drift velocity v_d and third one is for the high field is drifted and heated Maxwellian distribution where carrier temperature T_e is

significantly higher than lattice temperature. Carriers with velocities above v_{sat} can emit optical phonons and loose energy and higher the carrier energy higher the probability to emit phonon. This would cause a spread in the distribution to higher temperature T_e. The carrier mobility can be shown to reduce with the 'heating' or electric field (known as hot electron effect) and the carrier drift velocity v_{dH} saturates around v_{sat} at very high fields. An increased electric field would cause carrier heating rather than increase in its average drift velocity. This kind of symmetrical distribution is a good approximation if scattering is large and carrier momentum is randomized sufficiently. It is important to note that the distribution function is in fact distorted from Maxwellian. Also it is important to note here that we can no longer use simple description of E(**p**) (**p**=ħ**k**) like parabolic or non-parabolic variants for very high energies. This can be taken into account only in numerical simulations like Monte-Carlo/ Full band Monte-Carlo etc and in the case of analytical models we have to assume an approximate distribution as discussed above. The term τ_m and τ_E are the same as that we discussed in the section 2 for different scattering mechanisms, with τ_m for the momentum (used in the mobility expression) and τ_E for the energy relaxation (mainly due to optical phonon at high energy). The effective mass in the mobility term is also different from that used in the linear regime (which was the effective mass in the bottom of the band).

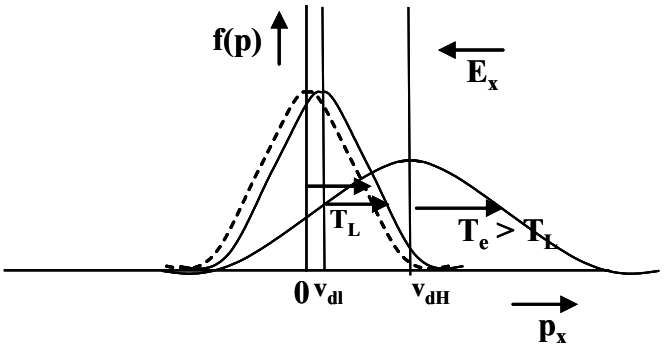


Figure 3.1

The distribution function f(p) in the momentum space p is shown for equillibrium (dashed), low field and high field (line). In the case of Low field f(p) is shifted by v_{dl} (drift velocity) but retains the lattice temperature while for high field the electron temperature T_e is significantly higher than the lattice temperature T_l but the distribution is assumed Maxwellian with a higher v_{dH}. This assumption is made for the sake of analytical solution for the balance equations also known as electron temperature approach. E_x is the electric field, p_x is the momentum in x-direction, 'x' being the transport direction. Carriers with velocity $>v_{dh} = v_{sat}$ can emit OP and loose energy.

Once we take into account the energy relaxation at high field due to the OP scattering, τ_m will get modified and to simplest order can be shown to be inversely proportional to the electric field as shown in Eq 3.3 [Lund, 1].

$$\tau_m = \frac{\sqrt{m^* \hbar \omega_0}}{qE_x}$$
$$v_{sat} = \sqrt{\frac{\hbar \omega_0}{m^*}} \approx 10^7 \, cm/s \qquad (3.3)$$

An f-type inter valley phonon scattering is said to dominate in the case of bulk Si [7] of energy ~ 54.4meV. Eventually at high field the electron drift velocity saturates. This saturation velocity for Si is ~ 10^7cm/s at room temperature and this has also been verified experimentally [1]. The transport process at high electric field is complex and expression in 3.3 is only for demonstrating a simple picture of what to expect at high fields. The saturated velocity is a function of temperature and is found experimentally to decrease with increasing temperature above 45K. Below 45K v_{sat} reaches a constant value of ~ 1.4 10^7cm/s. [1] To explain this we need to use simulations or for analytical purpose use simplifying assumptions like the electron temperature as shown in fig 3.1.

As discussed earlier, τ_m can be expressed to have power law dependence to energy. And we can express the electron temperatures in terms of electric field form the energy balance equation by knowing τ_E for the governing scattering mechanism for energy relaxation (optical phonon). The mobility by taking acoustic phonon scattering as dominant scattering mechanism for momentum relaxation τ_m, which is true for lightly doped Si where Coulomb scattering can be neglected, is given by Eq 3.4a. E_C is a critical field and we get at high electric field, v_{sat} as $E_C \mu_0 \sim 10^7$cm/s determined by the energy relaxation mechanism (inter valley scattering). μ_0 is the low field mobility (not to be confused with the low field mobility μ_0 of section 5. This μ_0 is same as the $\mu_{eff}(V_g)$ for low V_d). The generalization of the same is given in Eq 3.4b in terms of electric field as well as carrier temperature. For Coulomb scattering the carrier temperature dependence n = 1.5 and for acoustical phonons it is -0.5 (for mobility, we talk about momentum relaxation time which depends on the carrier temperature). For generalized expression for field dependence, E_C and β are fitting parameters. The saturated velocity v_{sat} temperature dependence is modeled by Jacoboni et.al [1] as Eq 3.4c for silicon.

$$\mu(E) = \mu_0 \left(\frac{T_e}{T_L} \right)^{-1/2} = \frac{\mu_0}{\sqrt{1 + (E/E_C)^2}} \qquad (3.4a)$$

$$\mu(E) = \mu_0 \left(\frac{T_e}{T_L} \right)^{n} = \frac{\mu_0}{\left(1 + (E/E_C)^{\beta}\right)^{1/\beta}} \qquad (3.4b)$$

$$v_{sat} = \frac{2.4\times10^{7}}{1+0.8e^{T/600}} \tag{3.4c}$$

Thornber [18] suggested a modified expression to 3.4b to express the drift velocity using more physically meaning full parameters such as saturation velocity v_{sat}, tangential field E and low field mobility μ_0 as shown in the Eq 3.4d. This avoids the need for defining a critical field E_C. The advantage of this model is that it separates the term E_C into $1/\mu_0$ that represents the scattering and v_{sat} that represents optical phonon energy. This would conveniently include any change in the scattering through mobility term where μ_0 is a function of effective field (for the MOSFET case), doping, temperature etc.

$$v_d = \frac{\mu_0 E}{\left(1+(\mu_0 E / v_{sat})^{\beta}\right)^{1/\beta}} \tag{3.4d}$$

The above properties are discussed with relation to bulk Si. In the case of MOSFETs we need to take into account the surface carriers as well as the electric field variation along the channel, which leads to electron velocity overshoot in very short devices. Further, in the absence of scattering, we have ballistic transport model for the saturation region. These are discussed in the following sections. Extraction of v_{sat} for MOSFETs is discussed in the section 5 and results on different devices are given in section 6.

3.2 Velocity Saturation in MOSFETs

As channel length reduces to sub-micrometer regime and further down to deca-nanometer regime, the classical MOSFET saturation will no longer be applied due to the high field effects. This effect has been shown to be present starting from few μm channel lengths and below [4-8]. The classical MOSFET pinch off voltage V_{Dsat} is V_g-V_t, but here the saturation occurs even before due to the velocity saturation effects. This is shown in fig 3.2 for FDSON device where $I_{DS} - V_d$ calculated from linear regime assuming long channel behavior (discussed in section 5) is compared to the measured I_{DS} - V_d (here we also have DIBL effect, see section 6 for full characterization results and discussion). The saturation region MOSFET drain current model is given in Eq 3.5 (DIBL effect included inside V_t term, see section 5).

$$I_{DS} = WC_{ox}\left(V_g - V_t\right)v_{sats} \tag{3.5}$$

In this case the drain current, I_{DS}, is expressed in the source region and v_{sats} is the saturated velocity at source [4]. Also in the source region the inversion charge density can be expressed as $C_{ox}(V_g - V_t)$. As

we see the expression shows that I_{on} is independent of channel length. This would mean that channel length scaling will not have I_{on} advantage, although C_{ox} scaling can still improve the on current per unit width. Also reduced transistor dimensions mean that load each transistor needs to drive (total gate capacitance) is smaller and so there is still performance advantage in scaling as well as advantage in terms of transistor density. Also we can improve I_{on} by applying channel uniaxial or biaxial strain improving the low field mobility and therefore I_{on} [2, 3].

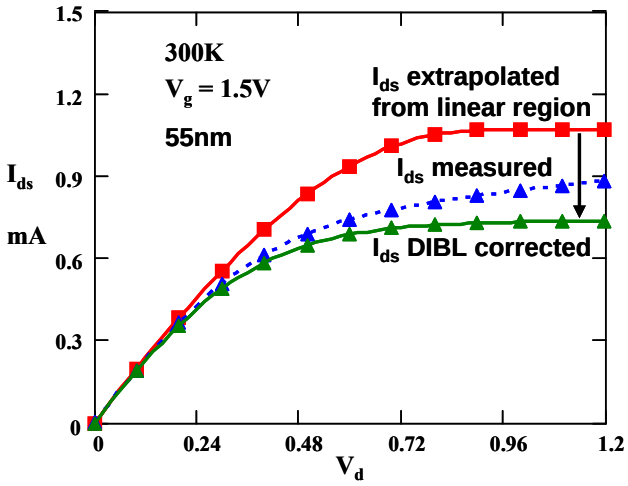


Figure 3.2

$I_{DS} - V_D$ for FDSON 55nm long and 1µm wide device at room temperature is shown (see section 6). Linear Regime $I_{DS} - V_g$ is used to extrapolate the $I_{DS} - V_D$ that is expected with long channel assumption. The measured I_{DS} is lower than the expected which is generally attributed to v_{sat} effect. DIBL corrected I_{DS} is also shown.

The channel mobility can be fitted as a function of electric field in a similar way as discussed in the bulk case (Eq 3.4 d) as given by Thornber [18]. Although Eq 3.4b is also widely used by considering $v_{sats} = \mu E_C$ and therefore using E_C as fit parameter. Mobility μ is the low field mobility which is dependent on the transverse effective field or gate voltage. The expression is shown in Eq 3.6. Note that the term v_{sats} is used here instead of v_{sat} of the bulk silicon. This has been studied by Modelli and Manzini [6] by experiment and simulation and found that using v_{sat} of bulk silicon does not properly recreate the measured drain current, but by using measured v_{sats} the measured drain current was best reproduced by simulation using the model of Eq 3.6.

$$v_d(E_{II}, E_\perp) = \frac{\mu(E_\perp)E_{//}}{\left(1 + (\mu(E_\perp)E_{//} / v_{sats})^\beta\right)^{1/\beta}} \tag{3.6}$$

The simulation was done by taking local values for parameters along the channel such as electric fields, mobility and doping concentration etc. v_{sats} is generally taken to be independent of transverse effective field although it is also weakly sensitive to the effective field or gate voltage. $E_{//}$ is the

electric field in the transport direction and $E_{\perp}$ is the transverse effective field E_{eff}, v_d is the drift velocity and $\mu(E_{\perp})$ is the mobility as a function of E_{eff} described in the section 2 and 4 for the linear regime operation. The effective field E_{eff} reduces the surface carrier mobility and therefore we need larger tangential field to attain the same drift velocity (modeled as scattering scaling by Thornber [18]).

The value of v_{sats} is only a fraction of 10^7cm/s as reported in different literatures [ref 5 for example where bulk saturation velocity and surface saturation velocity has been compared] and also shown in this thesis (see section 6) for channel lengths ~ 100nm and above. Fang and Fowler [7] measured v_{sats} for (100) Si MOSFET 10μ long MOSFET at high drain and gate voltage, so that lateral field ~ 10^4V/cm is achieved, as 6.5 10^6cm/s and (111) surface ~ 5.5 10^6cm/s and (110) surface ~ 4.5 10^6cm/s. In this measurement, large gate voltage is applied such that the carrier pinch off does not occur near the drain side while applying high V_d for obtaining high electric field. The saturation velocity as a function of temperature is similar to bulk in that v_{sats} decreases with increasing temperature. The measurement made by using resistive gate method by Coen and Muller [5] gives v_{sats} ~ 5 10^6cm/s for (100) surface. By using resistive gate a varying gate voltage along the channel is applied there by inducing a uniform inversion in the channel and this allows for the assumption of constant electric field along the channel as V_d/L (In gradual channel approximation, the average field along the channel is given by V_d/L [Sze, 10]). Considering the errors of extraction method this is fairly closer to the value given by Fang and Fowler. Modelli and Manzini [6] also obtained v_{sat} value of 6 10^6cm/s by resistive gate method and also results were validated by simulation results as explained above. The saturation velocity of holes is smaller ~ 2 10^6cm/s. The temperature dependence of hole- v_{sats} is shown to be much smaller than electron v_{sats} [5].

The saturation velocity for FDSON n-MOSFET is discussed in the section 6. The value for 0.24μ long MOSFET obtained is around 4 10^6cm/s, this is closer to the values mentioned before and this variation could be attributed to higher order effects like high K gate stack, high normal field ~ 1MV/cm, increased series resistances in FD SOI and FD SOI like structures and due to heating effect etc. The series resistance effect is mostly the dominant cause for under-estimation of v_{sats}. Temperature dependence shows the expected behavior.

The saturation velocity measured in the silicon depletion layers of optically induced charge packets [9] is shown to be much larger (close to 10^7cm/s) than v_{sats} discussed above (which is in the strong inversion). The drain current expression near the drain side is very complicated although electron drift velocity is taken to be v_{sat} at high $V_D > V_{Dsat}$ [4, 8]. The model by Sodini et al [4] explains the smaller value for saturation velocity v_{sats} using the expression Eq 3.7.

$$v_{sats} = \frac{gm_{sat}}{WC_{ox}} \approx \frac{v_{sat}}{1+\frac{2v_{sat}L\Theta_{eff}}{\mu_0}} \; for\,(V_g - V_t) >> E_C L \tag{3.7a}$$

$$v_{sats} = \frac{I_{Dsat}}{WC_{ox}(V_g - V_t)} = \left(1 - \frac{V_{Dsat}}{V_g - V_t}\right) v_{sat}$$
$$V_{Dsat} = \frac{E_C L(V_g - V_t)}{E_C L + (V_g - V_t)} \tag{3.7b}$$

In Eq 3.7a, v_{sats} is channel length dependent and also dependent on the linear regime effective mobility μ_{eff} parameters μ_0 and Θ_{eff} (given in section 4). The lateral field which tends to increase the drift velocity is compensated by the transverse field causing the drift velocity to saturate below the actual v_{sat}. The Eq 3.7b was given based on drain current using V_{Dsat} model under the assumption that the carriers are velocity saturated due to high drain field near the drain side and the this velocity is taken as v_{sat} of bulk. V_{Dsat} for long channel MOSFET is just V_g-V_t where the velocity saturation does not occur in the inversion channel region before pinch off point.

The formulation mentioned above is based on the assumption that the gradual channel approximation holds in the channel region even at high V_d. However in high V_d this assumption does not hold. In fact Muller and Eisele [10] attribute this "apparent" source drift velocity saturation v_{sats} to source electric field "pinning" rather than the physical velocity saturation at source justifying its low value. The current continuity imposes the condition that the velocity near drain should be higher than that at source as the gate over drive and therefore inversion carrier at drain side is smaller. This would mean a homogenous increase of electric field along the channel under the gradual channel approximation [Sze]. However as velocity reaches the saturation value v_{sat} the electric field has to be much higher to increase the velocity of the carriers and therefore drain side electric field will increase much faster and pins the electric field near source (or the increase of source side field with V_d is more gradual) thereby controlling the source injection (near source the inversion charge is maximum = $C_{ox}(V_g - V_t)$). Therefore the carrier apparent saturated velocity at source v_{sats} Eq 3.5 will be smaller than v_{sat}. They have also shown that v_{sats} increases as channel length decreases owing to the asymmetric field distribution along the channel. For longer channel lengths the pinned source field is shown to be smaller giving smaller saturated velocity v_{sats}. The simulation uses v_{sat} model Eq 3.6 with v_{sat} = 10^7cm/s and $\beta = 2$ and current continuity condition to obtain the results. Section 6 results on FDSON devices shows the channel length dependence of v_{sats} and we can see that v_{sats} decreases with increasing channel length.

Section 5 discusses method for the extraction of v_{sats}. We can still model the source drift velocity using Eq 3.6 by replacing v_{sat} with v_{sats} even though the physical meaning for v_{sats} is not the same as bulk v_{sat}. In the case of channel length ~ 55nm, velocity overshoot effect is seen (especially at low temperature). In the case of extreme short channel devices ~ 32nm region and below the transport is expected to be in quasi ballistic regime and I_{on}, is modeled accordingly as in the case of linear regime. These are discussed in the following sections.

3.3 Velocity overshoot

Velocity overshoot effect is caused by non-stationary/off-equilibrium transport due to large energy relaxation time. All the expressions in low or high fields in previous sections are derived for steady state condition using the balance equations where the carriers are in equilibrium for a given electric field (energy gain balanced by energy loss). For a uniform DC electric field applied at t=0, solution of Eq 2.2 (neglecting the diffusion terms, for large and uniform field, drift current is large compared to diffusion) will be as shown in Eq 3.8 [Lund, 11].

$$\begin{aligned} &< v_d >_x = -\frac{q\tau_m(E_x)}{m_{eff}} E_x \left(e^{-t/\tau_m} - 1\right) \\ &\therefore < \lambda >_x = -\frac{q\tau_m^2(E_x)}{m_{eff}} E_x e^{-1} \end{aligned} \tag{3.8}$$

Therefore electrons reach a drift velocity $<v_d>_x$ after few τ_m time traveling few $<\lambda>_x$ distance, which is the mean free path. The electron temperature is determined from the energy balance equation and can be expressed similar way as for momentum but τ_m replaced by τ_e. The time it takes or distance electron has to travel (inelastic mean free path) for the electrons to heat up is larger than that for the velocity drift. Also the momentum relaxation time decreases with electron temperature. If we consider an electric field transient region such that the change in the electric field is faster than the energy relaxation time or length, the electron temperature cannot increase as fast and so the mobility reduces slower (spatially) than the electric field. Therefore the electron drift velocity which is dependent on the local mobility and local electric field shoots up (assuming that the momentum relaxation time is very small) as it sees higher mobility than that it would have seen locally at steady state conditions. Therefore if we look at the transient period, the drift velocity would increase according to the electric field initially without much change in the distribution shape. After few energy relaxation times the shape gets broadens and therefore mobility reduces and the drift velocity would settle at a lower value. In this picture the peak velocity achieved in the case of $\tau_e >> \tau_m$ is $\mu_0 E$ where μ_0 is the low field mobility (μ_{eff}). This is shown in the fig 3.3. Figure is exaggerated but as we see, with time the drift velocity initially shoots up and then settles to a lower value. The fig 3.3 shows the field transient in time [11, 12].

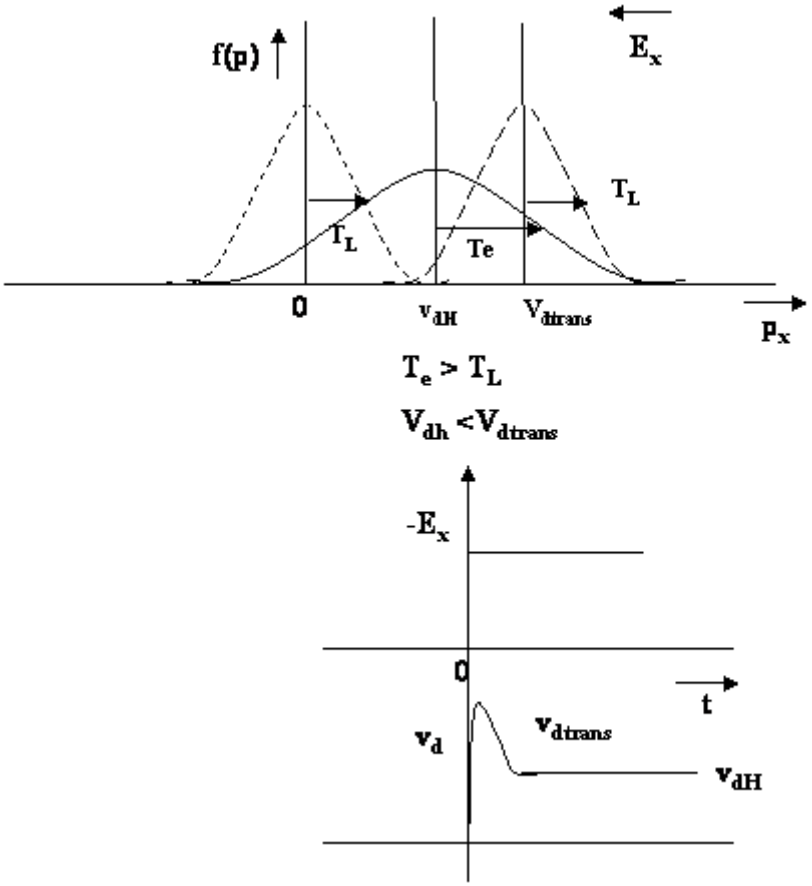


Figure 3.3

Top: shows the evolution of the distribution funtion for a step field with time. Initially the carrier acquire large transient drift velocity after few momentum relaxation time (assuming $\tau_m << \tau_e$). After ~ 1ps the distribution spreads and the drift velocity settles at a smaller value as the mobility is reduced at higher carrier temperature T_e. Bottom: Shows the E_x step and the drift velocity as a function of time

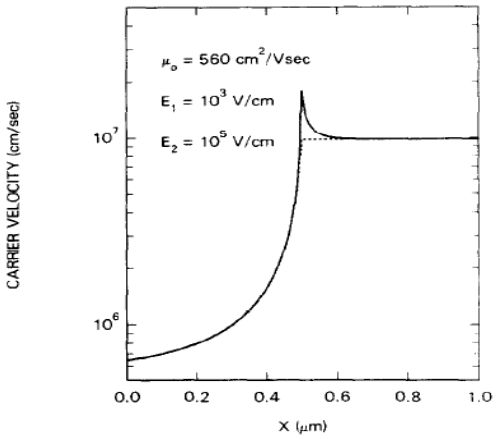


Figure 3.4

Carrier drift velocity as a function of channel distance from source for a field step at 0.5μm . We can see overshoot near the field step exceeding the bulk saturation value. Image from Baccarani and Wordeman [17].

Fig 3.3 on the top gives the distribution before applying electric field, just after applying electric field (drifted Maxwellian but same temperature as before) and heated distribution with temperature T_e after reaching the steady state. In the case of sub μm MOSFETs we have spatial transient of electric field near the drain. The transient velocity over shoot therefore happens near the drain side of the channel. If we look at the high field condition, most of the applied drain voltage drops near the drain region and therefore the electric field is not uniform in the channel. The electric field is very high at the drain side and the field changes very rapidly with distance in a small region near drain in the channel. Therefore, previous assumptions that near drain carriers move at v_{sat} will not hold in the presence of such field transients [14], although simulation by Kobayashi [14] shows that overall impact on 1μ long device is negligible as this effect is localized near the drain edge. The overshoot becomes more important as the channel length reaches 100nm [14, 15].

Simulation results by Baccarani and Wordeman [17] for a step field discontinuity in the channel is shown in the fig 3.4. We can see that velocity at the point of field transient exceeds the bulk saturation value and then settles to the saturation value after a short distance after the field discontinuity. Remember that here we have steady state response in time as field is constant with time and so the current is constant and current continuity applies throughout the channel. Therefore transient response in time and space are different and in space the field transient affects the distribution even before the point of field transient when time steady state current continuity applies. This means that there is a strong gradient in the carrier density around the transient region and therefore associated diffusion currents. The diffusion current increases the average velocity before the transient and decreases the velocity after transient point (as diffusion current will be opposite to drift here). Therefore this has the effect of settling the velocity quickly.

The distance electron travel before the distribution heats up and velocity settles down (in other words energy of carrier reaches equilibrium with the new field) is of the order of 10s of nm [11]. In the case of deca-nanometer transistors, the channel length is comparable to this distance or inelastic mean free path and therefore non-stationary/off-equilibrium transport becomes more pronounced in these devices. Results of simulation [eg: 14] shows that there is a significant deviation in the drain current with velocity overshoot effect compared to v_{sat}. This has an advantage in terms of on current as velocity overshoot effect gives much higher on-current than that compared to saturation effect as the channel length reduces. Therefore scaling is still advantageous even in the 100nm regime. But velocity overshoot effect also has been shown to have the impact of enhancing the DIBL [14] and therefore an increased off-state current for channel lengths ~ 100nm and below.

Ref [21] describes the drain current taking into account the non-stationary transport and models the velocity overshoot using the electron temperature approach. The saturation velocity model of Eq 3.6 is modified by using the energy balance equation with a characteristic mean free length λ_w. The electron temperature reaches a steady state after a few λ_w in a step field. However in the case of sub 100nm MOSFETs the channel length is comparable to λ_w or energy mean free path/relaxation length and so the transport becomes non-stationary and the drain current will be larger than the velocity saturated case. The expression however was derived for the case where β of Eq 3.6 = 1 and a linear potential gradient inside the channel with a field step at source or for slowly varying field with in the channel (so $E_{//} = V_d/L$). The expression 3.6 gets modified with a factor $1+2\lambda_w/L$ to Eq 3.9 for $V_d < V_{dsat}$.

$$v_d(E_{II}, E_{\perp}) = \frac{\mu(E_{\perp})E_{II}}{\left(1 + \frac{\mu(E_{\perp})E_{II}}{v_{sats}\left(1 + 2\lambda_w / L\right)}\right)} \tag{3.9}$$

The expression shows that the reduction of mobility with lateral field $E_{//}$ is slower with the non-stationary effect and becomes significant for L comparable to λ_w. For $L >> \lambda_w$ the equation retrieves the velocity saturation model for the sub μm MOSFETs. Remember that the mobility is still dependent on the lateral field and the saturation drain current is smaller than that we would expect by (see section 5) assuming long channel behavior.

Section 5 discusses extraction methods for extracting the saturation velocity based on Eq 3.6. Section 6 gives v_{sats} vs channel length down to 55nm for FDSON devices. We can see the velocity overshoot effect for channel lengths ~ 100nm and below. This is more pronounced at low temperatures. The drain current down to 55nm seems to be governed by the saturation and overshoot effects rather than the ballistic transport. Remember that in ballistic case the momentum relaxation does not occur in the channel. In the non-stationary case we still have concept of mobility and drift velocity due to momentum relaxation but as energy relaxation doesn't occur within the channel the saturation effect of drift velocity gets suppressed and we get larger drain current.

As the channel length goes below 32nm, the transport is better modeled by quasi ballistic approach. This sets up a maximum drive current limit at full ballistic transport regime beyond which channel length scaling cannot provide on current advantage. Ballistic limit and quasi ballistic drain current expression is explained in the next section.

3.4 Ballistic/quasi ballistic transport model

The ballistic transport in the linear regime was discussed in the section 2. The general expression for quasi ballistic transport in Eq 2.20a can be simplified for large V_d, ($qV_d >> kT$) using the property of the Fermi-Dirac integrals to obtain the saturation drain current expression I_{DSsat}. The expression is shown in Eq 3.10 [24, 30]. The injection velocity $\tilde{v}_{th}$ can be approximated using the properties of Fermi-Dirac integrals at strong degeneracy as eq 3.10c [30] for single sub band case. g_V is the valley degeneracy (2 for the twofold valley). Similar expression for full ballistic case was obtained by K Natori [31] as given in eq 3.10d taking into account multi sub band occupation by introducing a factor of 2.5 instead of g_V assuming 80% of carriers occupy the lowest sub band. As we see eq 3.10c and eq 3.10d are similar with eq 3.10c having an additional factor for back scattering. More detailed expression for multi sub band occupation in the case of quasi ballistic regime is given by V. Barral et al. [30].

$$I_{DSsat} = WC_{ox}\left(V_g - V_t\right)\left[\frac{1-r_{sat}}{1+r_{sat}}\right]\tilde{v}_{th} \tag{3.10a}$$

$$r_{sat} = \frac{l_{kT}}{\lambda_m + l_{kT}} \tag{3.10b}$$

$$\tilde{v}_{th} = B\frac{1}{\left(1+r_{sat}\right)^{1/2}}\sqrt{C_{ox}\left(V_g - V_t\right)}; \qquad B = \frac{8}{3\sqrt{\pi q g_V}}\frac{\hbar}{m^*} \tag{3.10c}$$

$$\tilde{v}_{th} = B\sqrt{C_{ox}\left(V_g - V_t\right)}; \qquad B = \frac{8}{3\sqrt{\pi q M_V}}\frac{\hbar}{m^*} \tag{3.10d}$$

The back scattering coefficient is replaced by r_{sat} and the saturation ballistic ratio BR_{sat} is given by $[1-r_{sat}]/[1+r_{sat}]$. The backscattering coefficient r or r_{sat} is dependent on the gate, drain bias and scattering mechanism. The drain current is controlled by a small portion at the source side known as ‘kT layer’ (region where the channel potential drops by kT/q). The carriers that are scattered with in this layer can return to source contributing to the back scattering. The carrier that crosses this barrier on average will make it to drain. The scattering on the later portion of the channel has the effect of modifying the potential and indirectly controls the drain current [23-26]. A theoretical interpretation is given by Lundstrom and Ren [24]. For elastic scattering case, the carrier scattered will contribute to the back scattering only if it has sufficient momentum in the transport direction so as to surmount the barrier (as we are dealing with 2D channel). In the case of inelastic scattering, the picture is simpler. Since carrier

loses energy during scattering, those that do not have enough energy (after scattering) to return back to the source will not contribute to the back scattering. Again this is further deterred from contributing to backscattering due to the same reason as mentioned for elastic scattering. Therefore in general the scattered carrier should have enough "longitudinal kinetic energy" ie KE in the direction of transport towards source in order to contribute to the back scattering. Therefore the backscattering coefficient r_{sat} can be expressed as eq 3.10b with λ_m the mean free path of linear regime (as field in this region is small) as given in eq 2.21 and a characteristic length L_{kT}. In the case of 1D transport (like nano-wire), elastic scattering contributes to back scattering throughout the channel length. The kT length is similar to Bethe criterion for thermionic-emission [24, 27] except that in MOSFETs the carrier injection is down the barrier and there is back scattering within the layer, as kT layer is comparable to λ_m. In the case of Bethe criterion, $L_{kT} << \lambda_m$ and therefore carriers travel without any scattering in this region up the barrier.

The physical significance of kT layer is studied in ref [23] by solving 1D Boltzmann transport equation using relaxation length approximation. Under the constant field assumption the back scattering coefficient is derived for both low field r_{lin}(or simply r) and high field conditions r_{sat}. The saturation back scattering ratio is expressed using L_{kT} and linear regime mean free path λ_m as given in eq 3.10b. They have also shown that this relation still holds with a minor modification for a parabolic channel potential profile. It is also shown in [25] through MSMC simulation that the back scattering coefficient r_{sat} can be expressed as 3.10b however the mean free path to be used is not same as λ_m. The fit value of λ using L_{kT} was found to be smaller as the back scattering was larger. In the case of analytical expression, we assume Maxwellian for backscattered flux [23] and in this case the mean free path is always λ_m. But simulated carrier distribution in [25] show that the carrier distribution is distorted from being Maxwellian and is heated giving a larger back scattering. However the need for a smaller mean free path was interpreted in [25] as need for a longer characteristic length L_{ceff} ($> L_{kT}$) rather than changing λ_m. L_{ceff} is defined as the centroid of spatial distribution of scattering contributing to r_{sat} is found to be $2.5L_{kT}$ in [25] showing the need for changing L_{kT} to L_{ceff} instead of λ_m. The use of L_{ceff} is somewhat consistent with [24] also where they have attributed L_{kT} as the length where scattering events contributing to r_{sat} is heavily weighted. Finally backscattering coefficient r_{sat} is shown to increase with degeneracy and this is found to be due to decreased λ_m also, apart from increased L_{ceff}. Therefore to conclude from the above, if we use λ_m from linear regime, the characteristic length obtained is not really L_{kT} but L_{ceff} ($> L_{kT}$). However as per [25] due to degeneracy dependences λ_m to be used in eq 3.10b is not so clear. There is a direct correlation between linear regime mobility and λ_m of eq 3.10b and an increase of channel mobility will have a positive impact on the saturation.

The effect of inelastic phonon scattering is said to improve ballisticity in the quasi ballistic MOSFETs as already discussed in section 2 [28]. As explained in previous paragraph, only a carrier with enough longitudinal KE can contribute to the back scattering. The inelastic scattering happens deeper inside the channel (after kT layer) losing energy and these carriers cannot return back to the source. The inelastic scattering is attributed to g-type OP with energy 61meV [28] and that's why this type of scattering cannot happen in the beginning of the channel especially inside kT layer. Therefore they contribute to forward flux and improve ballisticity. However it has been shown [28] that this advantage is lost as channel length increases (> 20nm [28]) since these carriers also has the effect of increasing the channel potential due to charge accumulation and thereby reducing the drain current.

Extraction methods for back scattering coefficient are described in ref [29, 30] and others. Extraction in terms of quasi ballistic model in saturation is not done in this thesis. The extraction using v_{sat} model has been performed for FDSON devices down to 55nm at different temperatures and result shows that performance of these devices can still be explained by v_{sat} and overshoot effects as shown in section 6.

3.5 Conclusion

On current is an important MOSFET performance parameter and it is therefore important to study the saturation regime. This chapter covered the saturation regime MOSFET drain current model for sub μm MOSFETs. The transport in high electric field is governed by high field effects namely saturation velocity in the case of sub μm MOSFETs and hence v_{sat} drain current model is used. The mobility is a function of electric field and is shown to reduce with electric field and is explained analytically in literatures through electron temperature approach. In the case of deep submicron MOSFETs near 100nm regime, transport is governed by non-stationary effects and carrier velocity in this regime can exceed the saturation velocity and is known as velocity over-shoot effect. Origin of this situation is explained by electron temperature approach. Carrier temperature takes a finite time or carrier has to travel a finite distance to heat up and reach a steady state so that v_{sat} effect comes into picture. In the case of channel lengths comparable to this energy relaxation length, carrier velocity can exceed the expected saturation values. Finally in the case of sub 32nm MOSFETs the transport is governed only by small amount of scattering and is better modeled using quasi ballistic theory. The Quasi ballistic and full ballistic theory is discussed in brief with important results along with the kT layer concept for explaining the saturation backscattering coefficient.

Reference

[1] C Jacobini, C Canali, G Ottaviani and A A Quaranta " A review of some charge transport properties of Silicon" Solid-State Electronics vol20 p77 1977

[2] M Saitoh, Ken Uchida "Universal relationship between Low Field Mobility and High Field Carrier velocity in high K and SiO_2 gate dielectric MOSFETs" IEEE IEDM, 10.1109/IEDM.2006.346757, 2006

[3] A. Lochtefield, D. A. Antoniadis "Investigating the relationship between electron mobility and velocity in deeply scaled NMOS via mechanical stress". IEEE EDL vol 22 n 12 p591 2001

[4] C G Sodini, P K Ko and J L Moll "The Effect of High Fields on MOS Device and Circuit Performances" IEEE TED vol 31, n10, p1386, 1984

[5] R W Coen and R S Muller "Velocity of Surface Carriers in Inversion Layers on Silicon" Solid State Electronics vol23 p35 1980

[6] A Modelli and S Manzini "High Field Drift Velocity of Electrons in Si Inversion Layers" Solid State Electronics vol 31 n1 p99 1988

[7] F F Fang, A B Fowler "Hot Electron Effects and Saturation Velocities in Si Inversion Layers" Journal of Applied Physics, vol 41, n4, p1825 1970

[8] V K Arora, M B Das, "Role of Velocity Saturation in Lifting Pinchoff Condition in Long-Channel MOSFET" EL vol25 n13 p820 1989

[9] J A Cooper Jr, D F Nelson, "High-Field drift velocity of electrons at the Si-SiO_2 interface as determined by a time of flight technique" J Appl Phys vol 53, n3, p1445 1983

[10] W Muller, I Eisele, "Velocity Saturation in Short Channel Field Effect Transistors" Solid State Communications, vol 34, p447 1980

[11] R S Huang and P H Ladbrooke "The physics of excess electron velocity in the submicron-channel FETs", Journal of Applied Physics, vol 48, n11, p4791, 1977

[12] J G Ruch "Electron dynamics in Short Channel Field-Effect Transistors" IEEE TED vol 19 n5 p652 1972

[13] G.G Shahidi, D A Antoniadis, H I Smith, "Electron Velocity Overshoot at Room and Liquid Nitrogen Temperature in Silicon Inversion Layers" IEEE EDL vol9 n2 p94 1988

[14] T Kobayashi, K Saito, "Two-Dimensional Analysis of Velocity Overshoot Effects in Ultrashort-Channel Si MOSFET's" IEEE TED vol32 n4 p788 1985

[15] S E Laux, M V Fischetti "Monte-Carlo Simulation of Submicrometer Si n-MOSFET's at 77 and 300K" IEEE EDL vol9 n9 p467 1988

[16] S Takagi, A Toriumi, "New Experimental Findings on Hot Carrier Transport under Velocity Saturation Regime in Si MOSFETs" IEDM p711 1992

[17] G Baccarani, M R Wordeman, "An Inverstigation of Steady-State Velocity Overshoot in Silicon", Solid State Electronics, vol28, n4, p407, 1985

[18] K K Thornber, "Relation of drift velocity to low field mobility and high field saturation velocity", Journal of applied physics vol 51 n4 p2127 1980

[19] C Jacoboni and Lino Reggiani "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials" Rev of Modern Physics, vol 55 n3 p645 1983

[20] R Stratton "Diffusion of Hot and Cold Electrons in Semiconductor Barriers" Physical Rev, vol 126, n6, p2002, 1962

[21] G Baccarani, S Reggiani "A Compact Double-Gate MOSFET Model Comprising Quantum-Mechanical and Nonstatic Effects" IEEE TED vol 46 n8 p1656 1999

[22] A Lochtefeld, D A Antoniadis "On Experimental Determination of Carrier Velocity in Deeply Scaled NMOS: How close to thermal Limit?" IEEE EDL vol 22 n2 p95 2001

[23] R Clerc, P Palestri, L Selmi "On the Physical Understanding of the kT Layer Concept in quasi ballistic Regime of Transport in Nanoscale Devices" IEEE TED, vol 53, n7, p1634, 2006

[24] M Lundstrom, Z Ren "Essential Physics of Carrier Transport in Nanoscale MOSFETs" IEEE TED vol49 n1 p133, 2002

[25] P Palestri et.al "Multi-Subband-Monte-Carlo investigation of the mean free path and of the kT layer in degenerated quasi ballistic nanoMOSFETs" IEDM, 10.1109/IEDM.2006.346940, 2006

[26] M Ferrier, et.al "Saturation Drain Current analytical modeling of Single Gate Fully Depleted SOI or SON MOSFETs in the quasi ballistic Regime of Transport ", IEEE Intl SOI conf proc, p91, 2006

[27] F.Berz, "The Bethe Condition for Thermionic Emission near an Absorbing Boundry" Solid State Electronics, Vol28, n10, p1007, 1985

[28] H tsuchiya, S Takagi "Influence of Elastic and Inelastic Phonon Scattering on the Drive Current of quasi ballistic MOSFETs" IEEE TED vol55 n9 p2397 2008

[29] M Chen et.al "Temperature Dependent Channel Backscattering Coefficients in Nanoscale MOSFETs" IEDM p39 2002

[30] V Barral et al "Experimental Investigation on the quasi ballistic Transport: Part I – Determination of a New Backscattering Coefficient Extraction Methodology" IEEE TED vol56 n3 p408 2009

[31] K Natori, "Ballistic metal-oxide-semiconductor field effect transistor", J. appl Phy, vol 76, n8 p4879 1994

4 Experimental Extraction in Linear Regime

This chapter is dedicated to state of the art extraction methods in the linear regime for the extraction of mobility. Different state of the art linear regime mobility- gate voltage models are being discussed. The impact of gate dependent series resistance in the advanced MOSFETs and a new improved extraction method by taking into account this effect is also discussed in this chapter. Other linear regime extraction methods that does not use any specific mobility-gate voltage models like split CV method for the extraction of μ_{eff}, and also for t_{ox}, L_{eff} and η parameter for effective field and magnetoresistance measurement for extraction of magnetoresistance mobility μ_{MR} are also been discussed. The magnetoresistance measurement in sub 32nm regime MOSFETs needs to be improved to take into account ballistic transport. Therefore an analytical model for magnetoresistance at low magnetic field for ballistic/quasi ballistic MOSFET is being developed and is discussed in this chapter in the context of previous simulation results.

INTRODUCTION:

The MOSFET drain current in linear regime (small drain-source bias) is expressed in general form as given in Eq 4.1.

$$I_{DS} = \frac{W}{L}\mu_{eff}(V_g)Q_{inv}(V_g)V_d$$
$$gm = \frac{dI_{DS}}{dV_g} = \frac{W}{L}V_d\left[\mu_{eff}\frac{dQ_{inv}}{dV_g} + Q_{inv}\frac{d\mu_{eff}}{dV_g}\right] \quad (4.1)$$

The term gm is the transconductance. In order to understand the transport mechanism, we extract the term μ_{eff}, which is the effective mobility and study the same versus channel length, temperature or electric field and separate the contributions of different scattering mechanisms. We already discussed different scattering mechanisms and ballistic transport for advanced MOSFETs in section 2. We can extract the mobility if we know the inversion charge Q_{inv}(C/cm^2), W (channel width) and L (channel length). One way to know the inversion charge is to measure the gate to channel capacitance (split CV) which is discussed in detail in the section 4.2. The inversion charge is given by Eq 4.2 or in high inversion it can be represented as $C_{ox}(V_g-V_t)$, where C_{ox} is the gate oxide capacitance per unit area, also measured from CV measurement as given in the section 4.2 and V_t is the charge threshold

voltage. The width W is normally large and we can directly take the mask defined width itself. However for channel length L we need to take effective channel length L_{eff} which has to be measured, which is also discussed in detail in section 4.2.

$$Q_{inv}(V_g) = \int_{Vin}^{Vg} C_{gc}(V_g) dV_g \tag{4.2}$$

The effective mobility can be described by a generalized mobility model for up to medium electric field and wide temperature ranges from liquid He to room temperatures and is given by Eq 4.3 [13, 14].

$$\mu_{eff}(V_g) = \frac{\mu_g \left[\theta\left(V_g - V_t\right)\right]^{n-2}}{1 + \left[\theta\left(V_g - V_t\right)\right]^{n-1}} \tag{4.3}$$

Equation 4.4 is the standard mobility law most widely used for room temperature and for low electric fields.

$$\mu_{eff}(V_g) = \frac{\mu_0}{1 + \theta\left(V_g - V_t\right)} \tag{4.4}$$

Equation 4.3 can retain this law by putting 'n' = 2. For liquid He temperature it is shown that 'n' should be 3 [14] in order to explain the increase of the mobility for low fields. The value of 'n' is obtained from generalized Y function explained in section 4.1.1.

However for high gate voltages (transverse electric fields ~ 1 MV/cm), the transconductance becomes negative [8, 13, 35] especially at low temperatures. In the case of advanced MOSFETs with t_{ox} ~ 3nm and below this effect is seen even in room temperature [35]. This is attributed to faster mobility roll off due to surface roughness at high gate voltages and this requires modification in Eq 4.3 in order to properly model the same. From Eq 4.1 it is clear that the term responsible for the negative value of gm is the derivative of effective mobility with V_g. This derivative is negative and if it is large enough, it can give a negative gm. This effect is generally attributed to surface roughness at interface between Si and gate oxide SiO_2 (see section 2), since inversion charge pile up closer to the interface with increasing gate voltage. In the high field/inversion charge region, the effective mobility rolls off as square of the electric field (section 2) or inversion charge. This would suggest that we need to include a square dependence in the mobility model. This is shown in the equation 4.5 below [13].

$$\mu_{eff}(V_g) = \frac{\mu_g \left[\theta\left(V_g - V_t\right)\right]^{n-2}}{1 + \left[\theta\left(V_g - V_t\right)\right]^{n-1} + \left[\theta_2\left(V_g - V_{t2}\right)\right]^n}$$

and

$$I_{DS} = \frac{\beta V_D \theta^{n-2}\left(V_g - V_t\right)^{n-1}}{1 + \left[\theta\left(V_g - V_t\right)\right]^{n-1} + \left[\theta_2\left(V_g - V_{t2}\right)\right]^n}; \beta = \frac{W}{L} C_{ox} \mu_g \quad (4.5)$$

The parameter θ_2 is extracted by making use of the expression Eq 4.6. If we plot LHS of Eq 4.6 vs V_g, it will be a linear function of V_g with a second threshold voltage V_{t2} and slope θ_2 [13]. In the absence of θ_2 this will be a constant. Therefore additional parameter V_{t2} is introduced in the second order term and its value is normally higher than the V_t [13].

$$\left[\beta V_D \theta^{n-2} \frac{\left(V_g - V_t\right)^{n-1}}{I_D} - 1 - \left[\theta\left(V_g - V_t\right)\right]^{n-1}\right]^{1/n} = \theta_2\left(V_g - V_{t2}\right) \quad (4.6)$$

The value of V_{t2} and slope θ_2 is a function of gate oxide thickness and as the thickness is reduced V_{t2} is shown to reduce and θ_2 is shown to increase. Parameter θ_2 along with θ is also shown to increase as temperature decreases in this model fit. The effect of series resistance in the source/drain is included inside the term θ [13].

A simpler model is given below in Eq 4.7 which has been shown to be effective down to liquid N_2 temperatures [8].

$$\mu_{eff}(V_g) = \frac{\mu_0}{1 + \eta_1\left(V_g - V_t\right) + \eta_2\left(V_g - V_t\right)^2} \quad (4.7)$$

This model is an improvement over Eq 4.4 to introduce the mobility roll off at high electric field for thin gate oxides and can explain the negative transconductance. However in this model a second threshold voltage is not mentioned [8]. The value of η_1 and η_2 in this model are dependent on the gate oxide thickness as $1/t_{ox}$ and $1/t_{ox}^2$ respectively. The model explained in Eq 4.3/4.5 is an improvement over Eq 4.4/4.7, generalized to all the temperatures and better explains low temperature behavior <77K. However extraction based on this model is more complex compared to Eq 4.4/4.7. The expression in Eq 4.5 will yield the same expression as Eq 4.7 for room temperatures where n = 2. This will have a second threshold voltage as shown in Eq 4.8 [10].

$$\mu_{eff}(V_g) = \frac{\mu_0}{1 + \alpha_1\left(V_g - V_t\right) + \alpha_2\left(V_g - V_t - \Delta V_t\right)^2}. \quad (4.8)$$

This can be simplified and can be finally rewritten as Eq 4.9 [10] to look like the expression in Eq 4.7.

$$\mu_{eff}(V_g) = \frac{\mu_0}{1+\Theta_1(V_g - V_t)+\Theta_2(V_g - V_t)^2}$$
$$\Theta_1 = \frac{\alpha_1 - 2\alpha_2 \Delta V_t}{1+\alpha_2 \Delta V_t^{\,2}}; \Theta_2 = \frac{\alpha_2}{1+\alpha_2 \Delta V_t^{\,2}}; \Delta V_t = V_{t2} - V_t \quad (4.9)$$

If we compare the eq 4.7 and 4.9, we can see that both the expressions are of the same type. The expression for the Θ_1 now shows that the model could yield negative values for the same if $\alpha_1 \langle 2\alpha_2 \Delta V_t$, owing to the second threshold voltage V_{t2}. The negative value of Θ_1 comes to be at low temperatures where there is an increase of mobility for low gate voltages mainly attributed to Coulomb scattering. The expression in Eq 4.5 models this by adding an additional term at the numerator (by putting n >2, 3 for liquid He). The model Eq 4.9/4.7 has been used down to liquid N_2 temperatures. However for temperatures below this, Eq 4.5 [13, 14] provides a better model fit for the effective mobility.

The empirical expression in Eq 4.7/4.9 of effective mobility can be linked to physical mobility model [16, 17]. The mobility model in ref [17] deals with surface roughness. The mobility attenuation parameters are expressed in terms of the surface micro roughness. The expression for the mobility terms for surface roughness model is given in Eq 4.10.

$$\mu_{eff}(V_g) = \mu_0 e^{-\frac{\theta Q_{inv}}{C_{ox}}} = \mu_0 e^{-\theta(V_g - V_t)} \approx \frac{\mu_0}{1+\theta s_1(V_g - V_t)+\theta s_2(V_g - V_t)^2}$$
$$\mu_0 = \mu_b e^{-\frac{q\Delta Q_{dep}}{kT\varepsilon_{Si}}}; \theta = \frac{\eta q \Delta C_{ox}}{kT\varepsilon_{Si}}; \theta s_1 = \theta, \theta s_2 = \theta / \sqrt{2} \quad (4.10)$$

The term μ_b is the bulk mobility, Δ is the characteristic interface roughness within which the mobility of the carriers vanishes [19], η is the weighting factor used in the expression of effective field (1/2 for electrons and 1/3 for holes), Q_{inv} is the inversion charge given by $C_{ox}(V_g - V_t)$ and Q_{dep} is the depletion charge, ε_{Si} is the permittivity of silicon, k is the Boltzmann constant, T is the temperature, q is the absolute electron charge. The exponential term can be expanded to look like Eq 4.7 giving a relationship of the first order and second order parameters in terms of surface roughness coefficient as shown in eq4.10. The suffix ‘s’ is included to separate these coefficients from those used in the models discussed so far and to stress that this is surface roughness coefficients alone and can only take positive values.

However this model cannot explain V_{t2} (Eq 4.8) or negative Θ_1/η_1 (Eq 4.9/4.7) (remember Eq 4.8 will give positive value for α_1 as we have introduced V_{t2}). In order to explain this we need to take in to account other scattering mechanisms like Coulomb especially at low temperatures.

The model used in ref [16] makes use of the semi-empirical formula from [20] for the mobility as shown in Eq 4.11 [19, 20] for room temperature (see section 2 also).

$$\frac{1}{\mu_{eff}(Q_{inv})} = \frac{1}{\mu_C} + \frac{a_{ph}(Q_{inv} + 2Q_D)^{1/n}}{\mu_{ph}} + \frac{a_{SR}(Q_{inv} + Q_D)^2}{\mu_{SR}} \tag{4.11}$$

It is important to note that μ_C is modeled independent of the inversion charge as the screening effect is negligible at high temperatures. For low temperatures (< 77K) μ_C is to be modeled differently and is proportional to inversion charge $(Q_{inv})^a$, (a = 1.6 to 2) (section 2) as the scattering is reduced by screening effect. Therefore for low inversion charge μ_{eff} increases with gate voltage (inversion charge) and is more precisely modeled in Eq 4.5 or manifests as Vt_2 (Eq 4.8) or negative Θ_1/η_1 (Eq 4.9/4.7).

Parameters on the right hand side of the Eq 4.11 are fitted. Q_D is the depletion charge, and $2Q_D$ is for electrons and for holes it should be $3Q_D$ in the phonon term denoted by subscript 'ph', denominators are fit values of respective mobilities and a_{ph} and a_{SR} are fit coefficients for phonon and surface roughness, n = 3 to 6 for phonon term and 3 if only lowest sub band and only intra sub band scattering is considered (acoustic phonon) (see section 2). The final expression for the parameters of mobility model of Eq 4.9 is as shown in Eq 4.12 assuming low doping, and low oxide/interface charges. (Complete expression, if above condition is not met is not shown here) see ref [16].

$$\begin{aligned} \mu_0 &= \frac{\mu_1}{a_{ph} Q_0^{1/n}}; Q_0 = \left(\frac{2n^2 - 3n + 1}{2n^2}\right)^n Q_{av} \\ \Theta_1 &= \frac{2}{n-1}\frac{C_{ox}}{Q_{av}}; \Theta_2 = \frac{\mu_0}{\mu_{SR}} a_{SR} C_{ox}^2 - \frac{C_{ox}^2}{(2n-1)Q_{av}^2} \end{aligned} \tag{4.12}$$

Q_{av} is the result of second order expansion of the term $(Q_{inv} + 2Q_D)^{1/n}$ about an average value Q_{av} within the region of gate voltage of interest to be fit. As we can see the terms Θ_1 and Θ_2 are inversely proportional to t_{ox} and t_{ox}^2 respectively as found in Eq 4.7 [8]. Also Θ_2 is proportional to μ_0 and therefore has the same temperature dependence as μ_0 considering the fact that the rest of the terms are more or less independent of temperature. The ratio Θ_2/μ_0 is proportional to surface roughness mobility alone. This approach is used in ref [21] to characterize surface roughness of FinFet side wall. Another thing to notice here is that, whether we use complete model of ref [16] or its simplified version (Eq

4.12), the term Θ_1 cannot be negative. At low temperatures however we have to include the inversion charge dependence of Coulomb scattering. The presence of series source drain resistance can be included in the Θ_1 as shown in the drain current expression in Eq 4.13.

$$I_{DS} = \frac{\beta V_{gt} V_d}{1 + \Theta_1 V_{gt} + \Theta_2 V_{gt}^2}; \beta = \frac{W_{eff}}{L_{eff}} C_{ox} \mu_0; \Theta_1 = \Theta_{10} + \beta R_{SD}(V_g)$$
$$V_{gt} = V_g - V_t; R_{SD}(V_g) = R_S(V_g) + R_D(V_g) \quad (4.13)$$

$R_{SD}(V_g)$ models series source-drain resistance as a general function of gate voltage. R_{SD} here is in Ω and term βR_{SD} is independent of width assuming that R_{SD} is inversely and β is directly proportional to device width W.

4.1 Extraction based on mobility model in linear regime

Extraction methods in linear regime like Shift and Ratio method [1] and references therein rely on properties, for example constant mobility with channel length, which were valid for large MOSFETs. This is not true in advanced MOSFETs and therefore is not used any longer for parameter extraction. Some of the important state of the art techniques that are explored in the thesis so far for extraction in linear regime for advanced MOSFETs (using Eq 4.13) are discussed in the following sections. These methods provide a fully analytical procedure to fit the experimental data without the need for a complex numerical fitting. Another advantage of methods discussed is that we can extract parameters for individual devices independent of variation of mobility and threshold voltage with different gate lengths. Even though we still need the physical dimensions especially L_{eff} and C_{ox} considering that W_{eff} is large enough, that it can be assumed to be same as mask width, to extract mobility μ_0, we can still directly measure β which could be used to represent μ_0 (see section 4.4). The effect of series resistance and its increasing importance in electrical extraction for the advanced CMOS devices is also being discussed. Improvement to the state of the art Y function method by taking into account the gate voltage dependence of the series resistance R_{SD} was made during this thesis and is explained in section 4.1.3. The extraction of t_{ox} and L_{eff} is explained in the section 4.2.

4.1.1 Y function and McLarty Method

The drain current expression with first order mobility dependence given in Eq 4.4 and constant R_{SD} can be rewritten as Eq 4.14. From this we can generate the Y function as given in Eq 4.14 [7].

$$I_{DS} = \frac{\beta V_{gt} V_d}{1 + \Theta_1 V_{gt}}; \beta = \frac{W_{eff}}{L_{eff}} C_{ox} \mu_0; \Theta_1 = \Theta_{10} + \beta R_{SD}$$
$$V_{gt} = V_g - V_t; R_{SD} = R_S + R_D \quad (4.14)$$

$$Y \equiv \frac{I_{DS}}{\sqrt{gm}} = \sqrt{\beta V_d}\left(V_g - V_t\right) \Rightarrow \Theta_1 = \frac{\beta V_d}{I_{DS}} - \frac{1}{V_g - V_t}$$

As we can see Y function is a simple linear function of V_g with slope giving the term β and gate voltage intercept giving the threshold voltage. In order to accommodate the low temperature behavior of Eq 4.3, Y function can be generalized as Eq 4.15.

$$Y \equiv \left(\frac{I_{DS}^2}{gm}\right)^{1/n} \propto \left(V_g - V_t\right) \quad (4.15)$$

The value of n is fitted in order to obtain linear Y vs V_g plot. Once the value of n is fitted we can extract μ_0 from transconductance peak gm_{max}. The extraction is more complicated using this model [14]. This method is not discussed in detail as we focus on the model of Eq 4.13 for temperature >77K.

The presence of Θ_2 was taken in to account by iteration method of [11]. The Y function gets modified due to the second order mobility attenuation as shown in Eq 4.16. The measured Y function therefore is nonlinear.

$$I_{DS} = \frac{\beta V_{gt} V_d}{1 + \Theta_1 V_{gt} + \Theta_2 V_{gt}^2} = \frac{\beta V_{gt} V_d}{1 + \Theta_{eff} V_{gt}}; \Theta_{eff} = \Theta_1 + \Theta_2 V_{gt}$$
$$Y_{mes} = \sqrt{\frac{\beta V_d}{1 - \Theta_2 V_{gt}^2}}\left(V_g - V_t\right) and\ \Theta_{eff}\left(V_g\right) = \frac{\beta V_d}{I_{DS}} - \frac{1}{V_g - V_t} \quad (4.16)$$

The extraction is done by iteration method as shown in the fig 4.1. An initial value of β and V_t is got from the measured Y function Y_{mes} at low V_g and it is used to calculate Θ_2 which is used to generate a new corrected Y function Y_{corr}. The parameters β and V_t are extracted again and the same operation is repeated until convergence is obtained as shown in fig 4.1. The final corrected Y function Y_{corr} would yield the good value of β and V_t and in turn Θ_1 and Θ_2 from Θ_{eff}.

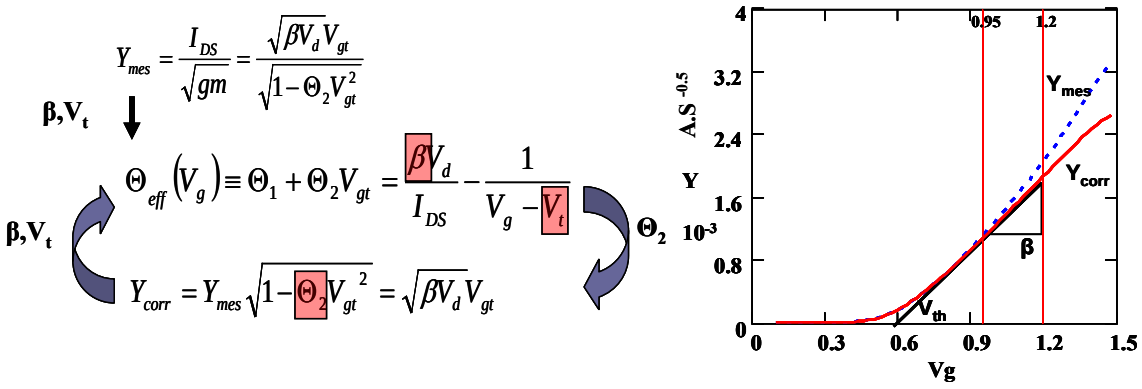


Figure 4.1

Left: Iteration method of ref [11].

Right: The measured Y function Y_{mes} and corrected Y function Y_{corr}.

Alternately, McLarty's method [6] avoids the need for the iteration. This method generates two functions as given in the Eq 4.17 A and B. (The term η shown here should not to be confused with the effective field parameter discussed previously and in the coming sections)

$$\frac{d\frac{1}{I_{DS}}}{dV_g} = \frac{1}{\beta V_d}\left(\Theta_2 - \frac{1}{(V_g - V_t)^2}\right) \qquad (A)$$

$$\eta = \left(\frac{d^2\frac{1}{I_{DS}}}{dV_g^{\,2}}\right)^{-1/3} = \left(\frac{\beta V_g}{2}\right)^{1/3}(V_g - V_t) \qquad (B) \qquad (4.17)$$

These two expressions directly provide the terms β, V_t and Θ_2 without the need for the Iteration method. However this involves second derivatives adding numerical error. We can obtain the Eq 4.17B from Y function also as shown below in Eq 4.18

$$\left(-\frac{d\frac{1}{Y^2}}{dV_g}\right)^{-1/3} = \left(\frac{\beta V_d}{2}\right)^{1/3}(V_g - V_t) = \eta \qquad (4.18)$$

The effect of gate voltage dependence of series resistance modifies Y/McLarty functions and it is discussed in the section 4.1.3. Y and McLarty functions, for FDSOI PMOS long channel MOSFETs is shown in the fig 4.2. Y function is clearly nonlinear due to the Θ_2 effect while McLarty function η is linear for same devices.

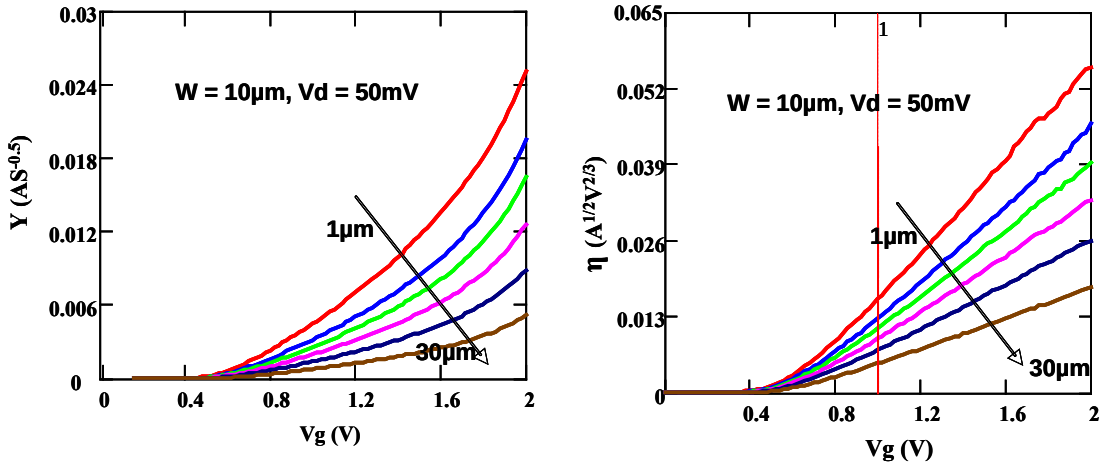


Figure 4.2

Y function and McLarty function η for same set of FDSOI PMOS Long channel MOSFETs is shown. We can see the effect of Θ_2 in the Y function causing nonlinearity while the function η is linear.

A modification to Y function based extraction for taking into account Θ_2 is given in ref [10]. Due to the Θ_2 term, the Y function extraction contains uncertainties in V_t and β. This method uses a different iterative process in order to minimize the error in V_t and β. The expression uses Y function as shown below in Eq 4.19.

$$\xi = \frac{1}{Y^2} = \frac{1}{\beta V_d}\left\{\frac{1}{(V_g - V_t + \varepsilon)^2} - \Theta_2\right\}$$
$$\approx \frac{1}{\beta V_d}\left\{\frac{1}{(V_g - V_t)^2} - \frac{2\varepsilon}{(V_g - V_t)^3} - \Theta_2\right\} \qquad (4.19)$$
$$V_{tinitial} = V_{text} = V_{g\max} - \frac{I_{DS\max}}{gm_{\max}}$$

For large V_g-V_t (linear regime) compared to the error ε in V_t we can expand the terms as shown in Eq 4.19. The terms V_t and ε makes these equations self-consistent. We can start with an initial value of V_t close to the correct value and continue the iteration until convergence is reached where error ε vanishes. Extrapolated threshold voltage V_{text} is generally used in this algorithm as the starting point $V_{tinitial}$. Extrapolated Threshold voltage is taken from maximum gm point, gm_{max} as shown in the Eq 4.19.

Reference book [bal] and references therein give other similar methods, for example YΦ method [15, bal] that involves the function F in Eq 4.20, by several iteration good value of Φ is found for which the function will be linear or its derivative with V_g vanishes.

$$F = \frac{1}{\sqrt{\frac{1}{Y^2} + \Phi}} = \sqrt{\beta V_d}\left(V_g - V_t\right); \Phi = \frac{\Theta_2}{\beta V_d} \tag{4.20}$$

It is very important to notice that the above methods are independent of R_{SD} provided R_{SD} is independent of gate voltage. However for the advanced devices R_{SD} also has to be considered as a function of V_g. This is discussed in the following two sections.

4.1.2 R_{SD} for advanced CMOS devices – gate voltage dependence

As MOSFET channel length is reduced, the influence of series source/drain (R_{SD}) resistance and the error induced by the same in the extraction of other MOSFET parameters becomes more and more important and has to be properly addressed. The series resistance in source or drain region can be divided into four parts as shown in the fig 4.3 below. [22, 23]

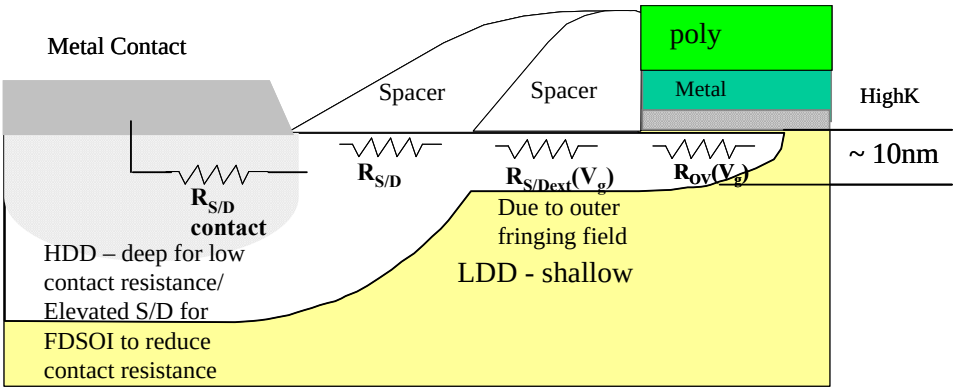


Figure 4.3

Schematic view of the Source/Drain region constituting the series resistance. The region is divided generally into four parts. The LDD extension in source and drain for advanced devices is shallow and constitutes main contribution to R_{SD}. The S/D overlap region works in accumulation and introduce a gate voltage dependence to the resistance. Heavily doped HDD region is designed to be deep enough (or elevated for FDSOI) to reduce the contact resistance.

For long channel bulk MOSFET, the impact of series source/drain resistance was negligible and considered constant with gate voltage. In this case the resistance components are only HDD and a relatively small gate overlap region. Although this will provide only an average value of the series resistance in the gate voltage region under study, it is still valid [1-3 and others] due to the absence of LDD and insignificant gate overlap region contribution compared to rest of HDD. The LDD region, which was introduced for sub μm channel lengths as shown in figure 4.3 to avoid high field effects like avalanche break down at drain and hot electron degradation, introduces higher series resistance. For deca-nanometric devices LDD is made shallow ~10nm [22] in order to avoid DIBL and this introduces even higher series resistance.

The overlap region now plays a bigger role and especially in deca-nanometric scale devices the overlap region is around 30% as long as the channel length [22]. In the case of ultra-thin Si (10nm and less) FDSOI devices the S/D extension LDD plays an important role in the series resistance (The source/drain HDD regions could be elevated to reduce HDD/contact resistances). The Source/Drain gate overlap regions are of great interest as these regions work in accumulation region and introduce strong gate voltage dependence to the series resistance. The LDD under the spacer region also have (but smaller dependence due to current spreading) gate voltage dependence due to the outer fringing fields especially for High K gate where fields are high due to small EOT (< 2nm). The resistance of HDD is also comparable even though it is heavily doped of the order of $10^{21}cm^{-3}$.

The contact resistance has become more significant for advanced devices especially for FDSOI. The specific contact resistivity is determined by the active dopant concentration at silicide/silicon interface and Schottky barrier height [22, 26]. The contact resistance therefore depends on the silicide thickness, doping concentration of the HDD region, the silicide material. The contact resistance reduces with increasing effective contact length up to a point (beyond which increasing the contact length will not reduce the resistance significantly). However the device dimension limitation for advanced MOSFETs put a lot of constraint on the contact size. The thickness of the silicide should be optimized so that maximum active doping concentration is obtained at silicide/silicon interface both for reducing the specific contact resistivity as well as sheet resistance [22]. New silicide material for lowering the barrier height is one way of overcoming the limitations on the maximum doping level and contact depth (in case of FDSOI) and dimensions (Transistor size).

Total S/D resistance is sum of the resistances of the four regions. Each of the regions has more than one parallel resistance components [22]. Overlap resistance R_{ov} is dominated by accumulation currents although it is somewhat reduced by current spreading component. The LDD resistance $R_{S/Dext}$ has an accumulation component due to fringing field but mostly the current is through the bulk LDD. This resistance is further reduced by current spreading as it approaches the deeper HDD region. HDD resistance again can be divided into spreading part (that converges in to the contact) and surface part (up to the silicide thickness depth were current transport is parallel to surface). The expressions in [22] are long and complicated but to first order we can see that R_{ov} and $R_{S/Dext}$ is inversely proportional to gate voltage. However $R_{S/Dext}$ overall sensitivity to gate voltage is small. All of these resistance components have significant contribution to the series resistance with overlap resistance and part of the S/D ext resistance contributing to its gate voltage dependence. This therefore has to be properly addressed in the extraction.

Extraction methods in linear regime like Shift and Ratio method [1] and references therein rely on constant mobility/threshold voltage with channel length. Also R_{SD} is assumed constant. Neither of these is true in advanced MOSFETs and therefore cannot be used any longer for parameter extraction. The ref [2] gives an iterative process which uses several channel lengths (such that they are close to each other and mobility is taken to be constant) and fits a constant R_{SD} and gate dependent effective mobility μ_{eff} (which needs several parameters to be fitted). Constant mobility method of ref [3] does not need prior information on gate oxide capacitance C_{ox}, effective channel length L_{eff}, and channel width W, makes use of the universal mobility curve at high electric field to extract a constant R_{SD} neglecting its gate voltage (V_g) dependence. Also this method is based on the effective field model for bulk MOSFETs and need to be studied properly before using the same for FDSOI (using 0.5 for the weighting factor η (of section 4.3) in $E_{eff,}$ model for example) and other advanced devices. The method of ref [5] requires a constant mobility with channel length to work and so can no longer be used. The extraction methods explained in the section 4.1.1 does not take into account the gate voltage dependence of R_{SD} either. Although McLarty function can eliminate up to linear $R_{SD}(V_g)$ dependence.

The total resistance R_{tot} vs 1/β method [4] does not need any prior knowledge of L_{eff} and mobility-L_{eff} dependence. R_{tot} is defined as the ratio V_d/I_{DS} in the linear regime as shown in Eq 4.21 below.

$$R_{tot} = \frac{V_d}{I_{DS}} = R_{SD}(V_g) + \frac{1}{\beta}\frac{1+\Theta_1(V_g - V_t)+\Theta_2(V_g - V_t)^2}{(V_g - V_t)} \tag{4.21}$$

The threshold voltage V_t and β are extracted using McLarty method discussed before to take into account up to linear gate voltage dependence of R_{SD}. Assuming same V_t and Θ parameters with different channel length we can plot R_{tot} vs 1/β and obtain the R_{SD}.

This method, however, does not take into account the threshold voltage (V_t)-L_{eff} dependence and also limits the extraction up to linear $R_{SD}(V_g)$ relationship. Therefore we need an improved extraction method to study the "real world" devices, especially for taking into account the gate voltage dependence of R_{SD} in nano-scale advanced MOSFETs with High K/ Metal gate structures.

4.1.3 New extraction method - Modified Y function for Advanced CMOS using $R_{SD}(V_g)$

Taking into account the gate voltage dependence of S/D series resistance, state of the art Y function technique is modified and a new extraction method has been developed and it is discussed below [12]. It is easy to show from Eq (4.13) that the measured Y-function Y_{mes} can be expressed as Eq 4.22,

$$Y_{mes} = \frac{I_{DS}}{\sqrt{gm}} = \frac{\sqrt{\beta V_d} V_{gt}}{\sqrt{1-(\Theta_2 + R'_{SD}\beta)V_{gt}^2}} \tag{4.22}$$

The prime denotes derivative with V_g. The term Θ_{eff} is then given by Eq 4.23 and derivative Θ'_{eff} by Eq 4.24.

$$\Theta_{eff}(V_g) \equiv \frac{\beta V_d}{I_{DS}} - \frac{1}{V_g - V_t} = \Theta_{10} + R_{SD}(V_g)\beta + \Theta_2 V_{gt} \tag{4.23}$$

$$\Theta'_{eff}(V_g) = R'_{SD}(V_g)\beta + \Theta_2 \tag{4.24}$$

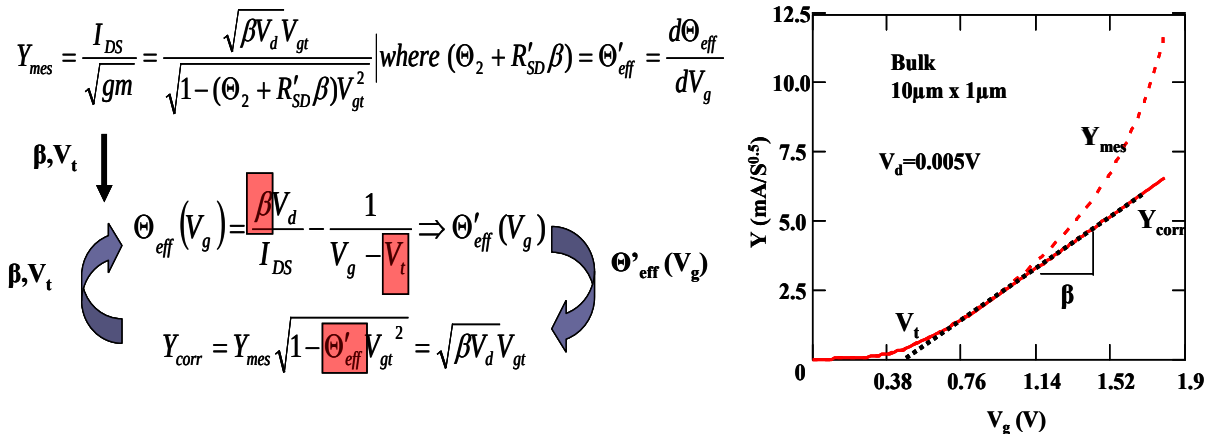


Figure 4.4

Left: The extraction procedure for β and V_t without the effect of the series resistance. Right: The extraction result on Bulk MOSFET 10µm wide and 1µm long device. Y_{mes} shows the effect of Θ_2 and R_{SD}. Y_{corr} is linearized after few iterations.

The extraction procedure is shown in the fig 4.4 (left). The initial β and V_t are extracted from Y_{mes} at low V_{gt} where the high field effect is negligible. The region for extracting the initial values from Y_{mes} for the iteration in a real device must be carefully chosen. The region should not be too close to the threshold voltage while keeping the criteria of low V_{gt}. The initial β and V_t are then used to calculate $\Theta_{eff}(V_g)$. The Θ'_{eff} (Eq 4.24) is used to correct the Y function, Y_{mes}, for large V_g and generate Y_{corr}. The terms β, V_t and $\Theta_{eff}(V_g)$ are re-calculated from the corrected Y function. After few iteration steps, a good linearity is achieved for $Y_{corr}(V_g)$ from which the final values of β, V_t and Θ_{eff} are obtained fig 4.4 (right). This has to be done for a set of devices with different channel lengths and same width.

We could also use β and V_t from McLarty method as the initial values for the iteration, which will simplify the issue of choosing data range. We can easily show that the McLarty function Eq 4.18 also gets modified by the gate voltage dependent series resistance as shown in Eq 4.25.

$$\eta_{mes} = \left(-\frac{d\ \frac{1}{Y_{mes}^2}}{dV_g} \right)^{-1/3} = \left(\frac{d^2 \frac{1}{I_{DS}}}{dV_g{}^2} \right)^{-1/3} = \left(\frac{\beta V_d}{2 + \beta (V_g - V_t)^3 R''_{SD}(V_g)} \right)^{1/3} (V_g - V_t) \tag{4.25}$$

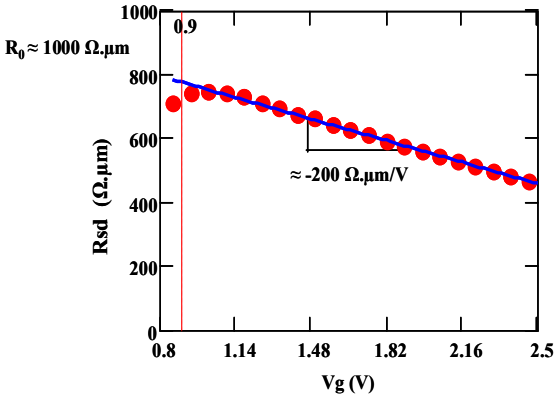
The McLarty function is also nonlinear but here the nonlinearity is due to second derivative of R_{SD} which sometimes vanishes or is small. Also it is a function of β and $V_{gt}{}^3$, and the order of magnitude of this term is 10^{-2} ~10^{-3} for low V_{gt} (< 0.5). For V_{gt} of 0.5 we have $V_{gt}{}^3 = 0.125$, and for 0.4 $V_{gt}{}^3 = 0.064$ and β R"$_{SD}$ ~$10^{-1} - 10^{-2}$). Note that the product βR_{SD} is independent of width. R"$_{SD}$ is in ΩV^{-2}. Therefore for low enough V_{gt} we can safely assume $2 + \beta (V_g - V_t)^3 R''_{SD}(V_g) \approx 2$.

We can see from the fig 4.2, in which the devices have R_{SD} effect too, that McLarty function is much more linear than Y function. Therefore the region from which we could get the initial value is much easier to choose from McLarty function than the Y function. We can choose small enough V_{gt} region to obtain initial values of β and V_t from this, for the iteration steps. Once β and V_t are finalized we can calculate the good Θ_{eff} and Θ'_{eff}. From the Θ'_{eff} vs β plot using eq 4.24 for different V_g, we get Θ_2 from the y-axis intercept and R'$_{SD}$ from the slope. The mean Θ_2 from the plateau of Θ_2 - V_g plot is used to correct Θ_{eff} for each gate voltage using eq 4.26 and by plotting the LHS of eq 4.26 vs β for each V_g, Θ_{10} (y-axis intercept) and $R_{SD}(V_g)$ (slope) are extracted. The plateau for Θ_{10} vs V_g plot is taken as the final value of Θ_{10}.

$$\Theta_{eff}(V_g) - \Theta_2 V_{gt} = \Theta_{10} + R_{SD}(V_g)\beta \tag{4.26}$$

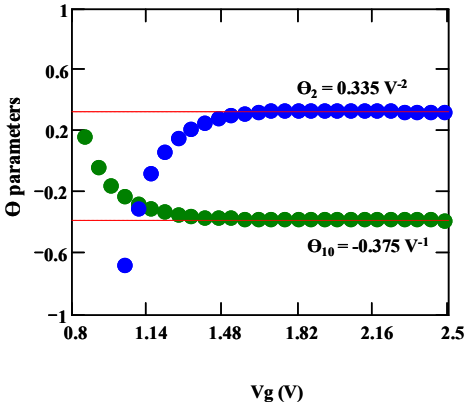
Extraction result on FDSOI PMOS long channel devices (same devices as fig 4.2) is shown in fig 4.5. Devices are obtained from LETI and are made by 3D CMOS sequential integration process, P.Batude et al [48 and references therein]. These devices are studied in terms of access resistance and other parameters. Fig 4.5a shows that the extracted $R_{SD}(V_g)$ is linear for these devices. This also explains the linearity of the McLarty function in fig 4.2. Fig 4.5b shows the plateau obtained for Θ_{10} and Θ_2 at high gate voltages. Fig 4.5d gives mobility μ_0 and threshold voltage V_t for different channel lengths. Fig 4.5c gives the reconstructed I_d-V_g showing excellent fit.

The extraction results using this method obtained for nano-scale (down to 55nm channel length) FDSON MOSFETs (ST Microelectronics) can be seen from ref [12]. The extracted R'$_{SD}(V_g)$ and $R_{SD}(V_g)$ can be used to find a quadratic (or even higher order) fit to the $R_{SD}(V_g)$.



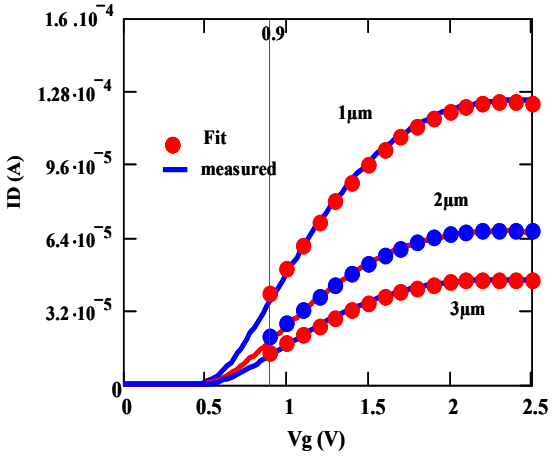
(a)

Extracted R_{SD} (in Ω.µm) vs Vg. In this case the R_{SD} is found to be a linear function of Vg for this set of devices



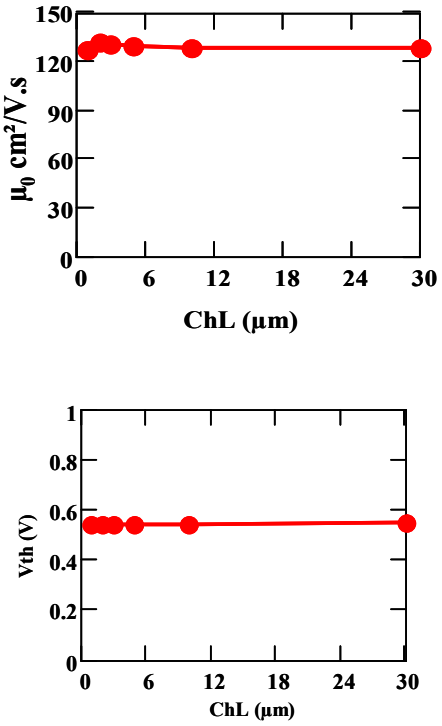
(b)

Extracted Θ_{10} and Θ_2



(c)

Measured Id-Vg (lines) and fit using the extracted parameters (symbols)



(d)

Extracted parameters μ_0 and V_t

Figure 4.5

Extraction results using modified Y function on PMOS long channel devices obtained from LETI [Ref 48]

We can express $R_{SD}(V_g)$, as shown in eq 4.27, as a quadratic function of V_g (and therefore R'_{SD} as a linear function). The result of such fit for the set of FDSON NMOS devices (same as ref [12]) is shown in fig 4.6.

$$R_{SD}(V_g) = AV_g^2 + BV_g + C$$
$$R'_{SD}(V_g) = 2AV_g + B \tag{4.27}$$

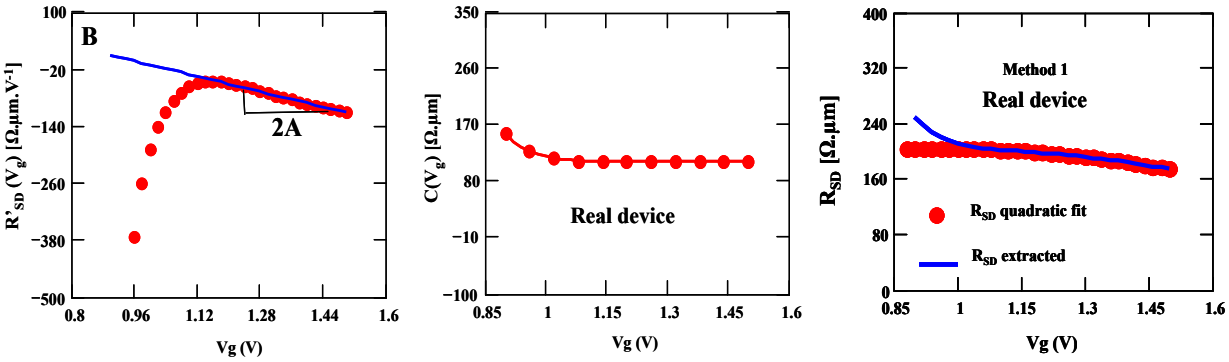


Figure 4.6

Quadratic fit for the $R_{SD}(V_g)$ for FDSON NMOS [12] devices. We can see a linear behaviour for the R'_{SD} with V_g from which we can extract the parameter A and B and from $R_{SD}(V_g)$ we can get the residue $C(V_g)$ which is a constant with V_g in case of quadratic behaviour. Figure on the right end shows the R_{SD} and the quadratic fit obtained using this method.

This method can be simplified by assuming a linear relationship for $R_{SD}(V_g)$.see [12]. The equation for Θ_{eff} is then re-written as shown Eq 4.28. We can use the iteration method as in [12] or McLarty Method for extraction of β, V_t and finally Θ_{eff} in this case as we are using only linear $R_{SD}(V_g)$. Using the expressions in Eq 4.28 we can then separate the terms.

$$\Theta_{eff}(V_g) = \Theta_{10} + R_0\beta + R'_{SD}V_g\beta + \Theta_2 V_{gt}$$
$$\Theta'_{eff}(V_g) = R'_{SD}\beta + \Theta_2$$
$$\Theta_{10} + R_0\beta = \Theta_{eff}(V_g) - R'_{SD}V_g\beta - \Theta_2 V_{gt} \tag{4.28}$$

It is to be noted that all of these extraction methods assume a constant Θ_{10} and Θ_2 for the set of devices under study. The effect of halo implant and neutral defect is to reduce μ_0 with reducing channel length and assumed not to affect Θ parameters. In reality this may not be true. Ref [16] gives Θ_2 as a function of mobility μ_0. Although the results obtained so far shows that we can still approximate Θ_2 to be constant for a given set of devices as seen from figures above.

4.1.4 Impact of Θ_2 dependence with mobility and L_{eff} dependence with gate voltage

Expressions in Eq 4.29 below are slightly modified versions of [16] by the addition of neutral defect (assumed constant) μ_N, and can be easily verified by following the same steps as [16].

$$\frac{1}{\mu_0}=\frac{1}{\mu_C}+\frac{a_{ph}Q_0^{1/n}}{\mu_{ph}};Q_0=\left(\frac{2n^2-3n+1}{2n^2}\right)^n Q_{av};\frac{1}{\mu_C}=\frac{1}{\mu_I}+\frac{1}{\mu_N}$$
$$\Theta_{10}=\frac{2}{n-1}\frac{C_{ox}}{Q_{av}}\left(1-\frac{\mu_0}{\mu_C}\right);\Theta_2=\frac{\mu_0}{\mu_{SR0}}a_{SR}C_{ox}^2-\frac{C_{ox}^2}{(2n-1)Q_{av}^2}\left(1-\frac{\mu_0}{\mu_C}\right) \quad (4.29a)$$

It is important to notice that the factor 1- μ_0/μ_c is retained unlike in ref 16. This is because unlike μ_I, with SiO_2/poly gate, which could be assumed large due to low interface charges, the advanced devices with high K and short channel lengths show additional Coulomb as well as S/D neutral defect scattering [section 2] for short channel lengths and therefore this factor is comparable to μ_0. The depletion charge is taken to be negligible especially in the case of FDSOI and similar devices. Without going too much into detail of this model, to 1st order, the term Θ_2 is a linear function of μ_0.

The variation of 1- $\mu_0/\mu_c = \mu_c(L)/[\mu_{ph}+ \mu_c(L)]$ term with channel length has to be analyzed properly. It can be totally neglected under two conditions. One, if the contribution of μ_c is dominant which would also mean that the term 1- μ_0/μ_c vanishes. Two, μ_c is large (long channel and thicker interfacial layer ~1nm) and μ_0 is dominated by phonon (high temperature), then 1- $\mu_0/\mu_c \approx 1$. In the intermediate stage the sensitivity of 1- μ_0/μ_c with channel length is smaller as the component responsible for the channel length dependence of μ_0 is μ_c.

The presence of halo implants, to higher order, can also modify Θ_{10} and Θ_2 for different channel lengths through the Q_0 and Q_{av} [16] terms in Eq 4.29a which is dependent on the channel doping. As channel length reduces average doping could rise and introduce this higher order channel length dependence (stronger electric fields due to depletion charge, Coulomb scattering). If we consider inversion charge order of magnitudes higher than the doping then we can approximate Q_{av} [see ref 16 for definition of Q_{av}] to depend only on the maximum inversion carrier density Q_{max} as shown in Eq 4.29b. This is easy to understand for FDSOI devices with un-doped channel.

$$Q_{av}=\left[\frac{\int_{Q_{min}}^{Q_{max}}\left(Q+2Q_D\right)^{1/n}dQ}{Q_{max}-Q_{min}}\right]^n-2Q_D\approx\left(\frac{n}{n+1}\right)^n Q_{max};\therefore Q_0\approx\left(\frac{2n^2-3n+1}{2n(n+1)}\right)^n Q_{max} \quad (4.29b)$$

In the case of Bulk MOSFETs, Q_{av} could be comparable to Q_D if we take low inversion region. Remember that eq 4.29b can still be used for Q_{av} but we cannot approximate the Q_0 term (full expression of Q_0 also contains Q_D [16]). One thing to be remembered here is that our discussion in this section concerns only device to device variation among the set of devices under study and not individual values. For the given set of devices, if halo implants do not change the average doping more than one order of magnitude then we can assume Θ_{10} to be constant for the chosen range of channel length and Θ_2 variation can be assumed to originate primarily from μ_0 variation. The order of magnitudes of the charges are, depletion charge (cm^{-2}) Q_D 10^{10} ~ 10^{12}, Q_{max} ~ 10^{13}, Q_{av} ~ 10^{12}). The dependency of Q_{av} and Q_0 with Q_{dep} is shown in fig 4.7 for maximum Q_{inv}, Q_{max} ~ $2x10^{13}$. For 1.8nm gate oxide and maximum V_{gt} of 1.5V, Q_{max} will be ~ 1.7 x $10^{13}cm^{-2}$. We can see that its sensitivity to Q_D range up to $10^{12}cm^{-2}$ considering order of magnitude is negligible. For depletion thickness of 10nm, Q_D of $10^{10}cm^{-2}$ corresponds to doping of $10^{16}cm^{-3}$.

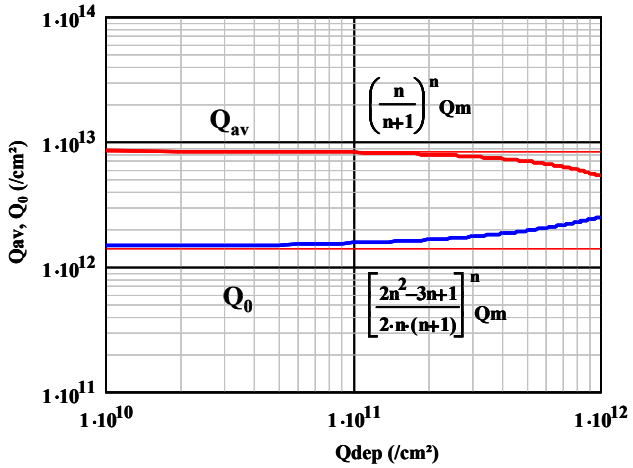


Figure 4.7

Figure shows the variation of Q_{av} and Q_0 with the depletion charge simulated using the complete expression in ref 16. The value for Q_m is taken as $2 .10^{13}$ cm^{-2}.n = 3. The depletion charge Q_{dep} from 10^{10} to 10^{12} cm^{-2} is shown. We can see that for Q_{dep} up to 10^{11} cm^{-2} we can use the expressions of 4.29b for Q_{av} and Q_0 and they are virtually insensitive to the depletion charge. This would mean that we can neglect the impact of halo implants (within a set of devices).

To sum up the above discussion we can safely assume Θ_{10} to be constant within the set of channel lengths under study. The variation between channel lengths can be taken into account in μ_0 due to Coulomb and neutral defects. The variation of Θ_2 due to μ_0 may be significant and has to be carefully handled. This will not affect the extraction of β and V_t and therefore Θ_{eff}. However this would affect the extraction methods for R_{SD}, Θ_1 and Θ_2. A plot of Θ'_{eff} with β (for different channel lengths and constant width) will be then nonlinear. From Eq 4.29c, as β increases with reducing effective channel length, the slope Λ decreases.

$$\Theta_2 = \Theta_{20} + \lambda\mu_0(L_{eff}); \lambda = \frac{\alpha_{SR}}{\mu_{SR0}} C_{ox}^2$$
$$\therefore \Theta'_{eff} = \Theta_{20} + \Lambda(L_{eff})\beta(L_{eff}) + R'_{SD}\beta(L_{eff}); \Lambda = \lambda \frac{L_{eff}}{C_{ox}W} \quad (4.29c)$$

EXTRACTION IN BULK – EXTRACTION OF μ_0 AND μ_{SR} WITHOUT THE NEED FOR L_{eff}

In order to accommodate Θ_2 variation into the extraction method we can take advantage of Eq 4.29c. The derivative of the Θ'_{eff} for each channel length can then be written as in Eq 4.30a.

$$\Theta'_{eff}(V_g) = \Theta_{20} + \lambda\mu_0 + R'_{SD}(V_g)\beta \quad (4.30a)$$

We can assume the terms λ, Θ_{20} and R_{SD} to be constant for the set of devices of varying channel lengths. We can take three long channel length devices where W_{eff}, L_{eff} and C_{ox} are known and therefore μ_0 is known from β and fit the three unknown parameters using Eq 4.30a and Eq 4.26. Once these parameters are known, we can use them for short channel devices to extract μ_0 either from Eq 4.29c or Eq 4.26. Also we see from Eq 4.30b that the term λ gives directly the surface roughness mobility.

$$\mu_{SR} = \frac{\mu_{SR0}}{\alpha_{SR}(Q_{inv} + Q_{dep})^2} \approx \frac{1}{\lambda(V_g - V_t)^2} \quad (4.30b)$$

This needs to be experimentally verified.

Extraction on FDSOI devices:

The experimental results for FDSOI and FDSON with T_{si} ~ 10nm show that Θ_{10} is negative and Θ_{20} positive fig 4.5 and [12] respectively and others, unlike classical bulk MOSFET case. The negative Θ_{10} cannot be explained therefore by these expressions alone and is not yet being fully understood. However for analytical purpose this procedure is still followed and is not expected to affect the issues we addressed above. This is substantiated by the results obtained in fig 4.5, 4.6 above and [12]. The physical meaning of Θ parameters in this case is therefore not known but mobility is still valid.

IMPACT OF $L_{eff} - V_g$ DEPENDENCE

The effect of L_{eff} gate voltage dependence will make the Y function method even more complicated. We have to rewrite the drain current expression as well as Y function as given in the derivation Eq 4.31.

$$I_{DS} = \frac{\beta(V_g)V_{gt}V_d}{f\Theta};\ where\ \beta(V_g) = \frac{C_{ox}W\mu_0}{L_{eff}(V_g)};\ f\Theta = 1 + \Theta_1 V_{gt} + \Theta_2 V_{gt}^2; V_{gt} = V_g - V_t$$

$$gm = \frac{dI_{DS}}{dV_g} = \frac{f\Theta\left[\beta(V_g)V_d + \beta'(V_g)V_{gt}V_d\right] - \beta(V_g)V_{gt}V_d\, f\Theta'}{f\Theta^2}$$

$$Y_{mes} = \frac{\sqrt{\beta(V_g)V_d}\,(V_g - V_t)}{\sqrt{1 - \Theta'_{eff}V_{gt}^2 - \frac{L'_{eff}}{L_{eff}}V_{gt}\,f\Theta}};\ \Theta'_{eff} = \Theta_2 + \beta' R_{SD} + \beta R'_{SD};\ \frac{\beta'}{\beta} = -\frac{L'_{eff}}{L_{eff}} \tag{4.31}$$

$$\approx \frac{\sqrt{\beta(V_g)V_d}\,(V_g - V_t)}{\sqrt{1 - \Theta'_{eff}V_{gt}^2}};\ L_{eff} >> L'_{eff}V_{gt}\,f\Theta$$

Important thing to notice in Eq 4.31 is that μ_0 is still a constant even though L_{eff} is modeled as a function of V_g. This is because μ_{eff} is dependent on the device's physical dimensions (due to proximity of halo, S/D, defects near S/D, strain etc) and its gate dependence is modeled in function $f\Theta$ for different scattering mechanisms and not on the actual channel length itself. We can model a linear dependence of L_{eff} with V_g to see a simplified expression as given in the Eq 4.32 for the Y function. L_{eff0} is the effective channel length at the initial point in the range of gate voltage of fitting.

$$Y_{mes} \approx \frac{\sqrt{\beta(V_g)V_d}\,(V_g - V_t)}{\sqrt{1 - \Theta'_{eff}V_{gt}^2}};\ L_{eff} = L_{eff\,0} + \ell V_{gt}$$

$$\beta = \frac{C_{ox}W\mu_0}{L_{eff\,0} + \ell V_{gt}} = \frac{C_{ox}W\mu_0}{L_{eff\,0}}\left(1 + \frac{\ell V_{gt}}{L_{eff\,0}}\right)^{-1} \approx \beta_0\left(1 - \frac{\ell V_{gt}}{L_{eff\,0}}\right) \tag{4.32}$$

$$\therefore Y_{mes} \approx \frac{\sqrt{\beta_0 V_d}\,(V_g - V_t)}{\sqrt{1 - \Theta'_{eff}V_{gt}^2}}\sqrt{\left(1 - \frac{\ell V_{gt}}{L_{eff\,0}}\right)}$$

We can see that from the Y function we extract more or less β_0 and V_t. The error in the evaluation of Θ_{eff} in Eq 4.23 using β_0 is given in Eq 4.33.

$$\Theta_{effeval}(V_g) \equiv \frac{\beta_0 V_d}{I_{DS}} - \frac{1}{V_g - V_t} = \left(\Theta_{eff} + \frac{\ell}{L_{eff}}\right)\left(1 + \frac{\ell V_{gt}}{L_{eff\,0}}\right) \tag{4.33}$$

We can rewrite the expression for Θ'_{eff} as shown in equation 4.34.

$$\begin{aligned}\Theta'_{eff} &= \Theta_{20} + \Lambda(L_{eff})\beta(L_{eff}) + R'_{SD}\beta(L_{eff}) + \beta'(L_{eff})R_{SD}; \\ &= \Theta_{20} + \left[\Lambda(L_{eff}) + R'_{SD} - \frac{L'_{eff}}{L_{eff}}R_{SD}\right]\beta\end{aligned} \tag{4.34}$$

From these expressions it is clear that the impact of $L_{eff}(V_g)$ is generally weighted by the ratio L'_{eff}/L_{eff} and this is neglected for practical purposes considering range of gate voltage of fit.

4.2 C-V method – Q_{inv}, L_{eff}, t_{ox}

In order to extract the mobility μ_0, we need to know inversion charge Q_{inv}, which in the previous section is taken as C_{ox} (V_g-V_t). V_t can be extracted as discussed above as well as the gain factor β. However we still need to know t_{ox} as well as L_{eff} to extract the μ_0 from β. Sub sections below discusses some of the state of the art extraction methods using C-V technique for the same.

4.2.1 Equivalent Oxide thickness t_{ox}

The oxide thickness t_{ox} used for C_{ox} gate capacitance per unit area defined as $\varepsilon_{SiO2} / t_{ox}$, seen in the drain current expression, can be measured from the C-V measurement. The term ε_{SiO2} is permittivity of SiO_2. It is to be noted here that t_{ox} value for calculating C_{ox} is different from the physical SiO_2 thickness t_{SiO2}. This is known as equivalent oxide thickness (EOT) denoted here as t_{ox}. This is due to the fact that inversion carriers have a finite thickness as well as for classical devices using the poly Si gate, we need to take into account the poly depletion effect (for thin oxides <10nm). Both have the effect of increasing this effective thickness t_{ox}. Therefore C_{ox} gets reduced accordingly, reducing I_{on} and thereby under estimation of mobility in measurements. In the case of advanced MOSFETs that needs t_{ox} ~ 2nm and less we use high K / metal gate stack. In this case we don't have the effect of poly depletion, but we still have inversion charge depth which is now comparable to the thickness. Also the gate stack have SiO_2 layer between the high K and Si making it a complex gate stack. Even if we are using high K, we still represent the oxide thickness in terms of SiO_2 as $t_{ox} = [\varepsilon_{SiO2}/ \varepsilon_{HK}]\, t_{HK}$ in order to conform to the classical MOSFET models and definitions and to have easier comparison with scaling requirements. Because the permittivity of high K, $\varepsilon_{HK} > \varepsilon_{SiO2}$, its SiO_2 equivalent thickness (EOT) $t_{ox} < t_{HK}$. This way the physical thickness of high K t_{HK} can be higher, thereby reducing the leakage, while achieving the requirement of thin t_{ox} <2nm of the technology node. Because of these reasons C-V measurement is necessary in order to determine t_{ox}.

In the case of bulk MOSFETs, C-V measurement from accumulation region was used to extract C_{ox}, doping concentration and flat band voltage V_{fb} from the Maserjian function technique [36] or Vincent et.al [37] or $1/C^2$ vs V_g method [37] and others in ref there in. However in the case of FDSOI devices, this is not possible. In this case we can make use of part of split CV technique by measuring the gate to channel capacitance C_{gc} in the inversion regime, sometimes referred to as $C_{sg}+C_{dg}$, the sum of gate to drain and gate to source capacitance [38 - 41]. In this case we connect the capacitance-meter 'High' to gate and 'Low' to Source and Drain. The substrate is biased to various voltages V_{Sub} as needed (normally grounded). The gate voltage is swept from the accumulation to high inversion obtaining a C-V plot.

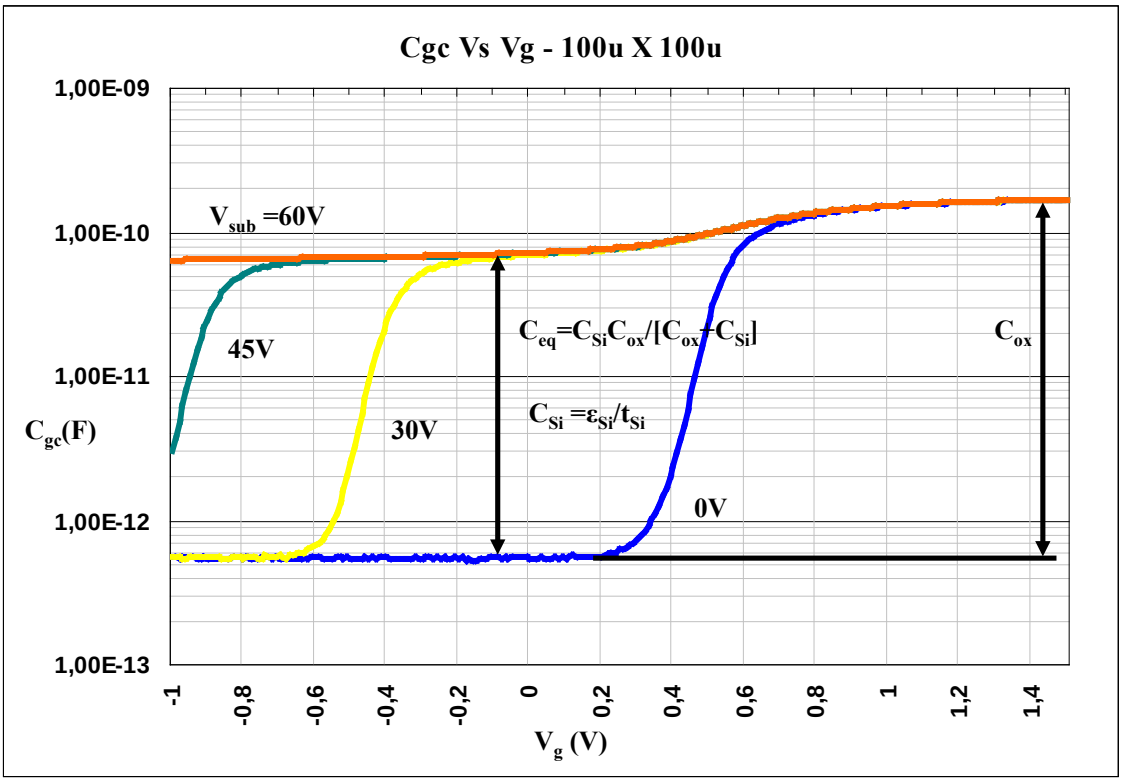


Figure 4.8

Figure shows gate to channel capacitance of FDSOI nMOS 100µmx100µm for different substrate bias V_{Sub}. The equivalent oxide thickness t_{ox} is obtained from the high inversion region for V_{Sub} 0V. The accumulation region for V_{Sub} = 0 is dominated by parasitic capacitance and will be subtracted from the C_{gc} maximum point in the high inversion to obtain the correct value of C_{ox}. Silicon thickness t_{Si} is obtained from C_{eq} by biasing the substrate enough to form inversion at the back interface as shown.

The overall equivalent thickness t_{ox} can be directly measured using C_{gc} using eq 4.35.

$$t_{ox} = \frac{WL}{C_{gcmeas}} \varepsilon_{SiO_2} \tag{4.35}$$

The product WL is the gate area. This is shown in fig 4.8 for the FDSOI case. C_{gc} at high inversion is corrected by subtracting C_{gc} in depletion region. C_{gc} in the depletion region (V_{sub} = 0) is the parasitic capacitance of the overall measurement setup independent of gate voltage (to first order). Fig 4.8 for FDSOI extracts t_{ox} as well as t_{Si} (Si thickness). The device has 145nm BOX, t_{si} ~ 10nm, large MOSFET 100µm x 100µm is used for the extraction in order to avoid higher order parasitic effects (gate dependence). However if back channel is activated by applying V_{Sub}, then we obtain capacitance in the negative gate voltages also denoted as $C_{eq,}$ from this silicon thickness t_{Si} can be obtained as shown in fig 4.8. The equivalent oxide thickness t_{ox} obtained using this method is 2.1nm and the silicon thickness from C_{eq} shown in the fig 4.8 is 9.88nm which is close to the expected value of 10nm.

4.2.2 Effective Channel length L_{eff}

Different terminologies for the channel length are used in different contexts namely L_{mask}, L_{poly}, L_{met} and L_{eff}. L_{mask} is the channel length designed on MASK. L_{poly} is the actual poly gate length after the process. L_{met} is the channel length below the gate stack. This is the physical distance between the source and drain extensions (LDD) after the process. This is different from the poly gate length due to the diffusion of the S/D dopant beneath the gate (gate overlap). L_{eff} is the electrical or effective channel length where the actual inversion layer is formed and is used in the drain current expression. This is shown in the fig 4.9.

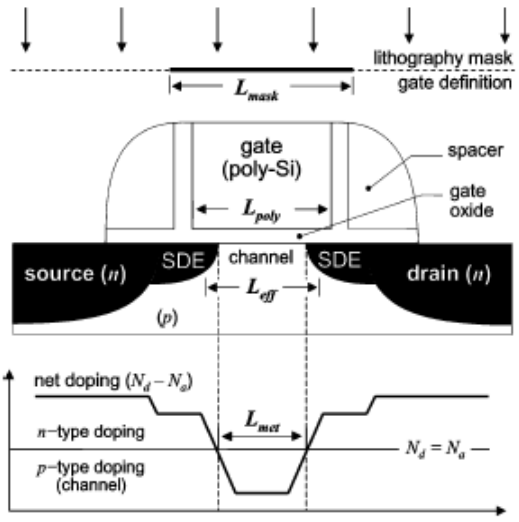


Figure 4.9

Definition of different channel lengths. Image taken from ref [40]

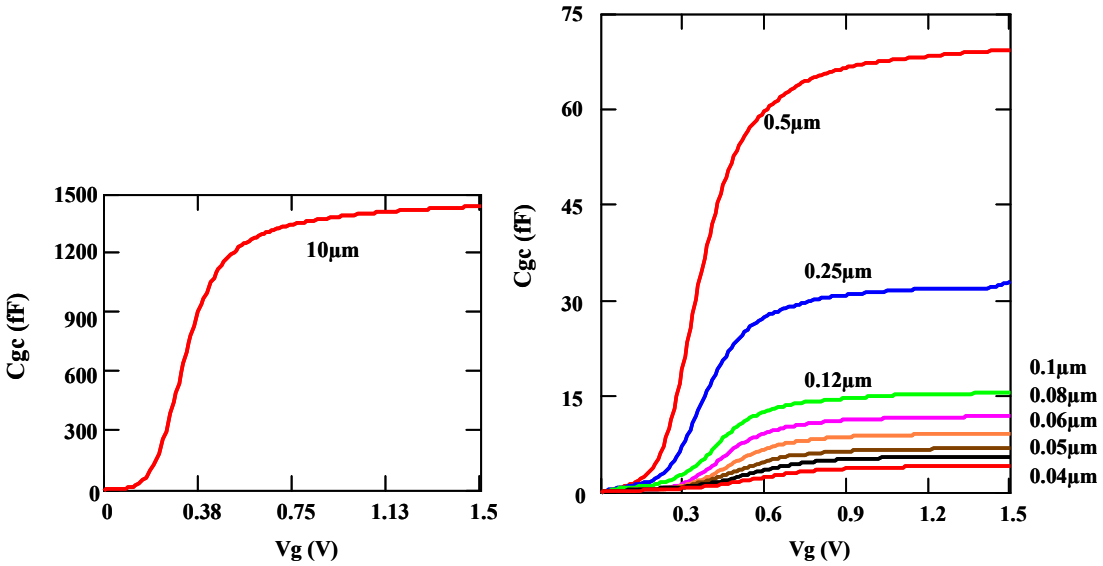


Figure 4.10

Gate to channel capacitance C_{gc} plot for bulk nMOSFETs with width 10µm and channel lengths from 10µm down to 40nm.

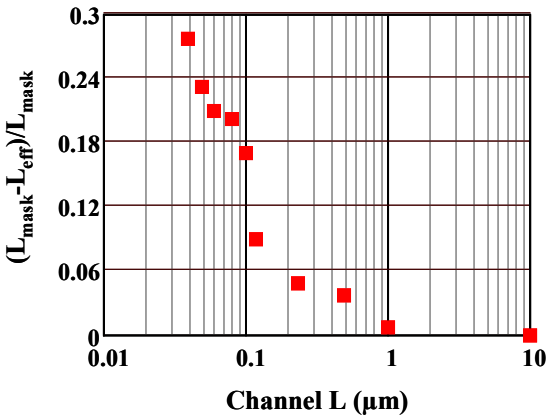


Figure 4.11

Relative difference between L_{mask} and L_{eff} plotted vs channel length. As we see for short channel devices <100nm, the difference in L_{Mask} and L_{eff} can be ~ 20-30%. This uncertainity in L_{eff} makes the mobility extraction difficult in short channel devices

In order to extract the effective channel length we can make use of C_{gc}, once t_{ox} is identified from the long channel device. Assuming uniform t_{ox} for the set of devices measured with constant width (large), the measured C_{gc} is proportional to L_{eff}. Therefore we can choose a reference device with known channel length and extract the effective channel length for sub µm devices using eq 4.36. L_{ref} is reference channel length of long channel device and is taken to be same as L_{mask}. Advantage of this method is that it is independent of R_{SD}, mobility and threshold voltage variation with channel lengths.

$$L_{eff} = \frac{C_{gc}}{C_{gcref}} L_{ref} \tag{4.36}$$

The gate to channel capacitance measured for a set of bulk nMOSFETs with channel length from 10µm down to 40nm and width 10µm is shown in fig 4.10. This shows how C_{gc} varies with channel length. Fig 4.11 shows the relative difference between L_{mask} and L_{eff} plotted against channel length. We can see the significance of effective channel length for deca-nanometric devices. For 40nm device the relative difference is around 30%, $\Delta L = L_{mask}-L_{eff}$ is ~ 10nm. For long channel lengths this difference is negligible. We can also see that the measured capacitance is very small for deep sub µm regimes and therefore the measurement noise and errors due to parasitic capacitances will also be larger. This is discussed in the next section.

4.2.3 Effect of parasitic capacitance on C_{gc} for short channel lengths (~ 200nm and below)

In the extraction mentioned above, we have, to first order, removed major parasitic capacitances by taking the difference between high inversion and accumulation regions. This is however not enough for measuring small capacitance as it is comparable to the fringing and overlap capacitances within the MOSFETs [40, 41].

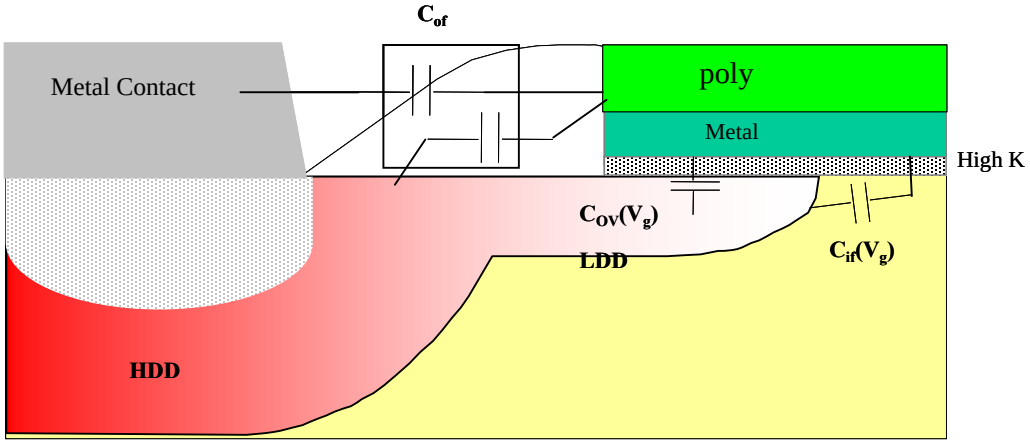


Figure 4.12

The three parasitic capacitances within the MOSFET shown in the above figure. C_{if} is the inner fringing capacitance due to electric field through the depleted channel and therefore vanishes at inversion, C_{ov} is the overlap region capacitance and reaches a constant value at V_g >0 as LDD will be in strong accumulation and C_{of} is the constant outer fringing capacitance between gate and S/D due to the outer fringing field from the gate through the spacer terminating at S/D.

The parasitic capacitances within the MOSFETs are divided into three. The inner fringing capacitance C_{if} is a function of gate voltage, overlap capacitance C_{ov}, which is also a function of gate voltage, and outer fringing capacitance C_{of} which is independent of gate voltage. These are shown in the fig 4.12. The inner fringing capacitance C_{if} is due to electric field between gate and S/D through the Si in the depletion region. This gets screened out by the inversion layer in the inversion region and therefore vanishes in the inversion region. The overlap region capacitance C_{ov} is between gate and the LDD overlap region and because this is in strong accumulation it will be at its maximum at inversion region.

But this region is also a function of gate voltage in the accumulation region. The outer fringing capacitance is due to the electric field between gate and S/D through the spacer. This capacitance is generally independent of gate voltage. C_{ov} will be parallel to the C_{gc} at high inversion and will be at its maximum. In the accumulation region it is parallel to C_{if} and function of gate voltage.

The fig 4.13 shows the measured $C_{gc} - V_g$ plot for short channel devices from 100nm down to 40nm. The region marked shows the fringing capacitances below threshold voltage. The fig 4.13 on the left shows the correction made by taking V_g=0 point as reference point for the extraction of the effective channel length. By doing so we are assuming that the effect of C_{if} (V_g) is small in this region [41] where C_{ov} has reached its maximum as LDD region is in accumulation and is constant and so is C_{of}. Fig 4.13 right shows that the capacitance is not a saturated parasitic value but a gate voltage dependent component mainly due to C_{if}. C_{gc}-V method for extraction of Q_{inv} and consequently effective mobility μ_{eff} is therefore limited to long channel devices where this higher order parasitic capacitance effects are negligible.

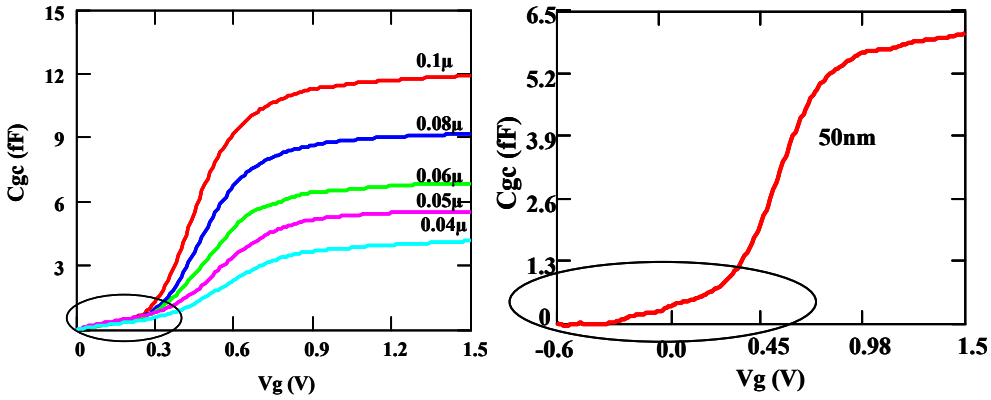


Figure 4.13

Left: The uncertainity in determining C_{gc} for short channel device (<100nm) is shown. Right: $C_{gc} - V_g$ plot for 50nm device. The parasitic capacitance dominates the accumulation region of the plot. The max C_{gc} is only a few fF and therefore uncertainity introduced is more significant. Also the noise in the meaurement also introduces even bigger uncertainity while measureing such small capacitances.

The parasitic region varies slowly with channel length, which means that for close enough dimensions we can take them to be the same. This is seen in fig 4.13 left where different channel lengths have almost same capacitance in the circled region. This has been made use in the ref [39] to modify the split C-V method by subtracting two C_{gc} of adjacent channel lengths (who's effective length is known by previously discussed method) and obtaining a corrected C_{gc} without the effect of C_{if} using Eq 4.37.

$$C_{gc}(L_{eff\,1/2}) = \frac{C_{gc1} - C_{gc2}}{L_{eff\,1} - L_{eff\,2}} L_{eff\,1/2} \tag{4.37}$$

The threshold voltage variation can be neglected for two close enough channel lengths. The inversion charge Q_{inv} is then used to obtain the μ_{eff}. This is shown in fig 4.14 below. In this figure the measured C_{gc} and its corrected C_{gc} for channel lengths 100nm and 80nm is shown. The effective channel lengths are 83nm and 64nm respectively calculated using Eq 4.36. Remember that C_{ov} and C_{of} along with external parasitic capacitance C_p are removed to first order by taking the difference between C_{gc} in high inversion and C_{gc} at V_g =0. The extraction of effective mobility is discussed in the next section.

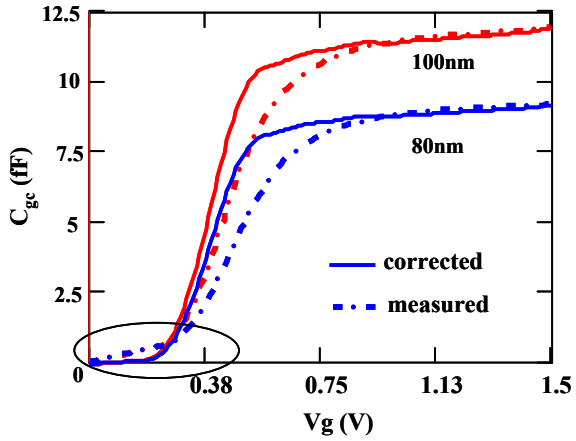


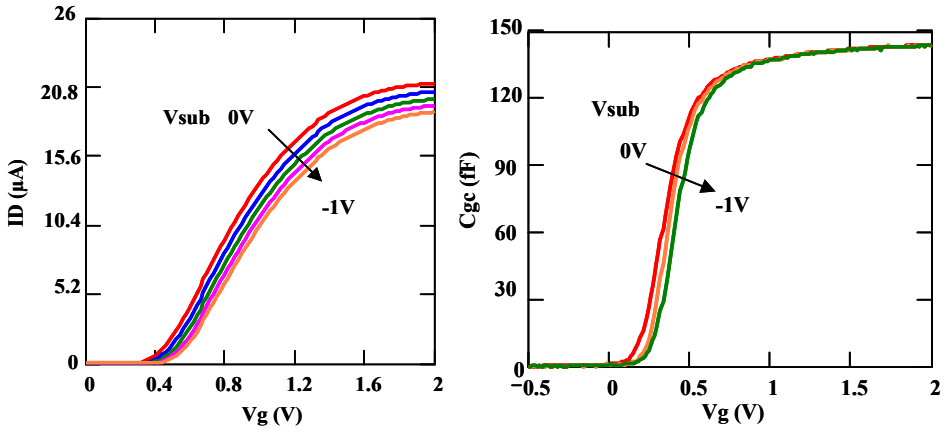
Figure 4.14

The measured C_{gc} and corrected C_{gc} using ref [39] for short channel devices 100nm and 80nm. Because these two are close by channel lengths the variation in threshold voltage can be neglected

4.3 Effective mobility from Split C-V and η parameter extraction

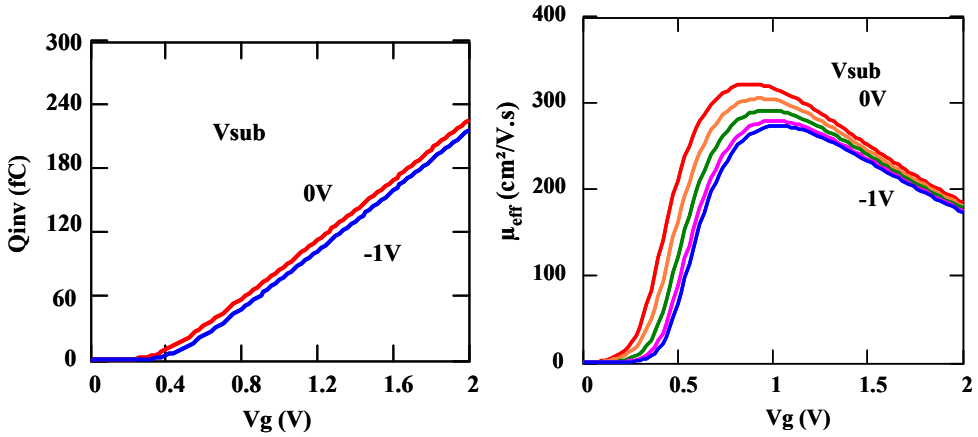
In the previous section we discussed the measurement of t_{ox} and L_{eff} and also parasitic capacitance effect in short channel devices. The gate to channel capacitance for long channel devices where the parasitic capacitance can be neglected can be used to obtain inversion charge directly as discussed in the beginning of the chapter Eq 4.2. The extraction result for bulk, high K/metal devices 10µm wide and 1µm long device is shown in fig 4.15. The effective mobility is studied against the effective field E_{eff} in order to understand the scattering mechanisms as discussed in section 2. The effective field is calculated by knowing the inversion charge Q_{inv} from C_{gc} measurement and depletion capacitance Q_{dep} calculated by knowing the doping concentration N_A, which can be extracted from the Maserjian function [36] or from static method by studying the V_t variation with substrate bias [Sze]. Alternatively it can be found from the split CV gate to bulk capacitance C_{gb} measurement and flat band voltage extracted from the Maserjian function [36, 43]. In the case of FDSOI depletion charge can be neglected for un-doped channel. For the bulk devices discussed above the doping concentration N_A obtained is ~ 4 x $10^{17}$$cm^{-3}$ from the C-V Maserjian technique. The doping level obtained from the

sensitivity of V_t with substrate bias V_{sub} gives a smaller value of ~ 1.5 x $10^{17}cm^{-3}$. The order of magnitude of results from both the methods is same.



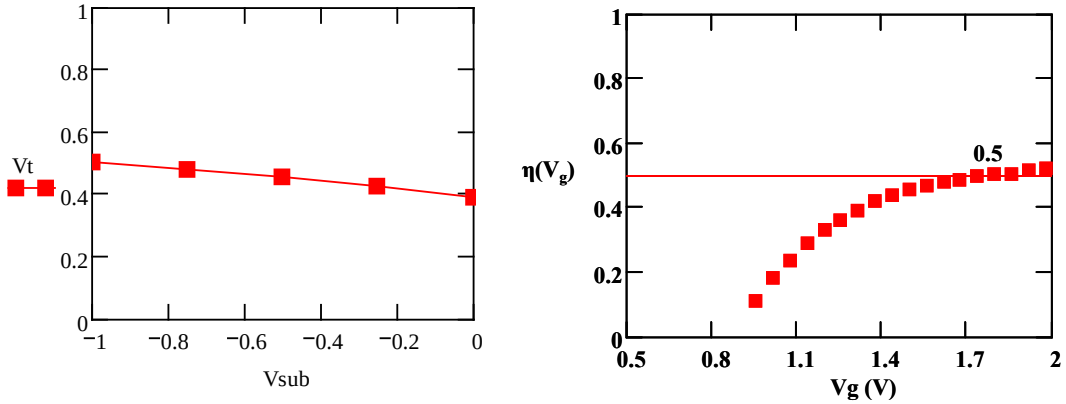
(a) (b)

(a) I_D -V_g for different substrate bias from 0V to -1V. (b) C_{gc} measured for the same substrate bias



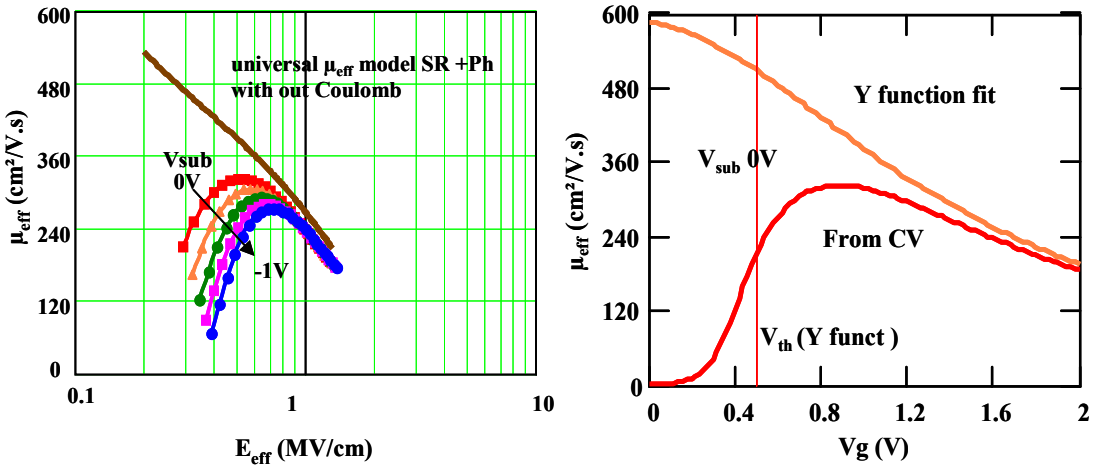
(c) (d)

(c) Q_{inv} -V_g from C_{gc} V_{sub} from 0V to -1V. (d) μ_{eff} – V_g using Q_{inv} and I_D for different V_{sub}



(e) (f)

(e) V_t –V_{sub} plot for obtaining ratio of C_{dep}/C_{ox} for the calculation of η. (f) η – V_g obtained shows plateau at 0.5 at high V_g



(g) (h)

(g) μ_{eff} $-E_{eff}$ for different V_{sub}. (h) μ_{eff} $- V_g$ from Y-function fit and CV method for $V_{sub} = 0$

Figure 4.15

Figures a to h shows the extraction results on the Bulk High-K Metal gate device 10µm wide and 1µm long. Comparison between μ_{eff} from CV method and Y-function fit is also shown

In this result the value obtained from the static method is used for the E_{eff} calculation however no big difference is seen by using the value of N_A from the Maserjian function technique. The method described in [42-44] is applied to extract the η parameter using the expression Eq 4.38a explained below.

The fig 4.15 (a) and (b) gives the I_D-V_g and C_{gc} for different substrate bias V_{sub}. Fig 4.15 (c) and (d) are Q_{inv} obtained from C_{gc} plot and μ_{eff} calculated from I_D and Q_{inv}. Fig 4.15 (e) is V_t vs V_{sub} made use to extract the C_{dep}/C_{ox} ratio from the slope as well as for calculating N_A [Sze]. Fig 4.15 (f) is the parameter η extracted from the C_{dep}/C_{ox} ratio and μ_{eff} sensitivity to V_g and V_{sub} using the method explained in [42-44]. Fig 4.15 (g) is the final extracted μ_{eff} – E_{eff} plot for different V_{sub} showing clearly the V_{sub} effect. The deviation from the universal curve at lower E_{eff} is explained in section 2 and is normally attributed to higher Coulomb scattering, due to high K, as well as due to artifact of split CV method itself at low inversion region. At high electric fields the curves converge to a universal curve due to SR scattering. The difference between the μ_{eff} obtained from CV and Y-function fit is seen in fig 4.15(h). Both the curves converge at high fields but show significant deviation at low V_g. Split CV method does not give a meaningful result for threshold and sub threshold regions, as here the capacitance is dominated by parasitic capacitances. Also the drain current cannot be represented using the linear regime expression alone in this region. Therefore split CV method is also limited to gate voltages well above the threshold voltage.

The parameter η has been studied by Bennamane et al on sSi nMOS [43] and extended the extraction method to FDSOI devices [44]. The extraction is based on the sensitivity of mobility to gate bias and substrate bias. The strained Si nMOS has shown a different value for η around 0.7 showing higher effective electric field. The parameter η for FDSOI is reported to be between 0.6-0.7. The expression used for η extraction is shown in Eq 4.38. The ratio C_b/C_{oxf} can be directly measured from the slope of V_t vs substrate bias V_b for both the bulk and FDSOI cases as $-dV_t/dV_b$. The term t_{si} is the silicon thickness and t_{box} is the BOX thickness for FDSOI/DG case. C_{oxf} is the front gate capacitance. C_{dep} for bulk MOSFET case is the depletion capacitance.

$$\eta = \frac{R}{1+R}; R = \frac{C_b}{C_{oxf}} \frac{\partial \mu_{eff} / \partial V_{g\,|V_b}}{\partial \mu_{eff} / \partial V_{b\,|V_g}}$$

$$C_b = \frac{C_{si} C_{box}}{C_{box} + C_{si}};\ for\ FDSOI; C_b = C_{dep}\ \ for\ bulk \tag{4.38a}$$

$$C_{si} = \frac{\varepsilon_{Si}}{t_{Si}}; C_{box} = \frac{\varepsilon_{ox}}{t_{Box}};$$

$$E_{eff} = \frac{\int E(z) n(z) dz}{Q_{inv}} = \frac{\left(Q_{dep} + \eta Q_{inv}\right)}{\varepsilon_{Si}}; Q_{inv} = \int n(z) dz \tag{4.38b}$$

4.3.1 η - Extraction result on sSOI

The extraction method for the FDSOI of ref [44] discussed above is applied to strained silicon on SOI (sSOI) from LETI and η parameter for sSOI is extracted for the first time. The devices measured have 145nm Box. The sSOI devices have a silicon thickness of ~ 10nm with a biaxial tensile strain of 20%. Devices have HfO_2 /TiN gate stack with EOT ~ 2 nm. The effective mobility is calculated from the I_D-V_g by taking inversion charge $Q_{inv} = C_{ox}(V_g - V_t)$ at high inversion. Fig 4.16 right shows the effective mobility for 10µm (W) x 2µm (L) device.

The fig 4.16 left shows the threshold voltage variation with substrate bias. The slope gives the capacitance ratio. From the ratio of derivative of effective mobility with gate voltage to derivative of effective mobility with substrate bias and the capacitance ratio we can calculate the parameter η at high inversion using the expression Eq 4.38 as explained. The fig 4.17 shows the results for η parameter for 10µm and 2µm long devices. The results show η ~ 0.75-0.8 which is significantly different from the conventional 0.5 value obtained for bulk and it is even larger than sSi result published previously by Bennamane et.al [43].

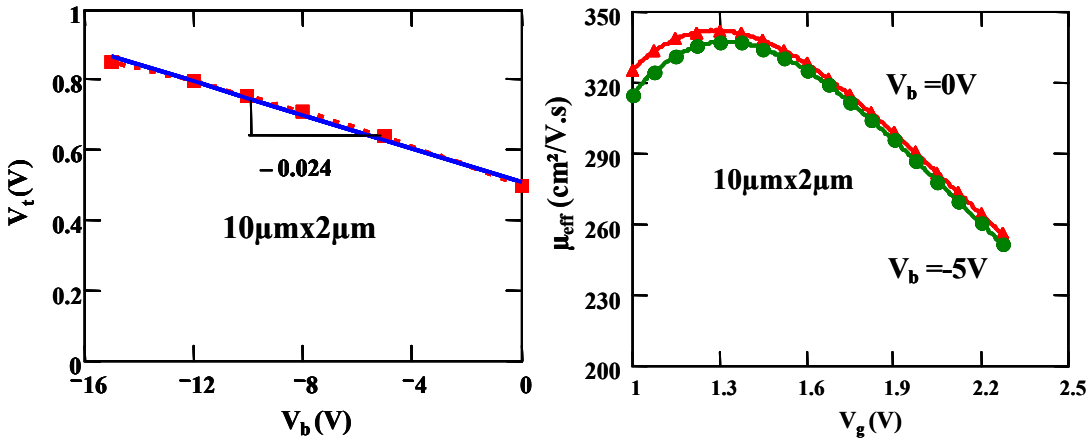


Fig 4.16

Left: Threshold voltage V_t vs substrate bias V_b giving the capacitance ratio. Right: Effective mobility vs gate voltage for two substrate bias 0V and -5V is shown in the figure. The device is 10µm wide and 2µm long sSOI 20% strain.

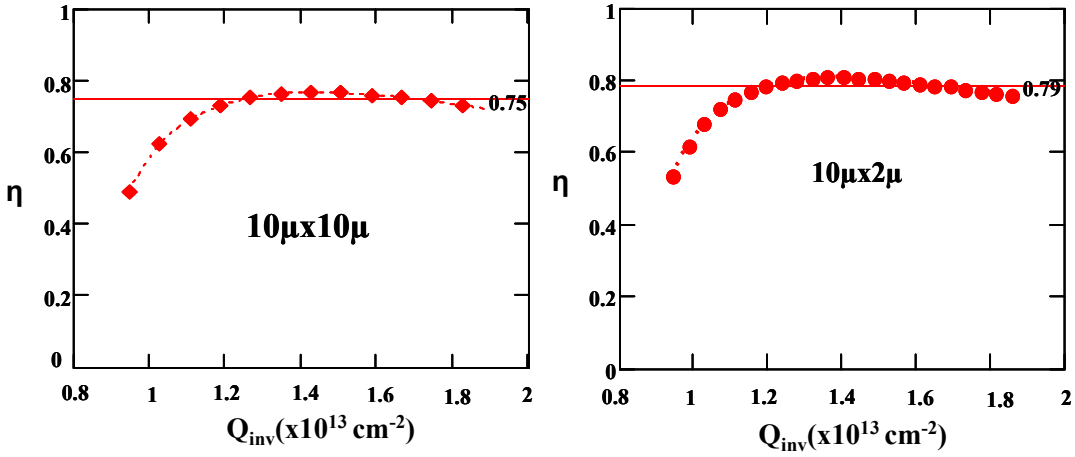


Fig 4.17

Left: η parameter for 10µm x 10µm device. Right: η parameter for 10µm x 2µm device.

The value of η for sSi or FDSOI or sSOI is found consistently different from classical value used for E_{eff} defined by triangular potential well approximation discussed in the section 2 [ref 8 of section 2]. To first order, we see that η for bulk, FDSOI, sSi and sSOI are 0.5, 0.6, 0.7 and 0.8 respectively. The origin of this difference is not yet fully understood. Based on the theory of preferential occupation of carriers in lowest sub band in sSi and sSOI compared to bulk case (see section 2) the carriers are more confined near the interface. The surface roughness and phonon scattering is larger (see section 2) and this corresponds to a higher effective field case under the bulk MOSFET triangular potential well model (or constant field model) which is the basis for the extraction method of η. Therefore effective field calculated from mobility point of view could be larger. This could explain why in the 10nm Si

FDSOI case the difference is not that large compared to strain cases as the preferential occupation effect in FDSOI is smaller. However in electrostatic point of view effective field is the charge weighted average of electric field distribution in the inversion layer and preliminary simulation did not show any difference in η due to sub band occupation. Deeper study supported by simulation is necessary for complete understanding of the origin of difference in the η parameter in terms of wave form shape due to confinement and sub band occupation.

4.4 Low temperature analysis of mobility

In order to separate the contribution of different scattering mechanisms, we can make use of their temperature dependence of mobility components discussed in the section 2. The different scattering mechanisms identified are phonon, Coulomb, SR, neutral defects for short channel length <100nm. Apart from these there is additional degradation of mobility due to High K/Metal gate stack for advanced device. This is attributed to Remote Coulomb RCS, increased interface states and remote phonon scattering. RCS is also attributed to S/D impurities for ultra-short devices sub 32nm. Also in sub 32nm we have to take ballistic transport model too. See section 2 for details. In brief we have 5 types of entities to look out for when we discuss about mobility namely phonon, Coulomb, SR, Neutral defects and ballistic transport. Therefore we can make use of their general trends to analyze the mobility by making use of Mathiessen rule. Limit of Mathiessen rule is discussed in section 2 and because of different energy dependence of scattering we cannot use it to obtain a full quantitative analysis. But it is still used for approximate analysis of the mobility components. The ballistic transport is taken into account by making use of Shur model discussed in section 2. From section 4.1, we can see that the low field mobility μ_0 is independent of surface roughness term. We can therefore make use of low field mobility μ_0 dependence with temperature and separate the remaining terms. The ballistic mobility is theoretically calculated and μ_{dd0} is separated from the apparent mobility μ_0 measured using Shur approach. From section 2 and section 4.1, we get Eq 4.39a (normalized to 300K).

$$\frac{1}{\mu_{0dd}} = \frac{1}{\mu_0} - \frac{1}{\mu_{bal}}$$
$$\frac{1}{\mu_{0dd}(T)} = \frac{1}{\mu_N} + \frac{T}{300\mu_{ph0}} + \frac{300}{T\mu_{C0}} \qquad (4.39a)$$

This method is made use in ref [45] where μ_N was found to be present in devices with channel lengths <100nm. The ballistic mobility μ_{bal} is calculated from the non-degenerate limit expression discussed in section 2. See section 6.1 for the extraction results based on this approach for FDSON devices.

The state of the art mobility vs temperature model of eq 4.39a is generalized to current gain factor β using eq 4.39b and rewritten as Eq 4.40.

$$\beta(T) = \frac{C_{ox} W_{eff} \mu_0(T)}{L_{eff}} \tag{4.39b}$$

$$\frac{1}{\beta(T)} = \left(\frac{300}{\beta_C T} + \frac{T}{300\beta_{Ph}} + \frac{1}{\beta_N} + \frac{T^{0.5}}{\beta_{Bal0}} \right) \tag{4.40}$$

The terms C_{ox}, L_{eff} and W_{eff} are assumed to be insensitive to temperature. In that case the only temperature dependence for β is due to the mobility μ_0. Therefore we can use β vs Temperature instead of μ_0 for studying the scattering mechanisms present in the device. Advantage of using β is that we can directly measure from I_{DS}-V_g, without the need to know C_{ox} or physical dimensions.

The term β_c is the contribution to gain factor by Coulomb scattering, β_{ph} is contribution due to phonon scattering, β_N is the contribution from Neutral defects [45]. The term β_{Bal0} can be expressed as eq 4.41.

$$\beta_{Bal0} = \frac{qL_{eff}}{\sqrt{2k\pi m^*}} \frac{W_{eff} C_{ox}}{L_{eff}} = \frac{q}{\sqrt{2k\pi m_0}} C_{Bal}; \quad \frac{q}{\sqrt{2k\pi m_0}} = 1.8 \bullet 10^9; m^* = \alpha m_0; C_{Bal} = \frac{W_{eff} C_{ox}}{\sqrt{\alpha}} \tag{4.41}$$

W_{eff} is in cm, while m_0 is in kg, C_{ox} is in F/cm². The constant has been calculated accordingly by taking appropriate unit for Boltzmann constant k. The term α can be calculated if we know the channel and surface orientations using eq 4.42 [section 2 ref 7]. In the case of m* = m_t (transverse effective mass), α = 0.19.

$$m_C^* = \frac{2}{\frac{1}{m_x} + \frac{1}{m_y}}; \tag{4.42}$$

The term m^*_C is the conductivity effective mass. The value of m_x (transport direction) and m_y (width direction) are taken according to the surface orientation. For silicon (100) surface lowest sub band m_x = m_y = m_t. Where m_t is the transverse effective mass = $0.19m_0$ and m_l is the longitudinal effective mass =$0.915m_0$ for silicon. This method however has to be carefully applied to obtain a sensible fit. For planar devices with large width we can directly calculate β_{bal0} as we know all the terms of the expression. Section 6 shows the result of applying this method on FinFet and GAA device. The extraction based on β is however difficult as we have too many unknowns. But the temperature

dependence of β can give a qualitative insight into the scattering mechanism present in advanced devices where physical dimensions are unknown.

Different electrical extraction methods mentioned above are dependent on a specific model for high inversion regime and they do not apply in the low inversion or sub-threshold. Magnetoresistance measurement is an alternative characterization method especially for advanced MOSFETs owing to its ability to make extraction without need for physical dimensions and can measure even in the sub-threshold. Magnetoresistance measurement is discussed briefly in the section below.

4.5 Magnetoresistance

The magnetoresistance mobility measurement provides a powerful tool to measure channel mobility of a MOSFET without the need for knowledge of physical dimensions. The magnetoresistance mobility can be directly measured in the linear regime of MOSFET from the relative change in the resistivity or drain current with the transverse magnetic field. The advantage of this method is that we can extract mobility directly from measurable quantities. In order for this to work W/L ratio of the MOSFET should be > 5 under the assumption that the hall voltage is cancelled in such structures. The expression used for the extraction of μ_{MR} is shown in Eq 4.43.

$$\left.\frac{I_{D0} - I_{DB}}{I_{DB}}\right|_{Vg,Vd} = \mu_{MR}^2 B_Z^2 \tag{4.43}$$

The term I_{D0} is the drain current at 0 field, I_{DB} is the drain current at B_z (subscript z is to denote that the magnetic field is in the z direction, normal to the 2D inversion carrier transport plane (x-y plane)). The magnetoresistance mobility μ_{MR} then can be extracted for different V_g, V_d at different temperatures. The extraction on 10μm wide 500nm channel length bulk MOSFET with High K / Metal gate is shown in fig 4.18. Measurement is made in LNCMI high magnetic field Lab, Grenoble. The magnetic field is varied from 0T to 11T, V_d 50mV and V_g from 0 to 1.2V. The mobility is larger at low temperature and this is also reflected in the I_D variation with magnetic field as seen in fig 4.18 right. The fig 4.19 shows the extracted magnetoresistance mobility μ_{MR} with V_g for 110K and 270K. The mobility is degraded at 270K and also more flat than at 110K due to phonon scattering. As discussed in section 2, the mobility at 110K is mostly composed of Coulomb and SR scattering. The mobility at low inversion or low V_g seems to be smaller at low temperature and this could be explained by the Coulomb mobility which is dominant in this region, L.Thevenod et al [50]. (Measurement at low V_g is affected by noise and so there is an associated uncertainty in this result).

The effect of series resistance can be removed using MR measurement using two channel lengths using the method proposed in ref [34].

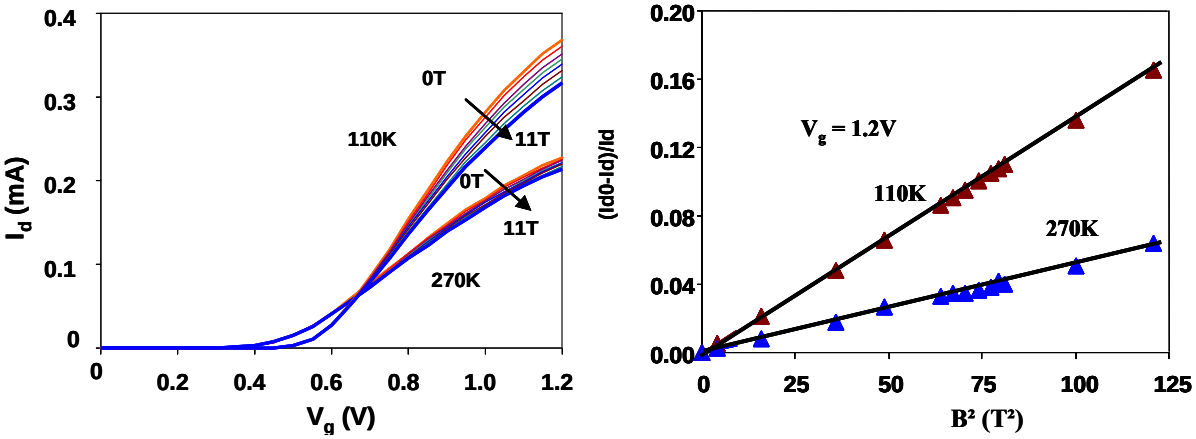


Figure 4.18

Left: $I_D - V_g$ for different magnetic field and temperature for bulk MOSFET with 10µm x 0.5µm High K metal gate. At low temperature I_D variation is larger as μ_{MR} is larger. Right: The relative I_D with B^2 showing straight line for V_g 1.2V for two temperatures showing μ_{MR} @ 270K < μ_{MR} @ 110K.

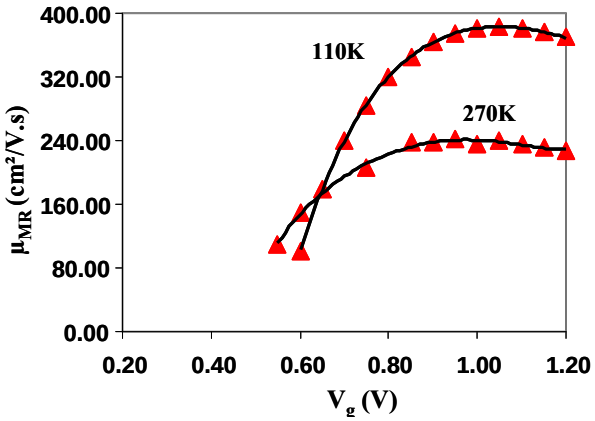


Figure 4.19

Extracted μ_{MR} in $cm^2/V.s$ with V_g at 110K and 270K.

The magnetoresistance mobility is related to effective mobility by Eq 4.44 [30 - 34].

$$\mu_{MR} = \sqrt{\frac{\int ED(E)\mu^3(E)\left(-\frac{df_0}{dE}\right)dE}{\int ED(E)\mu(E)\left(-\frac{df_0}{dE}\right)dE}} \qquad (4.44)$$

µ(E) is the energy dependent mobility q τ(E)/m^*, D(E) is the density of states and f_0 is the Fermi-Dirac distribution function. In the case where µ(E) is independent of E, then $\mu_{MR} = \mu_{eff}$ (or in the case of degenerate condition). As discussed in the section 2, we can represent the mean free time τ(E) as a

simple power law E^n with the exponent n depending on the scattering mechanism. For acoustic phonons scattering which generally have a power law n = -0.5 and for Coulomb scattering with n=1 μ_{MR} is ~ 60% higher than μ_{eff} in non-degenerate condition, Meziani et al [30]. In the case of Hall mobility, the relation is given in Eq 4.45. In this case μ_{Hall} is reported to be 20 – 40% higher than the μ_{eff}.

$$\mu_{Hall} = \frac{\int ED(E)\mu^2(E)\left(-\frac{df_0}{dE}\right)dE}{\int ED(E)\mu(E)\left(-\frac{df_0}{dE}\right)dE} \tag{4.45}$$

The magnetoresistance so far discussed are for linear regime operation and we have explored the MR in the saturation regime of operation for the first time. This is discussed in the section 5. MR in the case of sub 32nm node MOSFETs needs to be studied by taking into account ballistic transport. The section below discusses MR model for full ballistic case under small magnetic field regime and later the result is extended by including small amount of scattering for more realistic devices working in quasi ballistic regime in sub 32nm nodes. The analytical expression is then compared with already published MC and MSMC simulation results [33, 49].

4.6 MR in ballistic Transport

Ballistic MR model is derived for device operating in the linear regime with small magnetic field and electric field. The MOSFET width, W, is taken to be large and use expressions similar to that in section 2 for ballistic MOSFETs. The assumption for the below derivation is that electrons are either absorbed in drain or source before the completion of cyclotron motion. Source and Drain regions are assumed to have high levels of scattering. Assuming also $E_x/B_z << v(E)$ for majority of carriers, all the expressions below are derived by taking only magnetic field. This sets an approximate limit as shown in Eq 4.46.

$$r_C > L_{Ch};$$
$$r_C(E) = \frac{m^* v(E)}{qB} \tag{4.46}$$

The term r_c is the cyclotron radius (classical) and L_{ch} is the channel length. Expressions are derived at virtual source at the top of the barrier at the source region. Normally there is electric field in the channel region which has to be found self consistently.

The term v(E) is the carrier velocity as a function of energy, B is the magnetic field in Tesla, q is the electron charge absolute value and m^* is the effective mass. The electric field in -x direction is small. The well-known ballistic MOSFET approach is followed with modification for the magnetic field.

The electron current is divided as ID^+ and ID^-. The first term is the current flow by source injected carriers towards drain while the second one is due to the drain injected carriers in the reverse direction. Net current flow I_D is ID^+- ID^-. Only one sub band is considered in this derivation for simplicity. General expression for the drain current is given in Eq 4.47 for source to drain directed electrons.

$$ID^+ = \frac{2q}{h}\int_{E0s}^{\infty} M(E) f_1(E) dE \tag{4.47}$$

E_{0s} is the bottom of the first sub band at the virtual source. M(E) is the 2D mode and $f_1(E)$ is the Fermi-Dirac function with source Fermi level E_{f1}. Eq 4.48 shows the expanded expression taking 2D density of states.

$$ID^+ = \frac{2qW}{h}\int_{E_{0s}}^{\infty} \pi\hbar \frac{D_{2D}(E)}{2} v_x(E) f_1(E) dE \tag{4.48}$$

The source to drain directed flux occupies half of the 2D density of states (+kx) and the carriers injected from drain towards source (ID^-) occupy the other half in the absence of scattering (-kx) and magnetic field. ID^- however uses the Fermi-Dirac function $f_2(E)$ with the Fermi level of the drain side $E_{f2} = E_{f1}-qV_d$. $v_x(E)$ is the carrier velocity component in the transport direction denoted as x-direction (+x towards drain), W is the width of the channel, (ħ, h) are the reduced Planck's constant and Plank's constant respectively.

Fig 4.20 (top) shows the conduction band (first sub band) from source to drain. The virtual source region is also shown. The carrier distribution at virtual source for low V_d is shown just below. We can see that the distribution is asymmetric as $-v_x$ side is occupied by drain injected carriers and drain side Fermi level is qV_d smaller than source Fermi level. The 2D k-space distribution is shown at the bottom for the same region. The angle $\Theta_{max}(E)$ is an important parameter and it determines the amount of carriers back deflected by magnetic field as a function of energy. This angle also depends on the channel length. This is because the carriers injected into the channel at an angle Θ_{max} or smaller would circle back in the presence of the magnetic field and get collected at source and assuming symmetry the same effect is taken for drain injected carriers also (for low V_d). The carrier that re-enters the source, re-enters at $-\Theta$ angle, where it was originally injected into the channel at an angle Θ ($\leq \Theta_{max}$). Similarly for the drain injected carriers. Therefore, at virtual source, –kx space up to $-\Theta_{max}$ is filled up by source back deflected carriers and rest is filled by drain transmitted carriers (since we assume symmetrical behavior for drain). This is shown in fig 4.20 bottom.

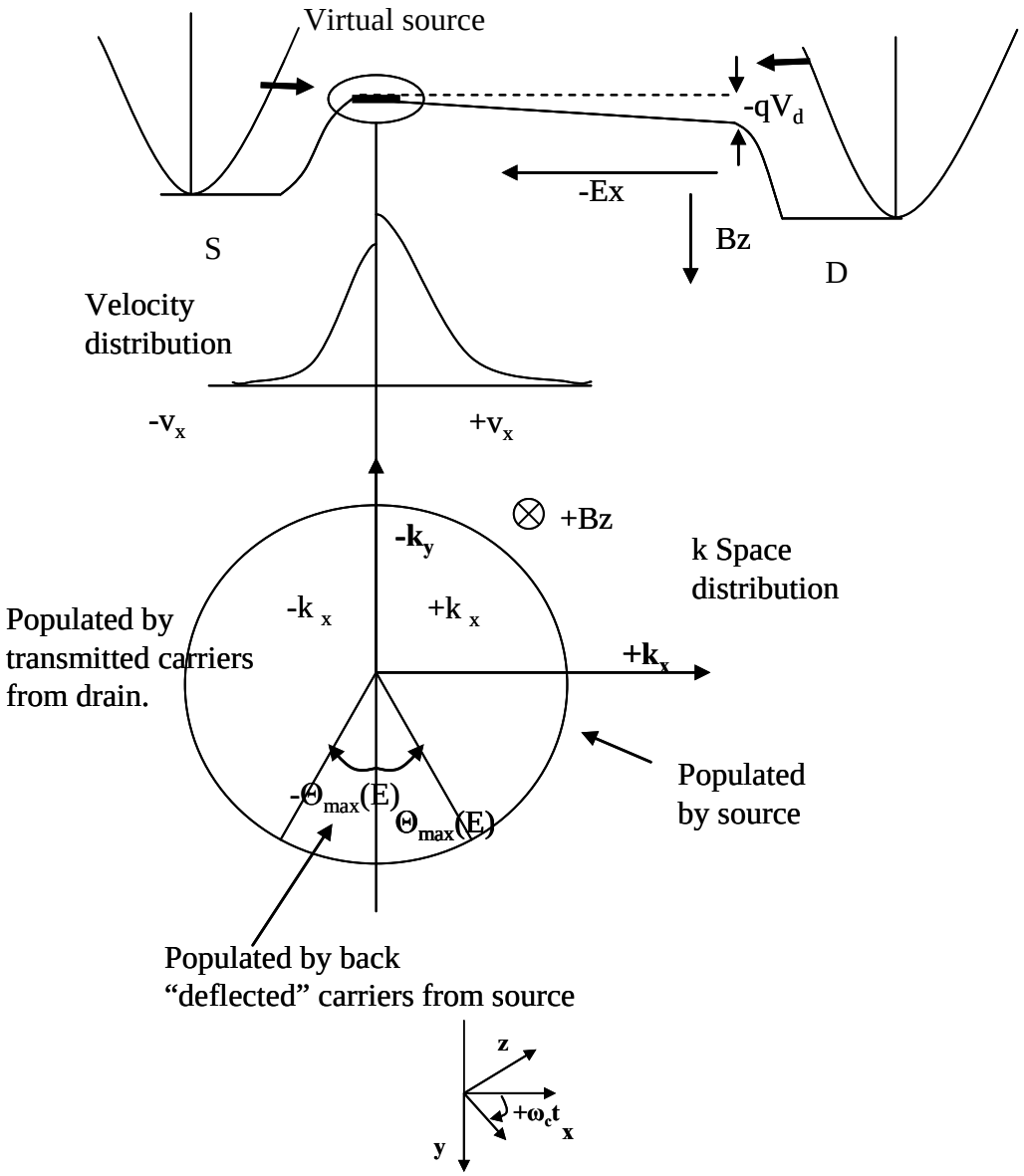


Figure 4.20

Top: shows the bottom of the first sub band with source and drain distribution assumed parabolic and circular (2D). Drain side Fermi level is reduced by a bias V_d by an amount qV_d. Middle: shows the distribution in the velocity space at the virtual source showing that the $+v_x$ side is filled by the injected carriers from source and $-v_x$ side is filled by drain injected carriers. The presence of magnetic field changes the scenario as shown in the bottom figure. Bottom: Shows the k – space distribution. The effect of the magnetic field is shown where a portion of region in $-k_x$ space is populated by the carriers that are injected with in an angle $\Theta_{max}(E)$ from source. $\Theta_{max}(E)$ is a function of energy. Carriers injected with a larger angle will get absorbed in drain. Magnetic field is in +z direction into the page. Bottom right shows the rotation direction.

From fig 4.21, the injected carrier that has the x-component of velocity v_x, when returned to source, will have $-v_x$. Θ_{max} (E) is the maximum injection angle up to which injected carriers will get collected back at source (similarly for drain) for a given energy, channel length and magnetic field. Following derivation is made under this assumption which works well for high energy carriers which meet the criteria given in Eq 4.46. The presence of space varying electric field changes this picture. This is not treated in this derivation assuming small V_d. Limitations of these assumptions are discussed in detail in section 4.6.2.

In fig 4.21, r_{cx} is different from the cyclotron radius r_c. r_{cx} is the radius of motion of a carrier injected with the x-component v_{injX} of the actual injection velocity $v_{inj}(E)$ but injected in x-direction. Let us take the $\Theta_{max}(E)$ case because any carrier injected above this angle for the given energy will be transmitted and below this angle will get collected at source itself. This gives Eq 4.49.

$$r_{CX}(E) = \frac{m^* v_{injX}(E)}{qB} = L_{ch};$$
$$v_{injX}(E) = v_{inj}(E)\sin\Theta_{\max}(E) \tag{4.49}$$
$$\therefore \Theta_{\max}(E) = \sin^{-1}\frac{qBL_{ch}}{m^* v(E)} \approx \frac{qBL_{ch}}{m^* v(E)}$$

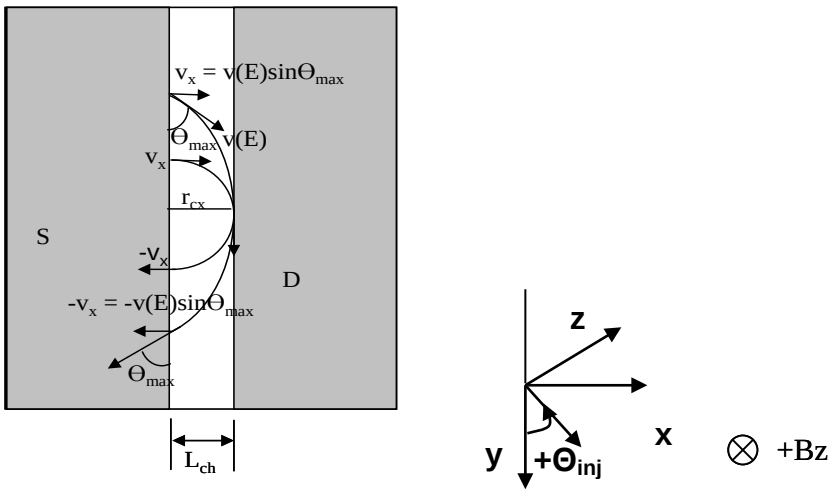


Figure 4.21

Figure shows trajectory of carrier, which is injected at a velocity v(E) at an angle Θ_{max} (E) that returns back to source and re-enters at $-\Theta_{max}(E)$. If we take the evolution of x-component of the velocity v(E), $v_x(E)$, along the trajectory we can see that the carrier re-enters the source with $-v_x(E)$. We can then picture a fictitious carrier injected with velocity $v_x(E)$ normally into the channel completing a semi-circle of radius $r_{cx} = L_{ch}$.

In the eq 4.48 we have to take the average velocity v_x. This averaging is in injection angle Θ. But as we discussed only carriers injected from Θ_{max} and above make it to the drain. So the averaging is only

done for this region. The 2D density of states that effectively contribute to ID^+ is also reduced to $D_{2D}^{+}(E) = D_{2D}(E)\frac{(\pi - \Theta_{\max}(E))}{2\pi}$ (for a parabolic and circular band) instead of $D_{2D}/2$ that is used for ballistic MOSFET expression. Therefore modified expression for ID^+ is given as Eq 4.50.

$$ID^{+} = \frac{2qW}{h}\int_{Ec}^{\infty} \pi\hbar \frac{D_{2D}(E)\left(\pi - \Theta_{\max}(E)\right)}{2\pi} \frac{\int_{\Theta_{\max}(E)}^{\pi} v(E)\sin\theta d\theta}{\int_{\Theta_{\max}}^{\pi} d\theta} f_1(E)dE$$
$$= \frac{2qW}{h}\pi\hbar\frac{D_{2D}}{2\pi}\int_{Ec}^{\infty}\left[1+\cos\left(\Theta_{\max}(E)\right)\right]v(E)\,f_1(E)dE \qquad (4.50)$$

Consider the energy range for the integration. As per the assumptions made we cannot really integrate from bottom of the conduction band. Only carriers with energy above $E_C + \frac{(qBL_{ch})^2}{2m^*kT}$ can be considered to follow this assumption as only these carriers satisfy the criteria of $r_C > L_{ch}$. The picture is more complicated for carriers with energy below this and for simplicity this condition is omitted in the following expression. See the Appendix A for full derivation. Eq 4.51 gives the final expression for ID^+ after integration by neglecting an incomplete Fermi-Dirac complex integral for the limit $\frac{(qBL_{ch})^2}{2m^*kT}$ <<1. See Section 4.6.2 for discussion on errors introduced by this assumption and for more accurate way to integrate. N_{2D} is $D_{2D}kTg_v$ and $F_{1/2}$ is the Fermi-Dirac integral of order 1/2. g_v is valley degeneracy. For sub band 0, 1 (twofold) etc g_v = 2 and 0', 1' (fourfold) etc g_v = 4 for a Si(100) surface.

$$ID^{+} = qW\frac{N_{2D}}{4}\sqrt{\frac{2kT}{m^*\pi}}\frac{2}{\sqrt{\pi}}F_{1/2}(\eta_{F1})\left[1+\frac{F_{1/2}(\eta_{F1}-\frac{(qBL_{ch})^2}{2m^*kT})}{F_{1/2}(\eta_{F1})}\right] \qquad (4.51)$$

The net drain current expression $ID^+ - ID^-$ becomes as Eq 4.52 in the limit $\frac{(qBL_{ch})^2}{2m^*kT}$ <<1.

$$ID = \frac{qW\frac{N_{2D}}{2}v_T\Im_{1/2}(\eta_{F1})}{\left[1+\left(\frac{(qBL_{ch})^2}{4m^*kT}\right)\frac{\Im_{-1/2}(\eta_{F1})}{\Im_{1/2}(\eta_{F1})}\right]}\left[1-\frac{\Im_{1/2}(\eta_{F2})}{\Im_{1/2}(\eta_{F1})}\right] \qquad (4.52)$$
$$\Im_{1/2}(\eta_{F1}) = \frac{2}{\sqrt{\pi}}F_{1/2};\ \Im_{-1/2}(\eta_{F1}) = \frac{d}{d\eta_{F1}}\Im_{1/2}(\eta_{F1})$$

Remember that $\eta_{F2} = \eta_{F1} - (qV_d/kT)$. From the k-space image in fig 4.20, we see that +kx region and region up to $-\Theta_{\max}$ (E) in the $-k_x$ space region are filled by Fermi level of source and rest is filled by transmitted carriers from drain and so the Fermi level of drain has to be used. Instead of evaluating the

integral directly we can approximate the same by assuming $\Theta_{max}(E)/\pi << 1$. Then the equation comes down to normal ballistic MOSFET expression. Considering degenerate (even for non-degenerate case it is valid if field is small enough) case where there are lots of high energy electrons this assumption becomes more valid. Therefore we have Eq 4.53.

$$\frac{qN_{2D}W}{2} = \frac{C_{ox}(V_g - V_t)W}{[\Im_0(\eta_{F1}) + \Im_0(\eta_{F2})]} \tag{4.53}$$

Substituting this in eq 4.52 we get the final expression for the drain current under weak magnetic field as given in Eq 4.54.

$$\begin{aligned} ID &= WC_{ox}(V_g - V_t)v_T \frac{\Im_{-1/2}(\eta_{F1})}{\Im_0(\eta_{F1})}\left[1 - \left(\frac{(qBL_{ch})^2}{2\pi(m^* v_T)^2}\right)\frac{\Im_{-1/2}(\eta_{F1})}{\Im_{1/2}(\eta_{F1})}\right]\frac{qV_d}{2kT} \\ &= \frac{WC_{ox}(V_g - V_t)}{L_{ch}} \frac{\frac{qL_{ch}v_T}{2kT}\frac{\Im_{-1/2}(\eta_{F1})}{\Im_0(\eta_{F1})}}{\left[1 + \left(\frac{(qL_{ch})^2}{2\pi(m^* v_T)^2}\right)\frac{\Im_{-1/2}(\eta_{F1})}{\Im_{1/2}(\eta_{F1})}B^2\right]}V_d \end{aligned} \tag{4.54}$$

The equation is written in two forms. The first one is of quasi ballistic form with linear *magneto - ballistic ratio* $BR_{LinMR} = 1\text{-}r_{linMR}$, where r_{linMR} is the linear *magneto - back scattering (or reflection) coefficient*. The second one is of the form ballistic mobility μ_{bal} and ballistic MR mobility μ_{balMR}. This expression is in degenerate inversion. We can directly take this expression in non-degenerate limit with small magnetic field and short channel lengths such that $r_C > L_{ch}$ for majority of the carriers and so $(qBL_{ch})^2/2m << kT$. We can apply the non-degenerate expression for the Fermi-Dirac integrals. Remember that the low energy electrons does not normally meet this criteria and assumption is that, provided B is small, this portion of inversion carriers is negligible and the above mentioned condition is applicable to majority of carriers and the integration from bottom of the conduction band as well as the approximation in the evaluation of inversion carrier density does not introduce too much error ($\Theta_{max} << \pi$).

From Eq 4.54 we have eq 4.55a.

$$\mu_{balMR} = \frac{1}{\sqrt{2\pi}}\frac{qL_{ch}}{m^* v_T}\left(\frac{\Im_{-1/2}(\eta_{F1})}{\Im_{1/2}(\eta_{F1})}\right)^{1/2} \tag{4.55a}$$

Eq 4.55b is the linear ballistic mobility.

$$\mu_{bal} = \frac{qL_{ch}v_T}{2kT}\frac{\Im_{-1/2}(\eta_{F1})}{\Im_0(\eta_{F1})};\Im_{-1/2}(\eta_{F1}) \approx \frac{2}{\sqrt{\pi}}\sqrt{\eta_{F1}};\Im_0(\eta_{F1}) \approx \eta_{F1}$$
$$\approx \frac{qL_{ch}v_T}{2kT}\frac{2}{\sqrt{\pi}}\frac{1}{\sqrt{\eta_{F1}}};highly\ degenerate\ case(linear) \quad (4.55b)$$

The ratio of μ_{balMR}/μ_{bal} is shown in Eq 4.55c.

$$\frac{\mu_{balMR}}{\mu_{bal}} = \sqrt{\frac{\pi}{2}}\frac{\Im_0(\eta_{F1})}{\sqrt{\Im_{1/2}(\eta_{F1})\Im_{-1/2}(\eta_{F1})}};\because 2kT = \pi m^* v_T^2$$
$$= \sqrt{\frac{\pi}{2}} \approx 1.253\ \ in\ non-degenerate\ case \quad (4.55c)$$
$$we\ have\ \Im_{1/2}(\eta_{F1}) = \frac{4}{3\sqrt{\pi}}(\eta_{F1})^{3/2}(highly\ degenerate\ case)$$
$$\therefore \frac{\mu_{balMR}}{\mu_{bal}} \approx \frac{\sqrt{3}}{2}\frac{\pi}{2} \approx 1.36$$

Eq 4.56 gives the linear magneto-back scattering coefficient and linear regime magneto-ballistic ratio BR_{linMR}. As we see the r_{linMR} is a function of Θ_{max} at velocity $v = v_T$.

$$BR_{LinMR} = 1 - r_{linMR} = 1 - \left(\frac{(qBL_{ch})^2}{2\pi(m^* v_T)^2}\right)\frac{\Im_{-1/2}(\eta_{F1})}{\Im_{1/2}(\eta_{F1})}$$
$$= 1 - \mu_{balMR}^2 B^2 \approx 1 - \Theta_{\max}^2(v_T)\frac{\Im_{-1/2}(\eta_{F1})}{2\pi\,\Im_{1/2}(\eta_{F1})} \quad (4.56)$$
$$where\ \ \Theta_{\max}(v_T) \approx \frac{qBL_{ch}}{m^* v_T}$$

Therefore the magnetoresistance mobility for a full ballistic MOSFET has same dependence to channel length as ballistic mobility term. In the non-degenerate limit, low magnetic field case ballistic MR is just ballistic mobility multiplied by a constant factor of $\sqrt{}$ (π/2). The finite device width and channel dependent electric field would change this factor. Also in the evaluation of inversion charges we have assumed $\Theta_{max} << \pi$. Limitation of this model is discussed in section 4.6.2.

4.6.1 Quasi ballistic case

The expression for quasi ballistic transport in the absence of magnetic field is discussed in section 2 using flux theory [46, 47]. In the presence of magnetic field, carriers take a more complex trajectory. For example (1) the carrier back deflected by magnetic field could encounter scattering, (2) the carrier back scattered could get deflected by magnetic field and get transmitted (or reflected back) depending on scattered angle and channel position. Remember that we cannot directly use the *magneto-back*

scattering coefficient r_{linMR} for the back scattered flux since it is valid only for the case of injection carriers from source and drain. For a simple derivation, we can replace the forward and reverse flux with the modified flux and injection velocity, derived in the previous section and follow the steps in ref [46]. The back scattering coefficient is not same as the back scattering coefficient of 0 magnetic field case. The validity of this assumption is analytically studied using a simplified model using two back scattering coefficients r_B and r'_B one for transmitted flux and one for back deflected flux due to magnetic field respectively (see figure B1 in appendix B). Since carrier trajectory is dependent on the magnetic field, we may have to consider r_B as a magnetic field dependent term apart from being different from electrical r_{lin}. The derivation based on r_B, r'_B is given in the appendix B for a given magnetic field. The final expression for drain current following the same procedure as [46] is given in Eq 4.57 with low V_d simplification of Eq B4 in appendix B.

$$ID = \frac{W}{L} C_{ox}\left(V_g - V_t\right) \frac{\left(1-r_B\right)\mu_{bal}\left(\eta_{F1}\right)}{\left[\left(1+\mu_{balMR}^2\left(\eta_{F1}\right)B^2\left[1-\frac{\left(1-r_B'\right)r_B'}{\left(1-r_B\right)}\right]\right)\right]} V_d \tag{4.57}$$

$\mu_{bal}(\eta_{F1})$ is the electrical ballistic mobility discussed in the section 2 for degenerate case. $\mu_{balMR}(\eta_{F1})$ is derived in the previous section. Therefore we can express a quasi-ballistic MR from Eq 4.57 with a simplification that $r_B = r'_B$ as in Eq 4.58.

$$\begin{aligned} &\mu_{QBMR}\left(\eta_{F1}\right) = \sqrt{(1-r_B)}\,\mu_{balMR}\left(\eta_{F1}\right) \approx (1-\frac{r_B}{2})\,\mu_{balMR}\left(\eta_{F1}\right) \\ &\therefore \mu_{QBMR} = \left[\frac{1}{\mu_{balMR}} + \frac{1}{\mu_{ddMR}}\right]^{-1};\ \mu_{balMR} \propto L \end{aligned} \tag{4.58}$$

We can express the magnetoresistance mobility in the quasi ballistic regime in a similar way as in the case of electrical quasi ballistic mobility using Shur like model in terms of drift diffusion like term μ_{ddMR} and ballistic term μ_{balMR}. This is in agreement with the MC simulation results obtained previously, discussed in section 4.6.3. The expression for the effective mobility therefore becomes as Eq 4.59 in the presence of magnetic field.

$$\mu_{eff}\left(B\right) = \frac{\mu_{effB0}}{1+\mu_{QBMR}^2 B^2};\ \mu_{effB0} = \left(1-r_B\right)\mu_{bal}\left(\eta_{F1}\right) \tag{4.59}$$

If r_B is magnetic field dependent then it will introduce a non-linearity in the plot $\Delta R/R$ vs B^2 in the quasi ballistic case. MC simulation results for 20nm DG MOSFET in ref [33 (fig 4)] does not show a

perceivable non linearity for the plot $\Delta R/R$ vs B^2. However the condition used for the simulation there does not match the condition required for the analytical model to work. This is discussed in section 4.6.3. Also it is important to note that μ_{ddMR} need not necessarily be μ_{MR} of the long channel as r_B relation to B and r_{lin} is unknown for the time being. Therefore the nature of r_B need to be studied. The quasi ballistic model, like the ballistic case, is limited to sub 50nm regime MOSFETs and low field in few T.

4.6.2 Limitations of ballistic MR model

In the absence of electric field, carriers with $2r_c \leq L_{ch}$ do not contribute to the transport. Carriers injected at injection angle π and $2r_c \leq L_{ch}$ will complete the cyclotron motion inside the channel itself (classical case). For injection angle $< \pi$ they cannot complete the cyclotron motion but instead get absorbed back into the source as shown in fig 4.22 A, B. In the intermediate case ie fig 4.22C, $L_{ch}/2 < r_c < L_{ch}$, carriers will have $\Theta_{max} > \pi/2$ and the expression derived with v_{injX} does not hold to these carriers. See fig 4.23. From the figure it is easy to see that $\Theta_{max}(E)$ has to be expressed as Eq 4.60.

$$\Theta_{\max}(E) = \sin^{-1}\left(\frac{qBL_{ch}}{m^* v(E)} - 1\right) + \frac{\pi}{2} = \cos^{-1}\left(1 - \frac{qBL_{ch}}{m^* v(E)}\right) \tag{4.60}$$

The energy range corresponding to different zones can be calculated from the expression of r_C. Since $L_{ch}/2 < r_C < L_{ch}$, we can find the energy limits for $r_C = L_{ch}/2$ and $r_C = L_{ch}$. This is given in Eq 4.61. For carrier energy below $(E\text{-}E_C)_{min}$ we have the situation given in fig 4.22A.

$$\begin{aligned}
&For\ r_C(E) = \frac{m^* v(E)}{qB} = \frac{L_{ch}}{2};\ v(E) = \sqrt{\frac{2(E - E_C)}{m^*}} \\
&\therefore (E - E_C)_{\min} = \frac{(qBL_{ch})^2}{8m^*} \\
&For\ r_C(E) = \frac{m^* v(E)}{qB} = L_{ch} \\
&\therefore (E - E_C)_{\max} = \frac{(qBL_{ch})^2}{2m^*}
\end{aligned} \tag{4.61}$$

The integration will be divided into three parts. From E_C to $[(E\text{-}E_C)_{min} = E_C + \frac{(qBL_{ch})^2}{8m^*}]$, the carrier does not contribute to transport in the absence of electric field. In the presence of electric field they drift to the edges and possibly contribute to edge current. For carriers from $(E - E_C)_{min}$ to $[(E\text{-}E_C)_{max} = E_C + \frac{(qBL_{ch})^2}{2m^*}]$, we have to use Eq 4.60 for $\Theta_{max}(E)$ and carriers with energy above $(E - E_C)_{max}$ obey the desired criteria used to formulate μ_{balMR}.

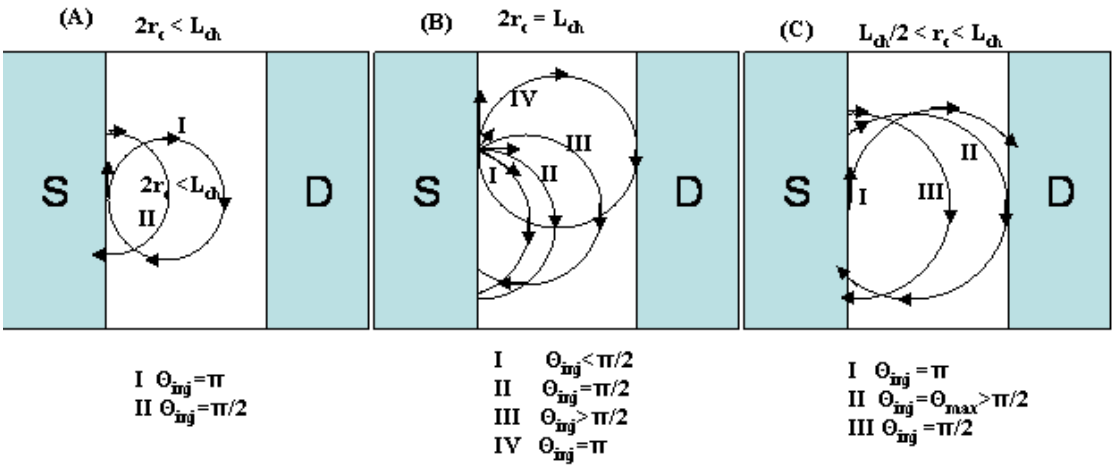


Figure 4.22

Figure showing situations that do not satisfy the assumption made for the μ_{balMR} derivation. The electric field is assumed 0. Fig (A) $2r_C < L_{ch}$: Carriers does not participate in the transport. However this is not the case in the presence of the electric field where there is a drift in the y (width) direction (classically) and therefore finite width has to be taken into account for a more realistic scenario. Fig (B): The carriers with $2r_C=L_{ch}$ for different injection angles. Fig (C) shows carrier with $L_{ch} > r_C > L_{ch}/2$, in this case $\Theta_{max} > \pi/2$ and therefore v_{injx} method needs to be modified.

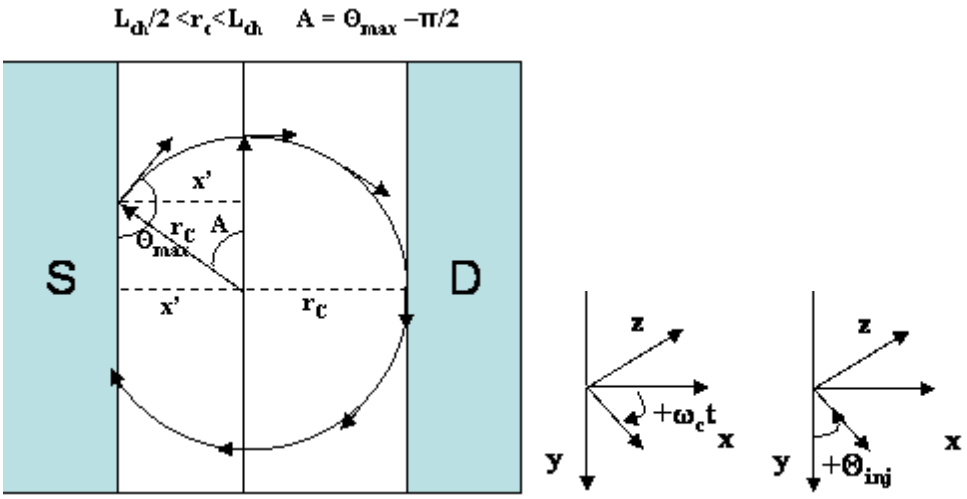


Figure 4.23

Figure showing the evaluation of Θ_{max} for the case C of fig 4.22. The center of the cyclotron is determined by the injection angle (0 electric field case). Figure on the right shows the convention used to define rotation and injection angle. The rotation is clock wise according to the definition of directions and field direction ($+B_z$). The injection angle is defined from y-axis from 0 to π. For injection angle > $\pi/2$ the center is inside the channel at a distance $x' = L_{ch}-r_C$ from the source (drain).

However this complicated scenario can be neglected as long as the contribution of carriers with energy below $(E\text{-}E_C)_{max}$ is negligible. The result obtained in the previous section and Appendix A strictly applies for the third region neglecting carriers below $(E\text{-}E_C)_{max}$. To analyze how these different complicated scenarios are going to affect the expressions derived, let us take typical values for the parameters involved in the expressions where assumptions have been made to simplify derivation. Let L_{ch} = 20nm, B = 0.1T, m = 0.19m_0, and kT at room temperature = 26meV. Let us take $v_T \sim 10^5$m/s which is the average velocity of electrons injected from source in non-degenerate case.

$$r_C(v_T) = \frac{m^* v_T}{qB} = 1.1\mu m >> L_{ch} \;;\; \frac{(qBL_{ch})^2}{2m} = 3\times10^{-25} J; @\, B = 1T = 3\times10^{-23} J$$
$$@\, B = 1T, r_C = 110nm \approx 6L_{ch} \qquad kT_{300K} = 4.14\times10^{-21} J \tag{4.62}$$

As we see from Eq 4.62, $r_C >> L_{ch}$ for majority of carriers, even at B =1T and non-degenerate case. The energy component $(E\text{-}E_C)_{max}$ is also 2 orders of magnitude smaller (0.2meV) at B = 1T (2μeV for 0.1T) than kT at room temperature. Therefore expressions derived here are valid for B ~ 1T and channel lengths ~ 50nm (using the value of $r_C \sim 2L_{ch}$ at 1T) or less for room temperatures. The percentage of the carriers, %Q, below $(E\text{-}E_C)_{max}$ relative to total carriers is evaluated as below assuming 2D density of states in Eq 4.63

$$\begin{aligned}
&(E - E_C)_{\max} = \Delta E_{\max} = \frac{(qBL_{ch})^2}{2m^*} = 0.2meV; \frac{\Delta E_{\max}}{kT} = 0.0077 \\
&\%Q = \frac{\int_{E_C}^{E_C+\Delta E_{\max}} D_{2D} \frac{1}{1+e^{\frac{E-E_{F1}}{kT}}} dE}{\int_{E_C}^{\infty} D_{2D} \frac{1}{1+e^{\frac{E-E_{F1}}{kT}}} dE}; \eta = \frac{E-E_C}{kT}; \eta_{F1} = \frac{E_{F1}-E_C}{kT} \\
&= \frac{\Im_0(\eta_{F1}) - \int_0^{\infty} \frac{1}{1+e^{\zeta-\left(\eta_{F1}-\frac{\Delta E_{\max}}{kT}\right)}} d\zeta}{\Im_0(\eta_{F1})}; \zeta = \eta - \frac{\Delta E_{\max}}{kT} \\
&= \frac{\Im_0(\eta_{F1}) - \Im_0\left(\eta_{F1} - \frac{\Delta E_{\max}}{kT}\right)}{\Im_0(\eta_{F1})} \approx \frac{\Delta E_{\max}}{kT} \frac{\Im_{-1}(\eta_{F1})}{\Im_0(\eta_{F1})}
\end{aligned} \tag{4.63}$$

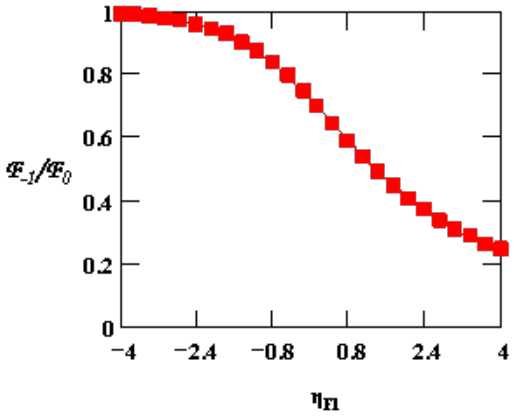


Figure 4.24

Figure shows the degeneracy factor vs Fermi level position η_F.

The degeneracy ratio $\frac{\mathfrak{I}_{-1}(\eta_{F1})}{\mathfrak{I}_0(\eta_{F1})}$ decreases with increasing η_{F1}. For negative η_{F1} in the non-degenerate case this approaches 1 hence from Eq 4.63, the percentage of carriers %Q below $(E\text{-}E_C)_{max}$ will be around 0.7% for the above mentioned case. For $\eta_{F1} = 1$ this is 50% smaller than non-degenerate case which can be easily verified from fig 4.24. Remember that this is $(E\text{-}E_C)_{max}$ and the carriers below $(E\text{-}E_C)_{min}$ is even smaller < 0.2%.

The mean free path in the source and drain region is assumed to be negligible compared to the channel length. If we include mean free path, then we have to add $2\lambda_m$ to the L_{ch} while comparing r_C to the channel length. This will modify Θ_{max} which was calculated by assuming that once carrier crosses the S/D border they get absorbed. In reality this is not true up to a distance (on average) of λ_m in source and drain.

The expressions derived so far assume a large width so we don't really consider the case of carriers injected near the channel edges. The carriers are assumed to be absorbed in drain (source) before they encounter the channel edges and the expression is derived based on this assumption. But as width becomes smaller, the non-ideality will be seen in the drain current expression as averaging of v_x along the injection angle from 0 to π will not be correct and carriers injected at small angles near 0 and large angles near π has to be treated differently. Although in the limit of large width we can simply take the 2D expression. Therefore requirement of large width for a ballistic MOSFET for using this magnetoresistance model is not really same as W/L requirement for MR in the drift diffusion case.

The effect of strong magnetic field is not explored in this derivation as this topic is more focused in the MR and not in the quantum effects.

It should also be remembered that all of the derivation above is made based on the assumption that the injected carrier velocity is larger than E_x/B_z so that all the expressions neglect the E_x/B_z component and therefore the range of drain voltage and magnetic field should also take this into account. Therefore drain voltages should be in μV while B should be in 0.1T – few T so that the overall criteria for the ballistic/quasi ballistic MR model to work is met. The term r_B, r'_B used in quasi ballistic case needs to be investigated for different kind of scattering entities.

4.6.3 Comparison with previous results

MC simulation results previously published by K Huet et al. [33] and M Zilli et al. [49] are compared with the analytical model derived for ballistic and quasi ballistic cases. Fig 4.25 shows the expected trends of MR (μ_{balMR} and μ_{QBMR}) and electrical mobility (μ_{bal} and μ_{QB}) from the analytical expression derived in previous sections. Based on the expression in Eq 2.21 and Eq 4.58, μ_{QB} (using μ_{bal} and μ_{dd}) and μ_{QBMR} (using μ_{balMR} and μ_{ddMR}) (all of them are shown in fig 4.25) is calculated by making use of result from Meziani et al [30] for taking the right pair of μ_{dd} and μ_{ddMR} where $\mu_{ddMR} \approx 1.6\mu_{dd}$. Non-degenerate condition expression is taken for the calculations. The channel length dependence is attributed only to ballistic effects. As per the analytical model derived it is expected that μ_{QBMR} should have a similar trend as μ_{QB}. This is well observed in MC simulation [33, 49] where the Mathiessen rule has been used by respective authors to explain the observed behavior of the μ_{QBMR}. Fig 4.26 gives the MC simulation results taken from ref [33]. As we see the MC simulations also shows the same qualitative behavior with channel length as predicted by analytical model. The differences observed can be explained as below.

Simulation uses V_d = 50mV ~ 2kT which is too large and therefore the analytical ballistic mobility expression is not applicable. The analytical expression for μ_{balMR} is valid only for small field with qV_d << kT. The drain voltage V_d = 50mV corresponds to maximum field of $E_x \sim V_d/L_{ch} \sim 2.5\ 10^4$V/cm. The carrier trajectory therefore is significantly different in the high field region of the channel and this is not taken into account in the analytical model derived in the previous section. Although in the case of 20nm we can go to few T of fields, the desired range is limited to 2T (see Eq 4.62 in the section 4.6.2). Since analytical model is not applicable in this condition, the simulation does not match the calculated values. The expression mentioned for the backscattering coefficient R (r_{linMR} in this thesis) in [33] is however different than that derived here but it still maintains the same trend with B and L_{ch}. This is because the assumption made for deriving the same $v_z(0) = 0$ is over simplified and therefore the expression obtained there has different constants. The expressions here and in [33] are otherwise similar. The limitation in using the expression is already discussed in ref [33].

Ref [49] gives multi sub band MC simulation of quasi ballistic case under magnetic field and extracts a channel length dependent component μ_L (ideally should be μ_{balMR}) by fitting Eq 4.64. See fig 4.27. Fig 4.27 (top) is in agreement with quasi ballistic MR μ_{QBMR} of Eq 4.58 in terms of its trend with channel length. The channel length dependent part μ_L for MR and electrical mobilities in ref [49] however have a ratio more than 2. However the simulation is done with a larger electric field 1KV/cm which at 1T gives $E_x/B_z \sim 10^7$cm/s which is large and comparable to v_T. This simulation is also done to channel lengths 100nm but the analytical expressions derived here are valid only up to few 10s of nm as shown in fig 4.25. Also analytical expression does not take into account multi sub band occupation.

Therefore a quantitative comparison with ref [33] as well as [49] is not possible as it does not really meet the criteria needed for observing the analytical value of μ_{balMR} / μ_{QBMR} but qualitatively we can see that the same trend with channel lengths is observed on both the references as predicted by the analytical model.

$$\frac{1}{\mu_{MR}} = \frac{1}{\alpha_L}\frac{1}{L} + \frac{1}{\mu_C}; \mu_L = \alpha_L L \tag{4.64}$$

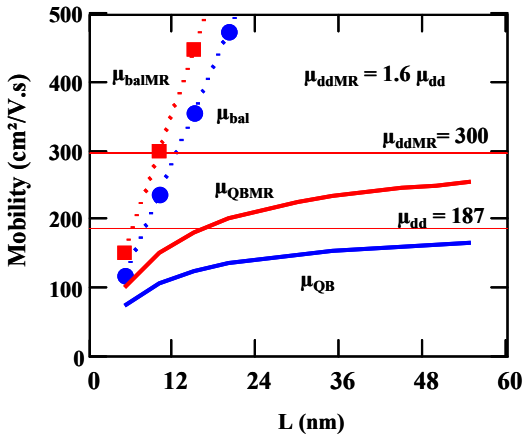


Figure 4.25

Simulated μ_{QB}, μ_{QBMR}, μ_{bal} and μ_{balMR} using the analytical expressions derived for μ_{QBMR} and μ_{balMR} and well known expression for μ_{QB} using Shur model. Value of μ_{ddMR} is taken 300cm^2.V^{-1}.s^{-1} arbitrarily and using the relation given by Meziani et. al [30] μ_{dd} is taken as μ_{ddMR}/1.6.

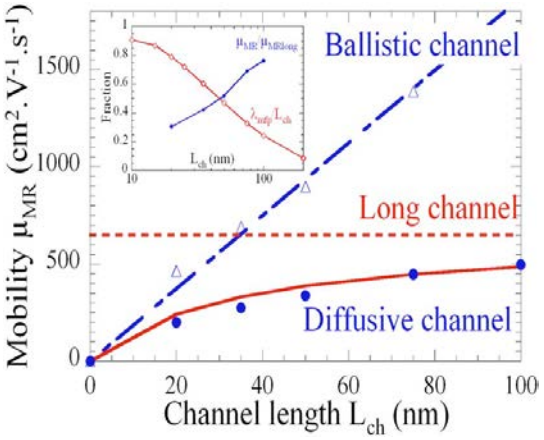


Figure 4.26

Result of MC simulation from ref 33 showing the linear dependence of μ_{balMR} to channel length L_{ch} in agreement with the analytical formula derived. Quoted from ref [33]: "Apparent magnetoresistance mobility as a function of L_{ch} in the MC simulated DG structures at T=300 K and N_S=10^{13} cm^{-2}. Solid circles: diffusive channel devices. Open triangles: ballistic channel devices. Dashed line: long channel magnetoresistance mobility. Solid line: Mathiessen-like approach. Inset: μ_{MR}/μ_{MRlong} ratio.(solid circles), mean free path λ_{mfp} to L_{ch} ratio. (open diamonds): as a function of L_{ch}, λ_{mfp} is extracted from the MC simulations at B=0 T."

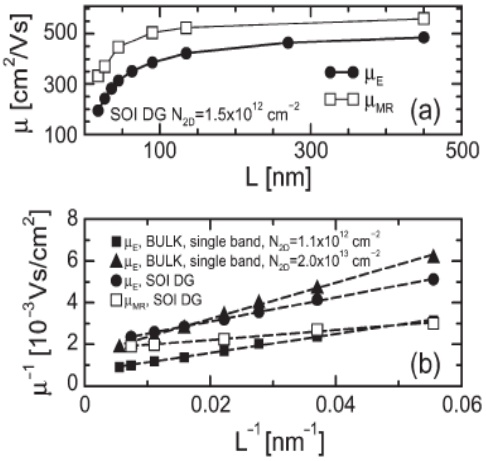


Figure 4.27

Figure taken from Zilli et.al [49]. Top: MR and electrical mobility is extracted from MSMC simulation showing similar behavior as predicted by the analytical models derived here. Bottom: Shows the channel length dependence better explained by Shur like model using a channel length dependent component μ_L and a channel length independent component. The magnitude of the same cannot be compared with the analytical expression as the simulation conditions and condition for the validity of the analytical expression are not the same.

Conclusion and future prospective

This chapter covered some of the state of the art extraction methods in the linear regime for the extraction of mobility. We saw different state of the art linear regime mobility- gate voltage models. The impact of series resistance and its gate voltage dependence in the advanced MOSFETs was discussed in this chapter. We discussed a new modified Y function based extraction method in the linear regime by taking into account the gate voltage dependent series resistance. We also studied the impact of higher order effects like variation of Θ parameters with channel lengths and others on the extraction method. We have also seen split CV method for the extraction of μ_{eff}, and also for t_{ox}, L_{eff}. We have discussed the extraction of η parameter for effective field for sSOI devices done for the first time and found it to be 0.8 while value of η for Bulk, FDSOI, and strained Si bulk is 0.5, 0.6 and 0.7 respectively. This was attributed to the carrier sub band occupation and resulting carrier confinement that causes increased SR and phonon scattering.

The mobility-temperature model is explored for extraction of scattering mechanisms by making use of low field mobility μ_0. This model is extended to gain factor β which is a directly measurable quantity. This approach is attractive for characterization in sub 32nm nodes as we do not need prior knowledge of physical dimensions in estimating the scattering mechanisms in the channel. This method with further improvement would enable a fully experimental extraction that would meet challenges for characterization in the future technology nodes.

Standard magnetoresistance measurement for extraction of magnetoresistance mobility μ_{MR} was also been discussed. Magnetoresistance measurement is attractive for future technology nodes as we can directly measure the channel mobility without the need for physical dimensions. The magnetoresistance model for sub 32nm regime MOSFETs however needs to be improved to take into account ballistic transport. Therefore an analytical model for magnetoresistance at low magnetic field for ballistic/quasi ballistic MOSFET is derived in this chapter in the context of previous simulation results under the limit of vanishing electric field valid in the linear regime. This new linear regime magnetoresistance model with further improvements will help in extending the magnetoresistance characterization to sub 32nm technology nodes.

Appendix A

We have the drain current expression for ballistic MOSFET as below. The injection velocity is averaged over the injection angle to get the average x-directed velocity of the carriers. This averaging is modified by taking Θ_{max} as the lower limit. This will take care directly the back scattered carriers due to magnetic field.

$$ID^{+} = \frac{2qW}{h}\int_{Ec}^{\infty} \pi\hbar \frac{D_{2D}(E)\left(\pi - \Theta_{\max}(E)\right)}{2\pi} \frac{\int_{\Theta_{\max}(E)}^{\pi} v(E)\sin\theta d\theta}{\int_{\Theta_{\max}}^{\pi} d\theta} f_1(E)dE$$

$$ID^{+} = \frac{2qW}{h}\pi\hbar\frac{D_{2D}}{2\pi}\int_{Ec}^{\infty}\left[1+\cos\left(\Theta_{\max}(E)\right)\right]v(E)\,f_1(E)dE$$

$$= \frac{2qW}{h}\pi\hbar\frac{D_{2D}}{2\pi}\int_{Ec}^{\infty}\left[1+\sqrt{1-\left(\frac{qBL_{ch}}{m^{*}v(E)}\right)^{2}}\right]v(E)\,f_1(E)dE \text{ using Eq 4.41.}$$

$$= \frac{2qW}{h}\pi\hbar\frac{D_{2D}}{2\pi}\int_{Ec}^{\infty}\left[v(E)\,f_1(E)dE+\sqrt{1-\left(\frac{qBL_{ch}}{m^{*}v(E)}\right)^{2}}\,v(E)\,f_1(E)dE\right]$$

$$= qW\frac{D_{2D}}{2\pi}\left[\int_{Ec}^{\infty}\left[v(E)\,f_1(E)dE\right]+\int_{Ec}^{\infty}\sqrt{1-\left(\frac{qBL_{ch}}{m^{*}v(E)}\right)^{2}}\,v(E)\,f_1(E)dE\right]$$

Taking v(E) in terms of $E\text{-}E_C$, using spherical parabolic band assumption.

$$= qW\frac{D_{2D}}{2\pi}\left[\int_{Ec}^{\infty}\left[\sqrt{\frac{2}{m^{*}}}(E-Ec)^{1/2}\,f_1(E)dE\right]+\frac{1}{m^{*}}\int_{Ec}^{\infty}\sqrt{2m^{*}(E-Ec)-(qBL_{ch})^{2}}\,f_1(E)dE\right]$$

The integration of the first term will give $F_{1/2}(\eta_{F1})$, where $F_{1/2}$ is the Fermi-Dirac integral of order ½. The integration of the second term needs to be analyzed more closely.

$$\int_{Ec}^{\infty}\sqrt{2m^{*}(E-Ec)-(qBL_{ch})^{2}}\,f_1(E)dE$$

$$= \sqrt{2m^{*}}\int_{Ec}^{\infty}\sqrt{(E-Ec)-\frac{(qBL_{ch})^{2}}{2m^{*}}}\,f_1(E)dE$$

Substitute $[E\text{-}E_c\text{-}(qBL_{ch})^2/2m^*]/kT = \eta^B$

$$\int_{Ec}^{\infty}\sqrt{2m^*(E-Ec)-(qBL_{ch})^2}\,f_1(E)dE$$

$$=\sqrt{2m^*}\int_{-\frac{(qBL_{ch})^2}{2m^*kT}}^{\infty}\frac{(kT)^{3/2}\sqrt{\eta^B}}{1+e^{\eta^B-\eta_{F1}{}^B}}d\eta^B$$

$$\eta_{F1}{}^B=\frac{E_{F1}-E_C-\frac{(qBL_{ch})^2}{2m^*}}{kT}$$

for $[qBL_{ch})^2/2m^*]/kT \ll 1$ the lower limit of the integral can be approximated to 0.

$$=\sqrt{2m^*}(kT)^{3/2}\int_0^{\infty}\frac{\sqrt{\eta^B}}{1+e^{\eta^B-\eta_{F1}{}^B}}d\eta^B$$

Defining $E\text{-}E_C = \eta$ and $(E_{F1}\text{-}E_C)/kT = \eta_{F1}$. Using change of variables

$$=\sqrt{2m^*}(kT)^{3/2}\int_0^{\infty}\frac{\sqrt{\eta}}{1+e^{\eta-\left(\eta_{F1}-\frac{(qBL_{ch})^2}{2m^*kT}\right)}}d\eta$$

$$=\sqrt{2m^*}(kT)^{3/2}F_{1/2}\left(\eta_{F1}-\frac{(qBL_{ch})^2}{2m^*kT}\right)$$

This is same as evaluating the integral from $(E\text{-}E_C)_{max}$ ie $E_C + \frac{(qBL_{ch})^2}{2m^*}$ instead of E_C.

Therefore we get,

$$ID^+=qW\frac{D_{2D}}{2\pi}\left[kT\sqrt{\frac{2kT}{m^*}}F_{1/2}(\eta_{F1})+kT\sqrt{\frac{2kT}{m^*}}F_{1/2}(\eta_{F1}-\frac{(qBL_{ch})^2}{2m^*kT})\right]$$

$$=qW\frac{N_{2D}}{4}\sqrt{\frac{2kT}{m^*\pi}}\frac{2}{\sqrt{\pi}}F_{1/2}(\eta_{F1})\left[1+\frac{F_{1/2}(\eta_{F1}-\frac{(qBL_{ch})^2}{2m^*kT})}{F_{1/2}(\eta_{F1})}\right]$$

The net drain current is then given by $ID^+\text{-}ID^-$.

$$ID=ID^+-ID^-=qW\frac{N_{2D}}{4}v_T\Im_{1/2}(\eta_{F1})\left[1+\frac{\Im_{1/2}(\eta_{F1}-\frac{(qBL_{ch})^2}{2m^*kT})}{\Im_{1/2}(\eta_{F1})}\right]\left[1-\frac{\Im_{1/2}(\eta_{F2})\left[1+\frac{\Im_{1/2}(\eta_{F2}-\frac{(qBL_{ch})^2}{2m^*kT})}{\Im_{1/2}(\eta_{F2})}\right]}{\Im_{1/2}(\eta_{F1})\left[1+\frac{\Im_{1/2}(\eta_{F1}-\frac{(qBL_{ch})^2}{2m^*kT})}{\Im_{1/2}(\eta_{F1})}\right]}\right]$$

The simplification of terms in the brackets in this expression is given below by expanding the Fermi-Dirac integrals involving small $\frac{(qBL_{ch})^2}{2m^*kT}$. $\Im_{1/2}(\eta_{F1})=\frac{2}{\sqrt{\pi}}F_{1/2}(\eta_{F1})$. Term $\eta_{F2} = \eta_{F1}\text{-}qV_d/kT$.

$$\left[1+\frac{\Im_{1/2}(\eta_{F1}-\frac{(qBL_{ch})^2}{2m^*kT})}{\Im_{1/2}(\eta_{F1})}\right]\approx\left[1+\frac{\Im_{1/2}(\eta_{F1})-\Im_{-1/2}(\eta_{F1})\frac{(qBL_{ch})^2}{2m^*kT}}{\Im_{1/2}(\eta_{F1})}\right]=2-\left(\frac{(qBL_{ch})^2}{2m^*kT}\right)\frac{\Im_{-1/2}(\eta_{F1})}{\Im_{1/2}(\eta_{F1})}$$

and

$$\left[1-\frac{\Im_{1/2}(\eta_{F2})\left[1+\frac{\Im_{1/2}(\eta_{F2}-\frac{(qBL_{ch})^2}{2m^*kT})}{\Im_{1/2}(\eta_{F2})}\right]}{\Im_{1/2}(\eta_{F1})\left[1+\frac{\Im_{1/2}(\eta_{F1}-\frac{(qBL_{ch})^2}{2m^*kT})}{\Im_{1/2}(\eta_{F1})}\right]}\right]=1-\frac{\Im_{1/2}(\eta_{F2})}{\Im_{1/2}(\eta_{F1})}\frac{2-\left(\frac{(qBL_{ch})^2}{2m^*kT}\right)\frac{\Im_{-1/2}(\eta_{F2})}{\Im_{1/2}(\eta_{F2})}}{2-\left(\frac{(qBL_{ch})^2}{2m^*kT}\right)\frac{\Im_{-1/2}(\eta_{F1})}{\Im_{1/2}(\eta_{F1})}}\approx1-\frac{\Im_{1/2}(\eta_{F2})}{\Im_{1/2}(\eta_{F1})}$$

Substituting these simplified expressions in the expression of ID we get,

$$ID=qW\frac{N_{2D}}{2}v_T\Im_{1/2}(\eta_{F1})\left[1-\left(\frac{(qBL_{ch})^2}{2\pi(m^*v_T)^2}\right)\frac{\Im_{-1/2}(\eta_{F1})}{\Im_{1/2}(\eta_{F1})}\right]\left[1-\frac{\Im_{1/2}(\eta_{F2})}{\Im_{1/2}(\eta_{F1})}\right]$$

$$=\frac{qW\frac{N_{2D}}{2}v_T\Im_{1/2}(\eta_{F1})}{\left[1+\left(\frac{(qBL_{ch})^2}{2\pi(m^*v_T)^2}\right)\frac{\Im_{-1/2}(\eta_{F1})}{\Im_{1/2}(\eta_{F1})}\right]}\left[1-\frac{\Im_{1/2}(\eta_{F2})}{\Im_{1/2}(\eta_{F1})}\right]$$

EXPRESSION FOR THE INVERSION CHARGE:

Since we know that the +kx half is filled by the Fermi level of the source, while in the –kx side up to -Θ_{max} is filled by the Fermi level of the source and rest by the Fermi level of the drain. Therefore the total inversion charge at virtual device can be calculated as follows

$$N^+=\frac{WL}{2}\int_{Ec}^{\infty}D_{2D}(E)\,f_1(E)dE=\frac{N_{2D}WL}{2}\Im_0(\eta_{F1})$$

$$N^-=\frac{WL}{2}\int_{Ec}^{\infty}\frac{\Theta_{max}(E)}{\pi}D_{2D}(E)\,f_1(E)dE)+\frac{WL}{2}\int_{Ec}^{\infty}\frac{\pi-\Theta_{max}(E)}{\pi}D_{2D}(E)\,f_2(E)dE)$$

$$N=N^++N^-=\frac{WL}{2}\int_{Ec}^{\infty}D_{2D}(E)\left[1+\frac{\Theta_{max}(E)}{\pi}\right]f_1(E)dE+\frac{WL}{2}\int_{Ec}^{\infty}D_{2D}(E)\left[1-\frac{\Theta_{max}(E)}{\pi}\right]f_2(E)dE$$

$\Theta_{max}(E)$ approximated as per Eq 4.41 for small $\Theta_{max}(E)$.

$$\therefore N=\frac{WL}{2}N_{2D}\left[\Im_0(\eta_{F1})+\frac{qB_zL_{ch}}{2kT}v_T\Im_{-1/2}(\eta_{F1})+\Im_0(\eta_{F2})-\frac{qB_zL_{ch}}{2kT}v_T\Im_{-1/2}(\eta_{F2})\right]$$

If $mv_T>>(qB_zL_{ch})/\pi$ (because $2kT=\pi mv_T^2$).

Then $qN=\frac{qN_{2D}}{2}\left[\Im_0(\eta_{F1})+\Im_0(\eta_{F2})\right]=C_{ox}\left(V_g-V_t\right)$

Therefore $\frac{qN_{2D}W}{2} = \frac{C_{ox}(V_g - V_t)W}{[\Im_0(\eta_{F1}) + \Im_0(\eta_{F2})]}$.

Substituting this in the drain current expression, we get the final drain current expression as below.

$$ID = \frac{WC_{ox}(V_g - V_t)v_T \frac{f_{1/2}(\eta_{F1})}{f_0(\eta_{F1})}}{\left[1 + \left(\frac{(qBL_{ch})^2}{2\pi(m^* v_T)^2}\right)\frac{f_{-1/2}(\eta_{F1})}{f_{1/2}(\eta_{F1})}\right]} \frac{\left[1 - \frac{f_{1/2}(\eta_{F2})}{f_{1/2}(\eta_{F1})}\right]}{\left[1 + \frac{f_0(\eta_{F2})}{f_0(\eta_{F1})}\right]}$$

For Low V_d

$$\begin{aligned} ID &= \frac{WC_{ox}(V_g - V_t)v_T \frac{f_{-1/2}(\eta_{F1})}{f_0(\eta_{F1})}}{\left[1 + \left(\frac{(qBL_{ch})^2}{2\pi(m^* v_T)^2}\right)\frac{f_{-1/2}(\eta_{F1})}{f_{1/2}(\eta_{F1})}\right]} \frac{qV_d}{2kT} \\ &= WC_{ox}(V_g - V_t)v_T \frac{f_{-1/2}(\eta_{F1})}{f_0(\eta_{F1})}\left[1 - \left(\frac{(qBL_{ch})^2}{2\pi(m^* v_T)^2}\right)\frac{f_{-1/2}(\eta_{F1})}{f_{1/2}(\eta_{F1})}\right]\frac{qV_d}{2kT} \\ &= \frac{WC_{ox}(V_g - V_t)}{L_{ch}} \frac{\frac{qL_{ch}v_T}{2kT}\frac{f_{-1/2}(\eta_{F1})}{f_0(\eta_{F1})}}{\left[1 + \left(\frac{(qL_{ch})^2}{2\pi(m^* v_T)^2}\right)\frac{f_{-1/2}(\eta_{F1})}{f_{1/2}(\eta_{F1})}B^2\right]} V_d \end{aligned} \quad \text{(A1)}$$

Therefore the magnetoresistance term is given by

$$\mu_{balMR} = \frac{1}{\sqrt{2\pi}} \frac{qL_{ch}}{m^* v_T}\left(\frac{f_{-1/2}(\eta_{F1})}{f_{1/2}(\eta_{F1})}\right)^{1/2};$$

while ballistic mobility for low V_d is $\mu_{bal} = \frac{qL_{ch}v_T}{2kT}\frac{f_{-1/2}(\eta_{F1})}{f_0(\eta_{F1})}$ and therefore the ratio

$$\frac{\mu_{balMR}}{\mu_{bal}} = \sqrt{\frac{\pi}{2}} \frac{f_0(\eta_{F1})}{\sqrt{f_{1/2}(\eta_{F1}) f_{-1/2}(\eta_{F1})}}.$$

ballistic ratio

$$BR_{LinMR} = 1 - r_{linMR} = 1 - \left(\frac{(qBL_{ch})^2}{2\pi(m^* v_T)^2}\right)\frac{f_{-1/2}(\eta_{F1})}{f_{1/2}(\eta_{F1})}$$

Appendix B

Magnetoresistance mobility in quasi ballistic case is derived using flux theory in the same way as ref [46] by using modified flux and injection velocity shown in Eq B1. The reflection coefficient is mentioned as r_B to distinguish the same from 0 magnetic field case as well as the magneto-back scattering coefficient r_{linMR}. F_{BS}^{+} is flux injected by source (taking into account the back scattering by magnetic field), $\tilde{v}_{TB}^{+}$ is the modified average injection velocity that determines the flux.

$$F_{BS}^{+} = \frac{ID^{+}}{qW} = \frac{N_{2D}}{2} v_T \Im_{1/2}(\eta_{F1}) \left[1 - \frac{(qBL_{ch})^2}{4m^* kT} \frac{\Im_{-1/2}(\eta_{F1})}{\Im_{1/2}(\eta_{F1})} \right] = F_S^{+} - F_{BS}^{-}$$

$$\tilde{v}_{TB}^{+} = v_T \frac{\Im_{1/2}(\eta_{F1})}{\Im_0(\eta_{F1})} \left[1 - \frac{(qBL_{ch})^2}{4m^* kT} \frac{\Im_{-1/2}(\eta_{F1})}{\Im_{1/2}(\eta_{F1})} \right]$$

And flux injected by drain (B1)

$$F_{BD}^{-} = \frac{ID^{-}}{qW} = \frac{N_{2D}}{2} v_T \Im_{1/2}(\eta_{F2}) \left[1 - \frac{(qBL_{ch})^2}{4m^* kT} \frac{\Im_{-1/2}(\eta_{F2})}{\Im_{1/2}(\eta_{F2})} \right] = F_D^{-} - F_{BD}^{+}$$

$$\tilde{v}_{TB}^{-} = v_T \frac{\Im_{1/2}(\eta_{F2})}{\Im_0(\eta_{F2})} \left[1 - \frac{(qBL_{ch})^2}{4m^* kT} \frac{\Im_{-1/2}(\eta_{F2})}{\Im_{1/2}(\eta_{F2})} \right]$$

If we see fig B1, the net flux F^{+}_{BS} encounters a back scattering with coefficient r_B. Similarly for F^{-}_{BD} which is drain injected net flux. The flux F^{-}_{BS} is that part of source injected flux (F^{+}_{S}) which is back deflected by the magnetic field (such that $F^{+}_{BS} = F^{+}_{S} - F^{-}_{BS}$). Similarly for drain ($F^{-}_{BD} = F^{-}_{D} - F^{+}_{BD}$). The sign (+) is for drain direction and (-) is for source direction. This flux has a complicated trajectory (semicircle to first order) and in order to have a simplified expression, its trajectory is divided into two halves. The upper half is a forward going flux and get back scattered with a coefficient r'_B back to source (similarly in drain) by definition. Remaining flux $1-r'_B$ continues the trajectory on average and constitutes reverse flux due to magnetic field in the lower half. Therefore lower half flux is a reverse flux and encounters a scattering with same probability r'_B as upper flux but this time get reflected towards drain (as per the definition of r'_B) and constitutes a forward flux (similarly for drain side it will be backward flux). From the figure we can see that r_B is similar to conventional r_{lin}, however r'_B is for a more complicated case. F^{+} is the net drain directed flux and F^{-} is the net reverse flux including the transmitted drain flux viewing at the source side.

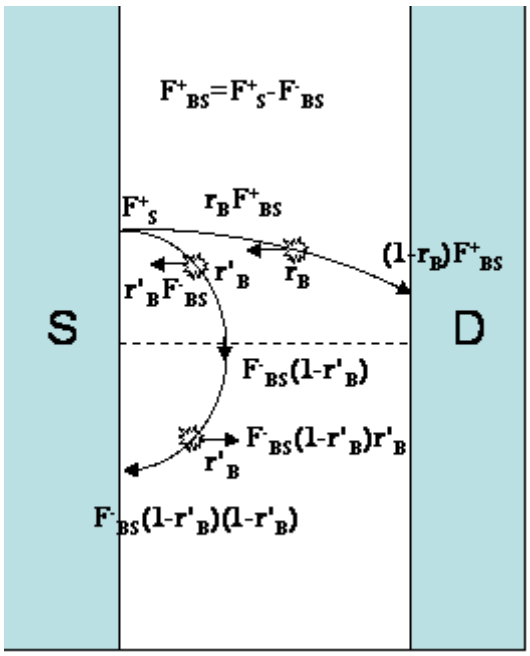


Figure B1

Different scenarios of scattering encounter for the carriers are shown. The carrier with high energy encounter a scattering with probability r_B and is somewhat similar to the conventional back scattering. For the carriers that are back deflected by magnetic field the scenario is more complex. For simplicity the total trajectory is divided into two parts. In the upper part the back scattering deflects the carrier back to the source itself with probability r'_B while in the bottom half the carrier is treated as reverse flux and the back scattering would cause the flux to go forward with probability r'_B. Similarly for drain induced carriers.

From figure B1 we have at the source side as given in Eq B2a.

$$F^+ = F_S^+ + r_B^{'}\left(1 - r_B^{'}\right)F_{BS}^-$$
$$F^- = r_B F_{BS}^+ + \left(1 - r_B\right)F_{BD}^- + r_B^{'}\left(1 - r_B^{'}\right)F_{BD}^+ + r_B^{'}F_{BS}^- + \left(1 - r_B^{'}\right)\left(1 - r_B^{'}\right)F_{BS}^- \quad \text{(B2a)}$$

For F^+ it is easy to see from the fig B1. In F^-, the first and second terms are obvious. The third term is similar to the second term of the F^+ case except this is from the drain side. Fourth and fifth terms are shown in fig B1. The term $[F_{BS}+\text{-}F_{BD}^-]>>[F_{BS}^-\text{-}F_{BD}^+]$ since magnetic field small. But $[F_{BS}+\text{-}F_{BD}^-]$ is comparable to F_{BS}^- since we are considering low V_d.

$$\therefore ID = qW(F^+ - F^-) = qW\left[\left(1 - r_B\right)\left(F_{BS}^+ - F_{BD}^-\right) + F_{BS}^-\left(1 - r_B^{'}\right)r_B^{'} + \left(1 - r_B^{'}\right)r_B^{'}\left(F_{BS}^- - F_{BD}^+\right)\right]$$
$$\approx qW\left(1 - r_B\right)\left(F_{BS}^+ - F_{BD}^-\right) + qWF_{BS}^-\left(1 - r_B^{'}\right)r_B^{'} \quad \text{(B2b)}$$

Therefore the drain current can be expressed with the modified flux. The expressions for different flux are given in B1. Therefore from B2b and B1 we have

$$
\begin{aligned}
ID &= qW(1-r_B)\left(F_S^+ - F_{BS}^- - F_{BD}^-\right) + qWF_{BS}^-(1-r_B')r_B' \\
&= qW(1-r_B)\left(F_S^+ - F_{BS}^-\left[1-\frac{(1-r_B')r_B'}{(1-r_B)}\right] - F_{BD}^-\right);\ \mu_{balMR}^2(\eta_{F1})B^2 = \frac{(qBL_{ch})^2}{4m^*kT}\frac{\Im_{-1/2}(\eta_{F1})}{\Im_{1/2}(\eta_{F1})} \\
&= qW(1-r_B)\frac{N_{2D}}{2}v_T\Im_{1/2}(\eta_{F1})\left(1-\mu_{balMR}^2(\eta_{F1})B^2\left[1-\frac{(1-r_B')r_B'}{(1-r_B)}\right] - \frac{\Im_{1/2}(\eta_{F2})}{\Im_{1/2}(\eta_{F1})}\left(1-\mu_{balMR}^2(\eta_{F2})B^2\right)\right) \\
&= qW(1-r_B)\frac{N_{2D}}{2}v_T\Im_{1/2}(\eta_{F1})\left(1-\mu_{balMR}^2(\eta_{F1})B^2\left[1-\frac{(1-r_B')r_B'}{(1-r_B)}\right]\right)\left(1-\frac{\Im_{1/2}(\eta_{F2})}{\Im_{1/2}(\eta_{F1})}\frac{\left(1-\mu_{balMR}^2(\eta_{F2})B^2\right)}{\left(1-\mu_{balMR}^2(\eta_{F1})B^2\left[1-\frac{(1-r_B')r_B'}{(1-r_B)}\right]\right)}\right) \\
&\approx qW(1-r_B)\frac{N_{2D}}{2}v_T\Im_{1/2}(\eta_{F1})\left(1-\mu_{balMR}^2(\eta_{F1})B^2\left[1-\frac{(1-r_B')r_B'}{(1-r_B)}\right]\right)\left(1-\frac{\Im_{1/2}(\eta_{F2})}{\Im_{1/2}(\eta_{F1})}\right)
\end{aligned}
$$

The ratio $\dfrac{\left(1-\mu_{balMR}^2(\eta_{F2})B^2\right)}{\left(1-\mu_{balMR}^2(\eta_{F1})B^2\left[1-\dfrac{(1-r_B')r_B'}{(1-r_B)}\right]\right)} \approx 1$

$$
\therefore ID = qW(1-r_B)\frac{N_{2D}}{2}v_T\Im_{1/2}(\eta_{F1})\left(1-\mu_{balMR}^2(\eta_{F1})B^2\left[1-\frac{(1-r_B')r_B'}{(1-r_B)}\right]\right)\left(1-\frac{\Im_{1/2}(\eta_{F2})}{\Im_{1/2}(\eta_{F1})}\right) \qquad \text{(B3)}
$$

The inversion carriers at virtual source can be obtained from the modified flux and injection velocity in the same way as [46] using Eq B2a. The total inversion carrier at source n(0) then can be approximated, assuming contributions due to F^-_{BS} and F^+_{BD} negligible. This is shown below (carriers contributing to $F_{BS}^- = \mu^2_{balMR}(\eta_{F1})B^2\, n_S^+$ and similarly for drain).

$$
\begin{aligned}
&n_S^+ = \frac{F_{BS}^+}{\tilde{v}_{TB}^+};\ n_D^- = \frac{F_{BD}^-}{\tilde{v}_{TB}^-}; \\
&n(0) = (1+r_B)n_S^+ + (1-r_B)n_D^- + (1-r_B)\left[\mu_{balMR}^2(\eta_{F1})B^2 n_S^+ - \mu_{balMR}^2(\eta_{F2})B^2 n_D^-\right] + r'_B(1-r'_B)\mu_{balMR}^2(\eta_{F2})B^2 n_D^- \\
&\quad \approx (1+r_B)n_S^+ + (1-r_B)n_D^- \\
&\quad = \frac{N_{2D}}{2}\left((1+r_B)\Im_0(\eta_{F1}) + (1-r_B)\Im_0(\eta_{F2})\right) \\
&qWn(0) = WC_{ox}(V_g - V_t) = qW\frac{N_{2D}}{2}\left((1+r_B)\Im_0(\eta_{F1}) + (1-r_B)\Im_0(\eta_{F2})\right) \\
&qW\frac{N_{2D}}{2} = \frac{WC_{ox}(V_g - V_t)}{\left((1+r_B)\Im_0(\eta_{F1}) + (1-r_B)\Im_0(\eta_{F2})\right)}
\end{aligned}
$$

Therefore substituting the expression into the drain current expression of Eq B3, we get the final expression as shown in Eq B4.

$$ID = WC_{ox}\left(V_g - V_t\right)\frac{\left(1-r_B\right)}{\left(1+r_B\right)}v_T\,\frac{\Im_{1/2}(\eta_{F1})}{\Im_0(\eta_{F1})}\left[\left(1-\mu_{balMR}^2(\eta_{F1})B^2\left[1-\frac{\left(1-r_B^{'}\right)r_B^{'}}{\left(1-r_B\right)}\right]\right)\right]\frac{\left[1-\frac{\Im_{1/2}(\eta_{F2})}{\Im_{1/2}(\eta_{F1})}\right]}{\left[1+\frac{1-r_B}{1+r_B}\frac{\Im_0(\eta_{F2})}{\Im_0(\eta_{F1})}\right]}$$

$$= \frac{WC_{ox}\left(V_g - V_t\right)\frac{\left(1-r_B\right)}{\left(1+r_B\right)}v_T\,\frac{\Im_{1/2}(\eta_{F1})}{\Im_0(\eta_{F1})}\frac{\left[1-\frac{\Im_{1/2}(\eta_{F2})}{\Im_{1/2}(\eta_{F1})}\right]}{\left[1+\frac{1-r_B}{1+r_B}\frac{\Im_0(\eta_{F2})}{\Im_0(\eta_{F1})}\right]}}{\left[\left(1+\mu_{balMR}^2(\eta_{F1})B^2\left[1-\frac{\left(1-r_B^{'}\right)r_B^{'}}{\left(1-r_B\right)}\right]\right)\right]}$$

$$= \frac{W}{L}C_{ox}\left(V_g - V_t\right)\left(1-r_B\right)\frac{\tilde{v}_T\,\frac{\Im_{-1/2}(\eta_{F1})}{\Im_{1/2}(\eta_{F1})}}{\left[\left(1+\mu_{balMR}^2(\eta_{F1})B^2\left[1-\frac{\left(1-r_B^{'}\right)r_B^{'}}{\left(1-r_B\right)}\right]\right)\right]}\frac{qL}{2kT}V_d;\ Linear \qquad \text{(B4)}$$

The expression for quasi ballistic MR can be therefore derived for linear regime using the quasi ballistic drain current expression of B4. From the quasi ballistic expression assuming $r_B = r_B'$, we have Eq B5 using Shur like model.

$$\mu_{QBMR}(\eta_{F1}) = \sqrt{(1-r_B)}\,\mu_{balMR}(\eta_{F1}) \approx (1-\frac{r_B}{2})\,\mu_{balMR}(\eta_{F1})$$

$$\mu_{QBMR} = \left[\frac{1}{\mu_{balMR}} + \frac{1}{\mu_{ddMR}}\right]^{-1};\ \mu_{balMR} \propto L \qquad \text{(B5)}$$

The expression for effective mobility will be as shown in B6.

$$\mu_{eff}(B) = \frac{\mu_{effB0}}{1+\mu_{QBMR}^2 B^2};\ \mu_{effB0} = \left(1-r_B\right)\mu_{bal}(\eta_{F1}) \qquad \text{(B6)}$$

Reference

[1] Y.Taur et.al "New shift and ratio method for MOSFET Channel-Length Extraction", IEEE EDL v13, n5, p267, 1992.

[2] William P.N. Chen et.al, "A New series Resistance and Mobility Extraction Method by BSIM Model for Nano-Scale MOSFETs" VLSI TSA, pp 143-144, 2006

[3] Da-Wen Lin, M.L. Cheng, S.W. Wang, C.C. Wu and M.J. Chen "A Constant-Mobility Method to Enable MOSFET Series-Resistance Extraction" IEEE EDL, vol28, n12, p1132, 2007

[4] D.Fleury, A.Cros, G.Bidal, J Rosa and G.Ghibaudo "A New technique to Exact the Source/Drain Series Resistance of MOSFETs" IEEE EDL vol32, n9, p975, 2009

[5] Clifford Y.Hwang, Tsung-Chia, and Jason C. S. Woo "Extraction of Gate Dependent Source/Drain Resistance and Effective Channel Length in MOS Device at 77K" IEEE TED, vol42, n10, p1863, 1995

[6] P K McLarty, S. Cristoloveanu, O Faynot, V Misra, J R Hauser and J J Wortman "A Simple Parameter Extraction Method for Ultra Thin Oxide MOSFETs" Sol Stat Elec, vol 38, n6, p1175, 1995

[7] G.Ghibaudo, "New Method for the extraction of MOSFET Parameters" Elec Letters, vol 24 n9, p543 (1988)

[8] Tong-Chern Ong, Ping K Ko, and Chenming Hu, "50 A° Gate-Oxide MOSFET's at 77K" IEEE TED, vol34, n10, p2129, 1987

[9] T Tanaka, K Goto, R Nakamura, S Satoh "Novel Extraction Method for size dependent Mobility Based on BSIM3-Like Compact Model", JJAP vol44, n4, p2424, 2005

[10] D. Fleury, A. Cros, H. Brut and G. Ghibaudo " New Y-Function Based Methodology for Accurate Extraction of electrical Parameters on Nano-Scaled MOSFETs" ICMTS, p160-165, 2008

[11] C.Mourrain, B.Cretu, G.Ghibaudo and P.Cottin "New Method For Parameter Extraction in Deep Submicrometer MOSFETs " ICMTS, p181-186, 2000

[12] N. Subramanian, G. Ghibaudo, M. Mouis " Parameter extraction of nano-scale MOSFETs using modified Y-function method " IEEE ESSDERC, p309-312, 2010

[13] K.Rais, F.Balestra, G.Ghibaudo, "On the high electric field mobility behaviour in Si MOSFETs from room to Liquid Helium temperature" Solid State Physics, p217,1994

[14] Ayoub Emrani, F.Balestra, G.Ghibaudo, "Generalized Mobility Law for Drain Current Modeling in Si MOS Transistors from Liquid Helium to Room Temperatures" IEEE TED vol.40 n3 p564 1993

[15] Takuji Tanaka, "Novel parameter extraction method for low field drain current of nano-scaled MOSFETs" IEEE ICMTS, p265, 2007

[16] G.Reichert, T.Ouisse, "Relationship Between Empirical and Theoretical Mobility Models in Silicon Inversion Layers", IEEE TED, vol 43, n9, p1394, 1996

[17] K.Rais, G.Ghibaudo, F.Balestra, "Surface Roughness Mobility Model for Silicon MOS Transistors" Solid State Physics (a), p853, 1994

[18] S.Takagi, "On the Universality of Inversion Layer Mobility in Si MOSFET: Part1 – Effects of Substrate Impurity Concenteration" IEEE TED, vol 41 n12 p2357 1994

[19] T Ando, A.B.Fowler, F.Stern, "Electronic Properties of 2D systems" Rev Modern Phy, Vol 54, n2,p437, 1982

[20] D.S.Jeon, D.E.Burk, "MOSFET Electron Inversion Layer Mobilities-A Physically based Semi-Empirical Model for a Wide Temperature Range" IEEE TED vol 36, n8, p1456, 1989

[21] Jae Woo Lee, D.Jang, M.Mouis, G T Kim, T.Chiarella, T.Hoffmann, G.Ghibaudo, "Experimental Analysis of Surface Roughness Scattering in FinFET Devices" ESSDERC, p305, 2010

[22] S D Kim, Jason C S Woo, "Advanced Model and Analysis of Series Resistance for CMOS Scaling Into Nanometer Regime – Part I-Theoretical Derivation", IEEE TED vol 49 n3 p457 2002

[23] S D Kim, Jason C S Woo, "Advanced Model and Analysis of Series Resistance for CMOS Scaling Into Nanometer Regime – Part II- Quantitative Analysis", IEEE TED vol 49 n3 p467 2002

[24] I.M.Hafeez, G.Ghibaudo, F.Balestra and M.Haond "Impact of LDD structures on the operation of silicon MOSFETs at low Temperature " SSE, vol38, n2, p419, 1995

[25] Fang-Shi J Lai, Jack Yuan-Chen Sun, "An Analytical One-Dimensional Model for Lightly Doped Drain (LDD) MOSFET Devices" IEEE TED vol32, n12, p2803, 1985

[26] K.Varahramyan, E J Verret, "A model for specific contact resistance applicable for Titanium Silicide – Silicon contacts" SSE vol39,n11,p1601 1996

[27] S.Mudanai et al, "Models for Electron and Hole Mobilities in MOS Accumuation Layers", IEEE TED, vol46, n8, p1749, 1999

[28] Jan Koomen, "Investigation of MOST channel conductance in weak inversion" Solid State Electronics, Vol 16, p801, 1973

[29] F.Boeuf et al, "Optimization of Bulk+/SON Integration for Low stand-by Power (LstP) Application" Proc.SSDM (Ext Abs), p1030, 2009

[30] Y M Meziani et al, "Magnetoresistance Mobility Measurements in Sub 0.1μm Si MOSFETs " IEEE ESSDERC, p157, 2004

[31] D C Look "Schottky-Barrier profiling techniques in semiconductors: Gate current and parasitic resistance effects" Journal of applied physics, vol 57, p377, 1985

[32] G Ghibaudo "Transport in the inversion layer of a MOS transistor: use of Kubo-Greenwood formalism" J Phys C: Solid State Physics, vol 19, p769-780, 1986

[33] K Huet et al, "Monte Carlo study of apparent magnetoresistance mobility in the nanometer scale metal oxide semiconductor field effect transistors" Journal of Applied Physics vol 104 – 044504, 2008

[34] Wipa.Chaisantikulwat et al, "Magnetoresistance Technique for Mobility Extraction in Short Channel FDSOI Transistors", ESSDERC, p569, 2005

[35] R Versari, B Ricco, "MOSFET's Negative Transconductance at Room Temperature " IEEE TED, vol46, n6, p1189, 1999

[36] G Ghibaudo et.al "Improved Method for the Oxide Thickness Extraction in MOS Structures with Ultrathin Gate Dielectrics" IEEE Trans on Semicon Manufacturing, vol13, n2 p152, 2000

[37] E. Vincent et .al "On The Oxide Thickness Extraction In Deep-Submicron Technologies" Proc IEEE int. Conference on Microelectronic Test Structures, vol 10, p105 1997

[38] J koomen, "Investigation of the MOST Channel Conductance in the Weak Inversion" Solid State Electronics, vol16, n7, p801 1973

[39] K Romanjek et al, "Improved Split CV Method for Effective Mobility Extraction in Sub-0.1 μm Si MOSFETs" IEEE EDL vol25 n8 p583 2004

[40] D Fleury et al, "Automatic Eextraction Methodology for Accurate Measurements of Effective Channel Length on 65-nm MOSFET Technology and below" IEEE Transaction on Semicondductor Manufacturing vol21 n4 p504 2008

[41] F Pregaldiny, C Lallement, D Mathiot "A simple efficient model of parasitic capacitances of deep – submicron LDD MOSFETs" Solid State Electronics vol 46, p2191, 2002

[42] A.Emrani, G Ghibaudo, F Balestra "On the Universal electric field dependence of the electron and hole effective mobility in MOS Inversion Layers" Solid State Electronics vol 37, n1, p111, 1994

[43] K Bennamane, M DeMichielis, G Ghibaudo and D Esseni, "Extraction of η parmeter characterizing μ_{eff} against E_{eff} curves in strained Si nMOS " Electronics Letters, vol44, n20, p1219, 2008

[44] K Bennamane, G Ghibaudo and A Benfdila, "Method for extraction of η parameter characterizing μ_{eff} against E_{eff} curves in FD-SOI Si MOS devices " Electronics Letters vol 45 n12 p655 2009

[45] A.Cros et al. "Unexpected Mobility degradation for very short devices: A new challenge for CMOS scaling", IEDM Tech. Digest, DOI:10.1109/IEDM. 2006.346872 2006

[46] A Rahman and M S Lundstrom " A compact scattering model for the nanoscale Double-Gate MOSFET " IEEE TED vol 49, n3 p481 2002

[47] J P McKelvey, J C Balogh, "Flux Method for the analysis of Transport Problems in Semiconductors in the presence of Electric Fields" Physical Rev, vol137, n 5A, p A1555, 1965

[48] P.Batude et.al "Advances in 3D CMOS Sequential Integration" IEDM, 10.1109/IEDM.2009.5424352, 2009

[49] M Zilli, D Essni, P Palestri and L Selmi "On the Apparent Mobility in Nanometric n-MOSFETs" IEEE EDL vol28 n11 p1036 2007

[50] L.Thevenod et.at "Magnetoresistance mobility extraction on TiN/HfO_2/SiO_2 metal-oxide-semoconductor field effect transistors" APL vol 90 – 152111, 2007

5 Experimental Extraction in Saturation Regime

The transport in the saturation regime is different from the linear regime due to high field effects as discussed in chapter 3. Since on current of MOSFET depends on its saturation regime parameters, it is important to study the saturation regime properties of the MOSFET. This chapter deals with saturation regime parameter extraction namely saturation velocity v_{sats}, and DIBL. Self-heating effect is also studied. Some of the state of the art extraction methods for v_{sat}, DIBL and self-heating effect explored in this thesis are discussed in this chapter. The impact of self-heating effect along with DIBL is taken into account and a new extraction method for DIBL and thermal resistance for self-heating is discussed in this chapter. Effect of magnetic field in saturation regime is studied for the first time and it is found that we can extract magnetoresistance mobility even in the saturation regime.

INTRODUCTION:

The extraction in linear regime discussed in the previous section does not directly reflect the saturation regime transport. The transport in saturation is different from linear regime due to high field effects as discussed in section 3. The MOSFET drain current model for sub μm regime is already discussed in section 3. The velocity saturation model is mainly used in this regime especially for deep sub μm MOSFETs. Effects like drain induced barrier lowering DIBL, transistor heating effects are also important parameters to study, earlier one being important for deep sub μm devices and the latter being important mainly for SOI devices.

5.1 Drain Induced Barrier Lowering - DIBL

Drain induced barrier lowering DIBL refers to reduced electrostatic control of the channel by gate and higher influence of drain bias leading to drain bias dependent I_{off} in the off state. Meanwhile in the saturation regime (on state) the drain current will not saturate as expected from long channel pinch off behavior. This is important especially for deep sub μm devices in terms of I_{on}/I_{off} ratio and low power / low leakage requirements and in the extreme case ability to turn off the MOSFET switch. This is said to be caused by channel barrier lowering at source with drain bias. In terms of electrical characterization, DIBL (as well as self-heating, discussed in later section) makes it difficult to extract the intrinsic MOSFET saturation regime parameter unless a proper model to account for DIBL is available. DIBL, however leads to a higher on current. Effect of DIBL in the sub threshold and

saturation regime is shown in fig 5.1 (I_d-V_d and I_d-V_g curves are for different devices). As we see, as channel length reduces DIBL effect increases. This channel length dependent behavior is known as short channel effect (SCE). In the saturation regime the drain current keeps increasing for 50nm device unlike 1µm device where drain current is saturated.

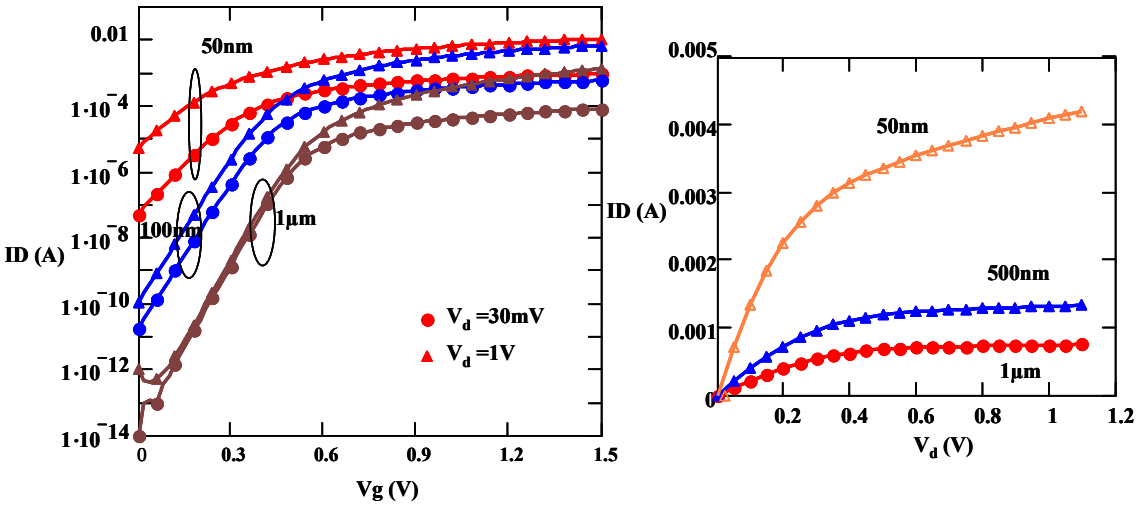


Figure 5.1

Left: The impact of DIBL in the Id-V_g is shown for long to short channel FDSOI devices. As we see 50nm is most affected and for 1µm long device V_d up to 1V has no significant impact on drain current. If we take same V_d, we see that the off-current is larger for short channel length. In the linear regime, the threshold voltage reduces with reducing channel length known as short channel effect. Right: The I_d-V_d characteristic for three channel length Bulk MOSFETs at V_g = 1.1V is shown. As we see for 1µm and 500nm the drain current saturates as expected. But for 50nm the drain current still increases. Therefore it is difficult to study the intrinsic saturation parameters of 50nm device unless this effect is modeled and extracted properly.

DIBL and SCE are shown to be suppressed by using shallow LDD and pocket (halo) implants as well as increased channel doping and ensure good electrostatic control of channel, Liu et al [10]. This has the effect of reduced mobility due to high effective field and impurity scattering. In the case of FDSOI, Si thinned to avoid this effect without the need for the halo and channel doping, thereby providing mobility advantage over bulk. Alternatively we have Double Gate (DG) devices, FinFet, GAA etc which has more than one gate to control the channel and these structures as well as ultra-thin body and box (UTB^2) FDSOI (thin BOX for avoiding BOX fringing fields) proposed for sub 32nm technology nodes. [8][9]

DIBL (λ) is modeled as drain dependent threshold voltage $V_t(V_d)$. Threshold voltage as a function of drain voltage is expressed to first order as $V_t = V_{t0} - \lambda V_d$. In a more general way it can be modeled as eq 5.1. g_{dmes} is the measured output conductance, g_{d0} is the intrinsic output conductance and g_m is the

measured transconductance. The parameter λ is proportional to the electrostatic integrity [4, 9] and reduces with LDD depth or Si thickness, depletion layer depth or channel doping.

$$g_{dmes} = g_{d0} + \lambda g_m \tag{5.1}$$

We can extract DIBL parameter λ generally from the output conductance g_d and transconductance g_m using eq 5.2 [4]. The intrinsic output conductance $g_{d0} = 0$ in saturation. Therefore by plotting the LHS with V_d at high saturation we can extract λ.

$$\frac{g_{dmes}}{g_m} = \frac{g_{d0}}{g_m} + \lambda \tag{5.2}$$

Fig 5.2 shows the ratio g_d/g_m for the bulk MOSFETs down to 40nm.

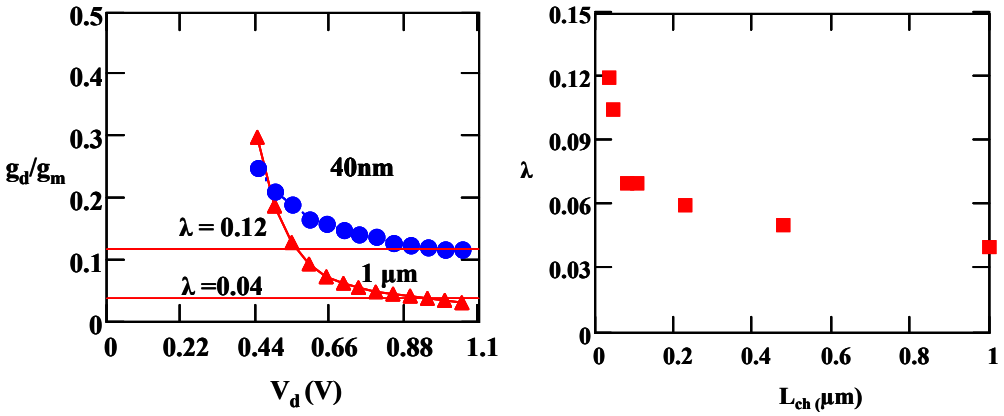


Figure 5.2

Left: g_d/g_m vs V_d for 40nm and 1µm channel length bulk MOSFETs is shown. (Same devices as fig 5.1right) As we see the curve reaches a plateau at saturation. Right: DIBL parameter λ (V/V) for different channel lengths from 1 µm to 40nm using the above method is shown.

As we see DIBL parameter λ increases as channel length reduces (SCE). The DIBL parameter λ for 40nm is around 120mV/V which is high. For well-behaved devices like FDSON bulk+ or GAA/FinFET etc discussed in section 6 this can be as small as 70mV/V. Once λ is obtained, we can extract g_{d0} and reconstruct the I_{DS}-V_d without DIBL using Eq 5.3. The result is shown in fig 5.3.

$$I_{DS}(V_d) = \int_0^{V_d} g_{d0}(u)du \tag{5.3}$$

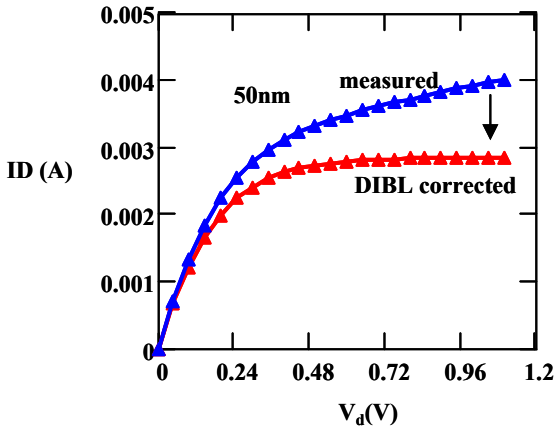


Figure 5.3

Measured I_{DS} -V_d and DIBL corrected Id-V_d for 50nm Bulk MOSFET (same device as fig 5.1 (right) and 5.2) is shown. As we see the DIBL corrected I_{DS} reaches a saturation value at high V_d.

5.2 Self-Heating effect

Due to low thermal conductivity of buried oxide, FDSOI devices suffer from self-heating effect in the On state. The heating effect is not that significant in bulk devices or with thick Si which can conduct the heat [7, 11]. The increased heating will degrade the performance of the MOSFETs as well as cause reliability issues. The thermal conductivity of BOX is 2 orders of magnitude smaller and so most of the heat is has to be conducted through the contacts which eventually heats up the metal interconnects [11]. This has the impact of increasing the delay time, power and reliability of interconnects as well. The simulation model in [11] also had shown that peak temperature inside the channel is much higher than the average channel temperature. However in the measurement point of view, all of these will manifest in the overall change in the I_{DS} and we obtain an equivalent value for the thermal resistance R_{th}. From the measured sensitivity of drain current to ambient temperature dI_{DS}/dT, we can extract the thermal resistance using the method explained in ref [7] using the Eq 5.4.

$$g_d = g_{d0} + R_{th}\frac{\partial I_{DS}}{\partial T}\left(I_{DS} + g_d V_d\right)$$

$$\frac{\partial I_{DS}}{\partial T} = \frac{\frac{dI_{DS}}{dT}}{1+\frac{dI_{DS}}{dT}R_{th}V_d};\; g_{d0} = \frac{\partial I_{DS}}{\partial V_d} \qquad (5.4)$$

$$\therefore R_{th} = \frac{g_{dsat}}{I_{dsat}\frac{dI_{DS}}{dT}};\; g_{d0} = 0\; at\; saturation$$

g_{d0} is the intrinsic output conductance without self-heating, R_{th} is the thermal resistance, $\frac{\partial I_{DS}}{\partial T}$ is the intrinsic sensitivity of drain current to temperature without self-heating. This includes sensitivity of the transport parameters as well as the threshold voltage.

5.3 Combining Self Heating and DIBL-new extraction and correction method for sub-µm MOSFETs

By following the same procedure as ref [7] and including the DIBL effect in the g_d as discussed in the previous section we can combine the two effects in the case of sub µm MOSFETs for the first time as shown in Eq 5.5. Derivation is given in Appendix A.

$$\frac{g_{dsat}}{g_m} = \lambda + R_{th}\frac{dI_{DSsat}}{dT}\frac{I_{DSsat}}{g_m} \tag{5.5}$$

The g_{d0} in the expression 5.4 is expanded to include the DIBL effect. g_m is the measured transconductance. Therefore the overall g_d/g_m ratio vs V_d for a given V_g, and T does not directly give the DIBL parameter λ but it is modified by the self-heating effect term and could even introduce a higher order temperature dependence (like dependence of R_{th} with temperature) to the ratio. We can plot the function g_d/g_m vs $\frac{dI_{DSsat}}{dT}\frac{I_{DSsat}}{g_m}$ for a given V_g and T and for different V_d and obtain the parameters R_{th} from the slope and λ from the y-axis intercept in the saturation region. The result for 55nm FD SON (1µm wide) device is shown in fig 5.4 (See section 6.1 for details of the device). The fig 5.4 (top left) gives the I_d-V_d. The top right figure shows the g_d/g_m ratio vs V_d. The fig 5.4 (bottom) shows the g_d/g_m vs $\frac{dI_{DSsat}}{dT}\frac{I_{DSsat}}{g_m}$ showing linear correlation at high V_d (saturation region). The slope gives the average thermal resistance R_{th} (K/W) and the y-axis intercept gives the DIBL parameter λ (V/V). The thermal resistance obtained in this example is ~ $4.2\text{x}10^5$K/W and corrected λ is 0.2V/V. The plateau in g_d/g_m vs V_d gave 0.15 (see section 6.1) at V_g 1.4V. The value of thermal resistance for these devices seems to be large compared to that reported in ref [7] but if we consider the devices heat dissipating area this value is comparable. I.e. if we take this device with the same area as the device in [7] which is 40µm wide and 0.8µm long then this gives a value of ~ 700K/W and the FDSOI devices in [7] is ~ 3000K/W. This is to be expected as FDSON devices have source and drain grown from the bulk. Therefore thermal conductivity is naturally larger as heat can dissipate through the S/D to the bulk. The buried oxide is present only under the channel region which is thin ~ 25nm (with t_{Si} 15nm) compared to the FDSOI given in the ref [7], which has ~ 400nm of BOX (Although with thick Si 80nm). The effect of series resistance is not studied here.

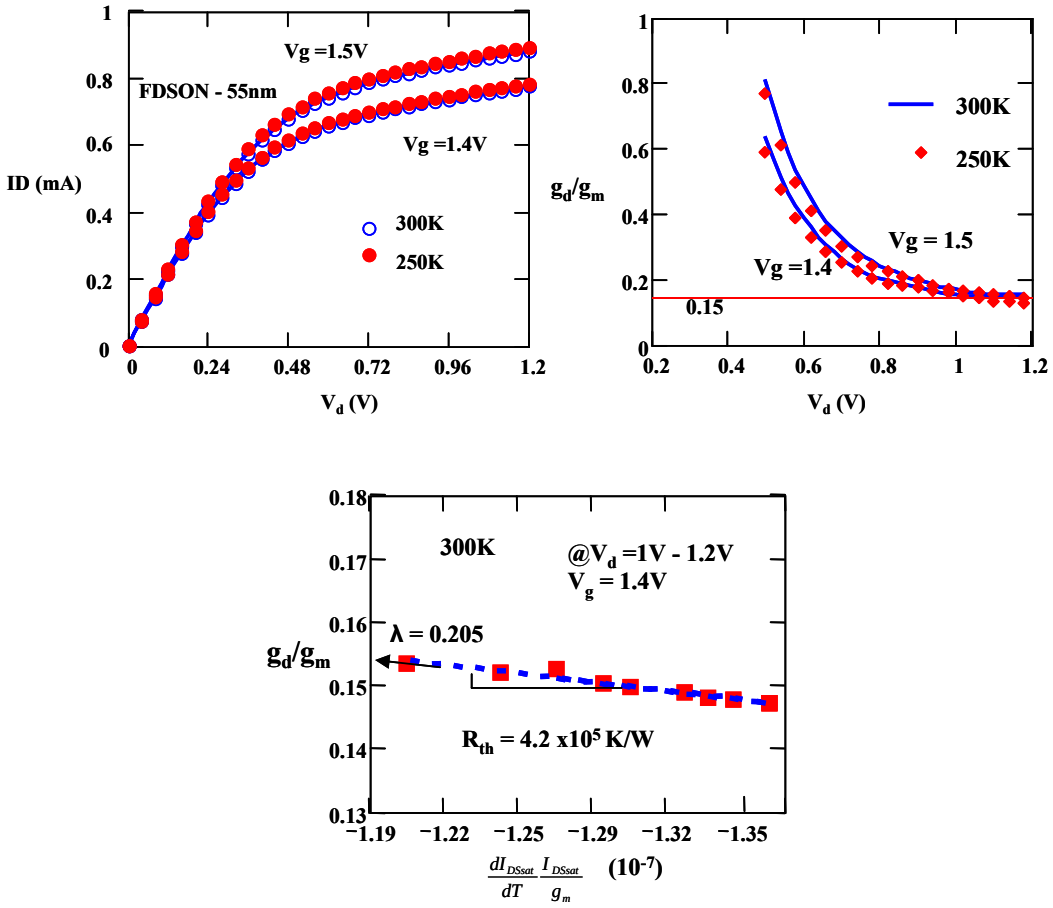


Figure 5.4

Extraction of DIBL and thermal resistance using the new method for 55nm FD SON device is shown. (See section 6.1). Top left: Shows the I_{DS}-V_d for two V_g 1.4 V and 1.5V and temperatures 300K and 250K. Top right: g_d/g_m ratio with V_d. Normally the plateau at high V_d is taken as the DIBL parameter λ. However this has to be corrected for the self heating effect. Bottom: g_d/g_m ratio plotted against the function $\frac{dI_{DSsat}}{dT}\frac{I_{DSsat}}{g_m}$ showing linearity at high V_d. From the slope we can obtain the thermal resistance R_{th} (K/W) and intercept gives the corrected DIBL parameter λ (V/V).

The drain current correction discussed in the previous section therefore has to be modified. Correction for heating effect is more complicated but we can make use of eq A-1 in Appendix A after determining λ and R_{th} and using dI_{DS}/dT as shown in Eq 5.6. The result for the correction is shown in fig 5.5.

$$g_{d0} = g_d - \lambda \frac{g_m}{1+\frac{dI_{DS}}{dT}R_{th}V_d} - R_{th}\frac{\frac{dI_{DS}}{dT}}{1+\frac{dI_{DS}}{dT}R_{th}V_d}\left(I_{DS}+g_dV_d\right) \qquad (5.6)$$

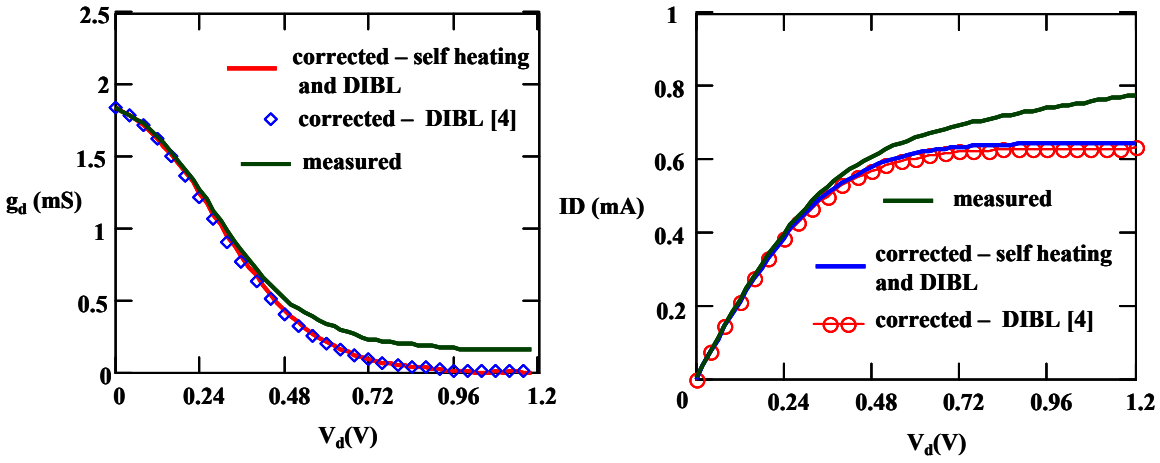


Figure 5.5

Left: The measured g_d and the corrected g_d is shown for FDSON 55nm device (same as fig 5.4). The conductance g_d is corrected using [4] for DIBL using the plateau g_d/g_m and the new method using the eq 5.6. As we see the plateau of g_d/g_m gives a good approximation to the correction for DIBL and self heating. Right: The measured and corrected drain current is shown. The drain current correction using ref [4] under estimates the drain current. The self heating effect correction shows slightly more drain current which is to be expected as self heating generally has the effect of reducing the drain current. However the difference between the to corrections methods is very small. V_g = 1.4V, W = 1µ and L_{ch} = 55nm at 300K.

The fig 5.5 shows the comparison between the correction methods of ref [4] and Eq 5.6. The plateau of g_d/g_m has both the effect of DIBL as well as the self-heating effect. Therefore this value on average can correct for both the effects and give an approximate estimation of I_{DS}. This can be seen from fig 5.5 (right) where the corrected drain currents are very close and the difference is small enough and can be neglected in this case. In the practical scenario the correction in section 5.1 is more convenient compared to the new method.

5.4 Saturation Velocity– v_{sats}

The drain current for sub-micron MOSFETs is discussed in the section 3. The drain current at saturation is characterized by a saturated velocity term v_{sats}. In the case of ballistic transport, this is characterized by the injection velocity and a ballistic ratio to take the back scattering into account for quasi ballistic case. Ref [6] and others define effective saturated velocity v_{sats} as g_m/WC_{ox}. The transconductance is corrected for the series resistance R_S and R_{SD} using the Eq 5.7. g_m is the measured transconductance and g_{mi} is the intrinsic transconductance, g_d is the measured output conductance.

$$g_{mi} = \frac{g_m}{\left(1 - R_S g_m - g_d R_{SD}\right)}; \tag{5.7}$$

Ref [11, 12] uses resistive gate which allows applying a varying gate voltage along the channel length. This ensures a uniform channel and therefore we can take a constant velocity and electric field throughout the channel and calculate the drift velocity and electric field (V_d/L). This method therefore requires special structures and such structure is not feasible in the case of deca-nanometer MOSFETs. Therefore this method can no longer be used. Another method is by using V_{dsat} model [15] discussed in section 3 (Eq 3.7). V_{dsat} model is based on the assumption that the carriers move at saturated velocity near drain. Also the expression is based on gradual channel approximation and this is not true for high V_d. Extraction method based on field model in Eq 3.6 (chapter 3) is proposed in ref [2]. The drift velocity at source v_{ds} is obtained from the drain current model using Eq 5.8a.

$$v_{ds} = \frac{I_{DS}}{WC_{ox}\left(V_g - V_t\right)} \tag{5.8a}$$

The drain current has to be corrected for DIBL/self-heating using the method discussed in the previous section. Pulse IV measurement is another way to avoid the self-heating problem but this measurement is very difficult compared to static measurement.

The electric field at source Es is calculated from the extrapolated I_{D0}-V_d, calculated from linear regime I_{DS}-V_g assuming long channel behavior in the absence of v_{sat} effect using Eq 5.8b and Eq 5.8c.

$$E_s = \frac{I_{D0}}{Lg_{d0}} \tag{5.8b}$$

$$I_{D0}(V_d)\Big|_{V_g} = \int_0^{V_d} g_{dlin}(V_g - u)du \tag{5.8c}$$

L is the channel length and g_{d0} is the conductance corrected for DIBL/Self heating from measured I_{DS}-V_d. $g_{dlin}(V_g)$ is ohmic conductance transfer characteristics calculated from linear regime I_{DS}-V_g measurement. Once the drift velocity and electric field is known, the curve is fitted to the model (Eq 3.6 of chapter 3) in Eq 5.8d at low E_s region where expression of E_s holds good. The saturation velocity v_{sats} is given by μE_C. Mobility μ is extracted from the $v_{ds} - E_s$ from the slope at small E_s in the limit $E_s << E_C$. The result of this extraction is shown in section 6.1 for FDSON devices.

$$v_{ds} = \frac{\mu E_s}{\left[1+\left(\frac{E_s}{E_C}\right)^{\beta}\right]^{1/\beta}} \tag{5.8d}$$

5.5 Magnetoresistance in Saturation Regime

The effect of magnetic field in the saturation regime of MOSFET operation has been studied for the first time with different channel lengths, 1μm down to 50nm, temperature and gate voltage. Study of MOSFETs in the saturation regime helps us understand the performance in terms of I_{on} in its actual context unlike measurements in the linear regime. As explained in the section 3, I_{on} is governed by high field effects in the channel. Both linear and saturation regime extraction have their own merits and demerits and therefore both are necessary in order to have a complete understanding of MOSFET performance. On one side the saturation regime model of the MOSFET is not helpful to understand the underlying scattering mechanisms which is possible in linear regime, while on the other side we can analyze v_{sats} to understand the phenomenon of overshoot, ballistic transport etc that govern the on current which is not possible in the linear regime.

The advantages of MR measurement in the linear regime is already known, that we can directly measure the mobility without any extraction limitations especially in determining the physical dimensions, numerical errors, limitations of model etc. In this context, for the first time, the effect of magnetic field in the saturation regime for sub μm MOSFETs is being investigated and found that we can extract μ_{MR} as a function of drain voltage from linear μ_{MRlin} to saturation μ_{MRsat} μ_{MRsat} is almost independent of the drain voltage V_d. Hess [16] had studied MR effect in hot electron regime previously. However it was not done in the velocity saturation region as the electric field was limited to ~ 200V/cm. For the devices measured here this is not the case as devices are characterized all the way into saturation regime. We have further gone into the regime of non-stationary transport for deep sub-micron MOSFETs and further down to 50nm where perhaps the transport is in quasi ballistic regime. Although measurement result down to 50nm could be explained by v_{sats} model and the observed magnetic field characteristics is explained under this context.

Measurement is done in Grenoble high magnetic field lab, LNCMI, with field ranging from 0T to 11T at near room temperature 270K as well as at 100K and 200K. Bulk MOSFETs with width W = 10μm, high K /Metal gate (EOT 2.4nm), and channel lengths from 1μ down to 50nm is used (so W/L > 5). The self-heating effect (bulk) and series resistance (R_{SD} ~ 8Ω/10μm) are neglected.

Fig 5.6 shows μ_{MR} extraction for 10μm x 0.5μm device with V_d at high inversion for three V_g and temperatures. $I_{DS} - V_d$ is shown in fig 5.6 (top left) at 200K. We can see that I_{DS}-V_d curve varies with magnetic field B from 0T to 11T. Saturation drain current I_{dsat} reduces with field. The relative variation of I_{DS} with B^2 shows linear behavior for both in linear (V_d = 50mV) and saturation region (V_d = 1V) shown in fig 5.6 (top right). The extracted μ_{MR} with V_d for three temperatures is shown in fig 5.6 (bottom) and it shows that μ_{MR} decreases initially with V_d and then saturates in the saturation

region. The gate voltage influence is also seen for μ_{MRsat}. We can see that μ_{MR} increases with decreasing temperature for the whole range of V_d as well as V_g.

The saturation drain current in sub μm regime is governed by velocity saturation effect. Electrical extraction using v_{sats} model described in section 5.3 (Eq 5.8) is performed on the same device for different magnetic fields at 270K. See fig 5.7. Fig 5.7 (left) shows extracted v_{sats} vs B showing a parabolic behavior. Therefore a relative $v_{sats,}$ $\Delta v_{sats}/v_{sats}(B)$ vs B^2 is plotted (similar to $\Delta I_{DS}/I_{DS}$), shown in fig 5.7 (right), and it also shows a linear behavior. The value of the slope gives a μ_{MRsat} ~ 220cm^2/V.s which is close to μ_{MRsat} obtained from drain current in saturation region.

Fig 5.8 shows μ_{MR} extracted as a function of V_d for different channel lengths at 270K and V_g = 1V. We can see that along with the linear regime μ_{MRlin} the saturation region μ_{MRsat} also reduces with the channel length. The linear region μ_{MRlin} at V_d = 50mV and saturation region μ_{MRsat} at V_d = 1.1V, V_g = 1V at 270K are plotted against channel length in fig 5.9 (left). As we see both mobilities are reducing with the channel length in the same manner. In order to understand the correlation between them, μ_{MRsat} vs μ_{MRlin} is plotted as shown in Fig 5.9 (right). Correlation between them is clearly seen however at short channel lengths where non stationary transport effects are expected to set in, μ_{MRsat} is not reducing at the same rate as μ_{MRlin} with channel length. It is visible in the fig 5.9 (right) where μ_{MRsat} shows a tendency to saturate at low μ_{MRlin} (for short channel lengths < 100nm). The threshold voltages of these devices are almost the same all the way down to 50nm with slight RSCE for channel lengths ~ 120nm and less due to halo implants. It is already known that μ_{MRlin} reduction with channel length is mostly due to reduction in mobility due to halo implants, neutral defects, RCS etc.

One possibility for non-linearity in fig 5.9 (right) is that since we are using same V_d in linear regime for comparison of all the channel lengths, we may be observing a field effect with reducing channel length, since the electric field with V_d = 50mV for 50nm is 10^4 V/cm which is already strong, μ_{MRlin} will roll off even faster with channel length. This is however smaller than saturation field especially in these high K based devices with smaller low field mobility, the saturation field is in the order ~ 6 10^4 V/cm (using v_{sats} of 0.65 10^7cm/s with high K for 50nm, fig 5.10 (left) and low field mobility ~ 100cm^2/V.s). This is also confirmed from the fact that the drain current saturates at V_d ~ 0.3-0.5V (~ 5-10 times larger). Low field electrical mobility (μ_0) from Y-function method for a similar 50nm device is ~ 100 cm^2/V.s (@V_d = 5mV). So value obtained here for μ_{MRlin} is comparable but we may observe non-linearity due to additional μ_{MRlin} roll off with channel length due to field effects.

Another possibility for non-linearity is non-stationary effects for sub 100nm. The carrier velocity saturation is suppressed and mobility is larger as discussed in section 3.3 (Eq 3.7). v_{sats} for Si/SiO_2 long channel RG MOSFETs is ~ 0.6-0.8 10^7cm/s [13-15] and in the case of 100nm standard MOSFET

devices, overshoot is already visible with $v_{sats} \sim 10^7 cm.s^{-1}$ [5, 6]. However fig 5.10 (left) does not show this level of overshoot effects for the high K devices used here.

Comparison is made between μ_{MRsat} and v_{sats} for different channel lengths at 0 magnetic field, is shown in fig 5.10 at 270K. v_{sats} here is extracted from $g_m/C_{ox}W$ plateau at high V_d. We can see from fig 5.10 (left) that v_{sats} shows a velocity overshoot like behavior as channel length reduces but it is not significant as in [5, 6]. The magnetoresistance μ_{MRsat} on the other hand reduces just like in the linear regime. Fig 5.10 (right) shows μ_{MRsat} vs v_{sats} and shows that μ_{MRsat} reduces linearly with increasing v_{sats}.

For now fig 5.7 could be explained as below. v_{sats} extracted electrically, which is significantly smaller than the bulk saturation velocity, is just the drift velocity at source at high field. As per [3] the saturation is attributed to source electric field pinning and so the source-side drift velocity is still dependent on the mobility and so magnetic field. This is because at high fields the electric field in the channel is non-uniform for a regular MOSFET and source side electric field does not increase significantly with V_d in saturation creating an apparent velocity saturation. (Note: In the case of RG MOSFETs [13-15] a resistive gate is used in order to have a gate bias gradient along the channel and so assumed to have a uniform channel condition even at high V_d). So when the mobility changes with the magnetic field, we observe the same in extracted v_{sats} with magnetic field.

Fig 5.10 also can be explained with the same argument neglecting non-stationary effect here as overshoot is small. Non linearity of Fig 5.9 (right) can be attributed to field effect in the linear regime mobility for short channel devices due to increased electric field with reduced channel length as V_d is kept constant and large enough. In the saturation region also, reduction of channel length will have the effect of increased magnitude for pinned source field [3] and therefore increased v_{sats}. As the drift velocity at source approaches the saturation, mobility reduces. This will get reflected in μ_{MRsat}. This could explain the fact that measured μ_{MRsat} is decreasing with v_{sats}.

Further study supported by simulation is necessary for complete understanding of these effects as physics of MR in this region is not fully understood. We still do not know the effect of velocity saturation or non-stationary transport on the magnetoresistance. With the experimental results shown here, it is now important to have a deeper study of magnetic field effect in this regime so that we can exploit the full advantages of MR extraction technique especially for future technology nodes.

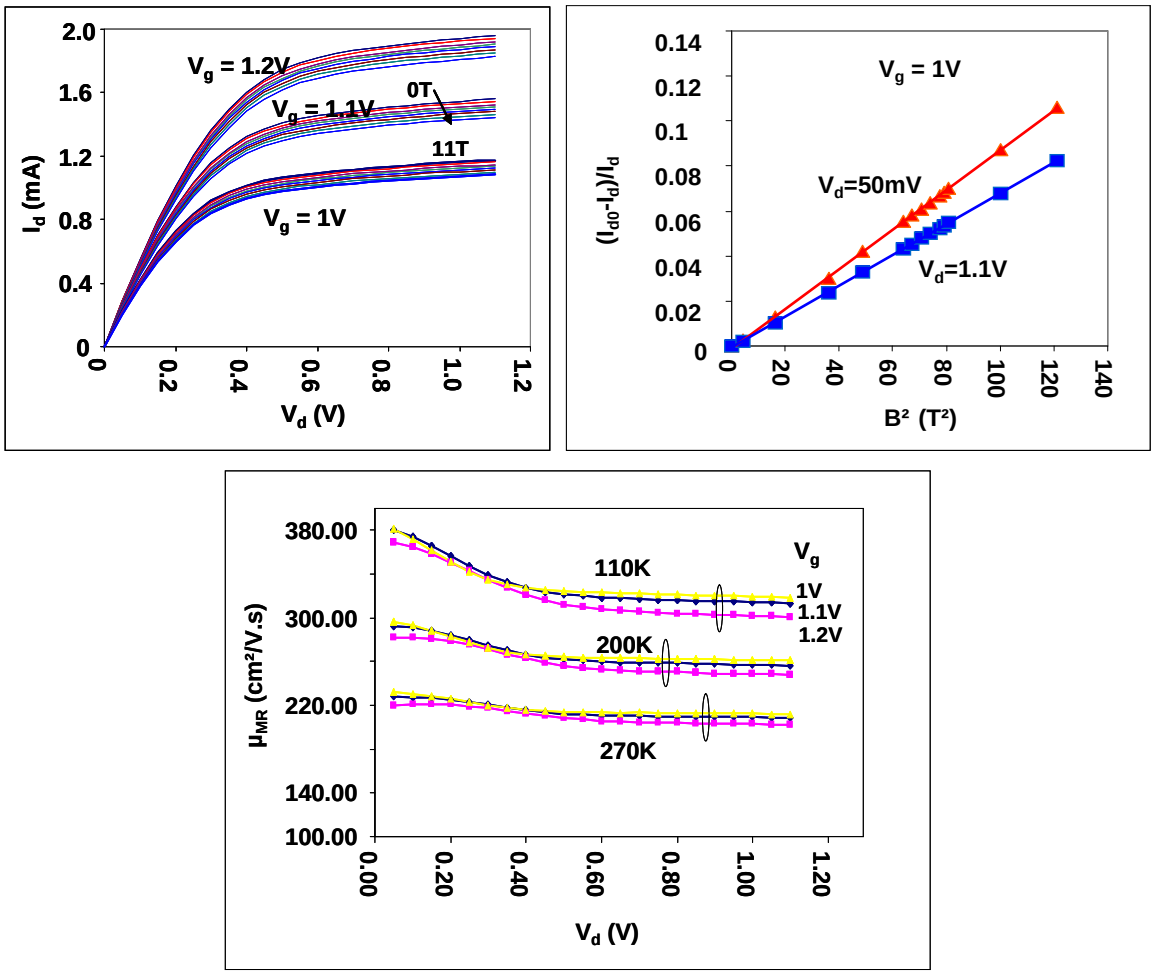


Figure 5.6

Top: Left I_{DS} -V_d for V_g = 1V, 1.1V, 1.2V at 200K, B = 0T to 11T, Top right: The relative I_{DS} with B^2 showing linearity both at V_d = 50mV and 1V. Bottom: Extracted μ_{MR} in $cm^2/V.s$ vs V_d at 110K , 200K and 270K. for V_g = 1V, 1.1V, 1.2V.

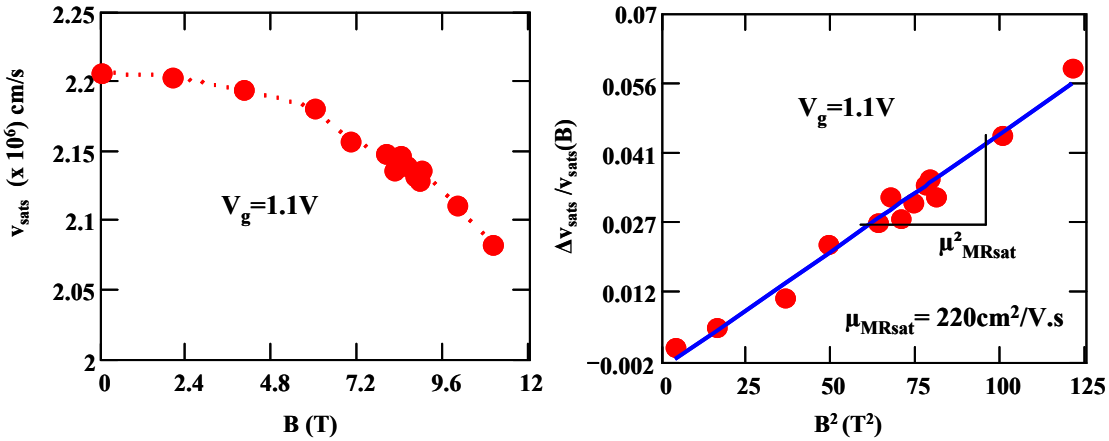


Figure 5.7

Left: Extracted v_{sats} at V_g = 1.1V with B at 270K for 0.5µm device (same as Fig 5.6) Right: relative v_{sats} with B^2 showing linear trend and slope gives μ_{MRsat} in agreement with Fig 5.6.

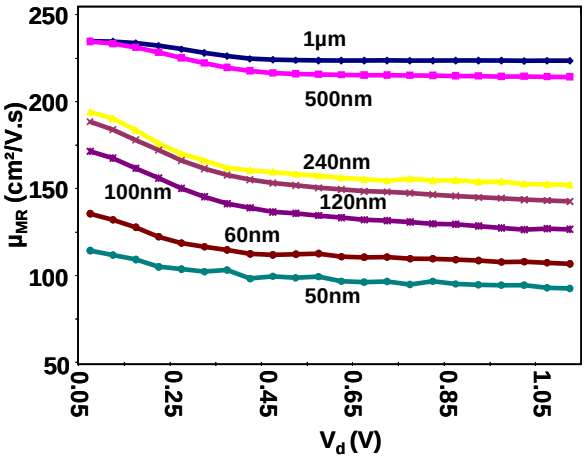


Figure 5.8

Extracted μ_{MR} in $cm^2/V.s$ with V_d at 270K. for V_g = 1V for different channel lengths

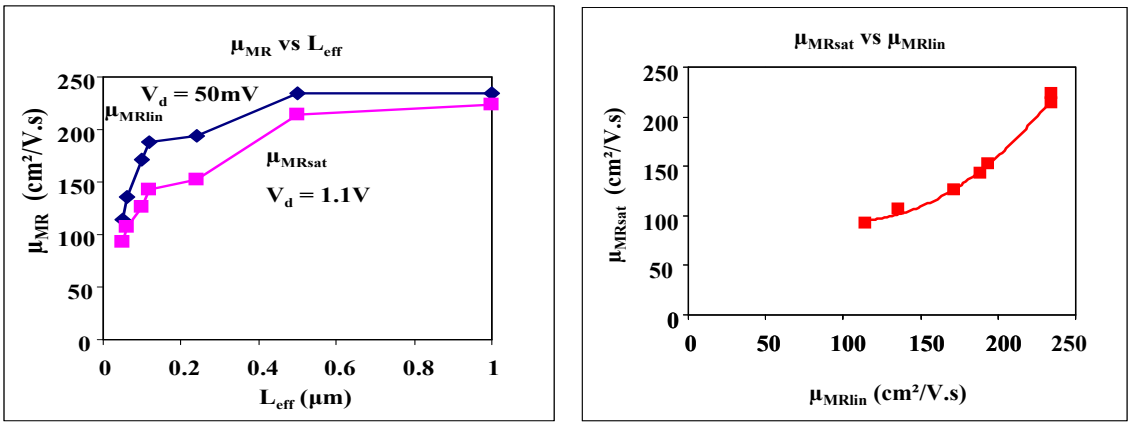


Figure 5.9

Left: Extracted μ_{MRlin} at V_d = 50mV and μ_{MRrsat} at V_d = 1.1V in $cm^2/V.s$ with channel length L_{eff} (µm) at 270K for V_g = 1V.

Right: μ_{MRsat} vs μ_{MRlin}

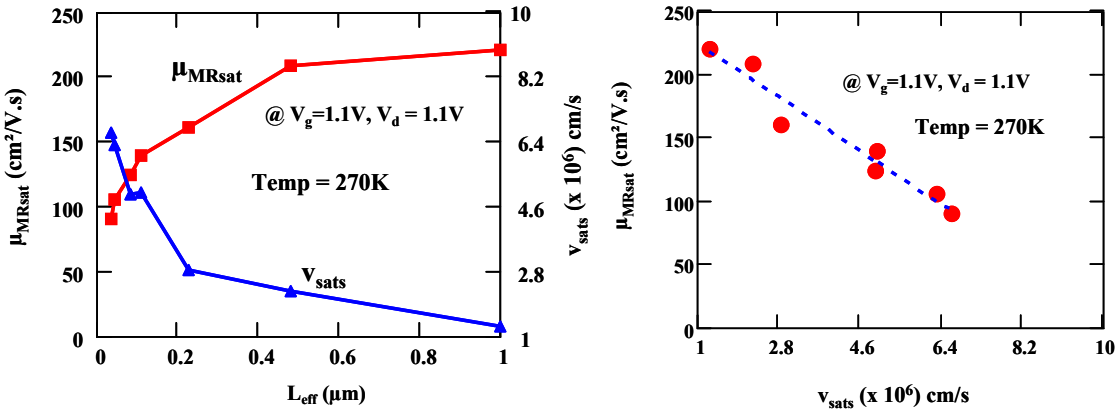


Figure 5.10

Left: Plot μ_{MRsat} and v_{sats} at B =0 vs channel length L_{eff}. v_{sats} is extracted from $gm/C_{ox}W$ plateau at high V_d.

Right: Extracted μ_{MRrsat} at V_d = 1.1V in $cm^2/V.s$ with v_{sats} at 0 field at 270K. for V_g = 1.1V.

Conclusion and future prospective

This chapter focused on the extraction in the saturation regime of MOSFET operation. The on current of MOSFET is one of the main performance parameters and analysis in this regime helps us understand the same in its actual context. Saturation for the sub µm MOSFETs is modeled by using saturated velocity v_{sats} due to high electric fields. However this regime of operation is affected by several factors that make the extraction of key performance parameter v_{sats} difficult. Main factors addressed in this chapter are DIBL and self-heating effect. Current state of the art extraction method is improved by taking into account both DIBL and heating effect for deep sub µm MOSFETs.

The use of magnetoresistance measurement is explored for the first time in the saturation regime and found that we can extract μ_{MR} in the saturation region shown as μ_{MRsat}. μ_{MRsat} shows similar behavior as linear regime μ_{MRlin} with channel length as well as temperature. However correlation between μ_{MRsat} with μ_{MRlin} is nonlinear and apart from the impact of using constant V_d for all the channel lengths we have to take into account non-stationary effects too. The physics of MR in the saturation region is still not fully understood. However experimental results show that it is important to have this study so that we can take full advantage of MR characterization.

Study of both linear and saturation regime is necessary to have a complete understanding of MOSFET performance.

Appendix A

The effect of self-heating and DIBL is combined using same procedures as [4, 7] for the sake of deca-nanometer devices. Therefore the output conductance g_d can be expressed as below including the DIBL and self-heating. The term ΔT is the small average temperature rise in the channel due to power dissipation in the saturation regime. Therefore the output conductance will have three components as shown below, intrinsic g_{d0}, DIBL term through λ and self-heating term through R_{th}. The transconductance will have now two components, intrinsic g_{m0} and self-heating term through R_{th}. The effect of series resistance is not taken into account here.

$$g_d = \frac{dI_{DS}}{dV_d} = \frac{\partial I_{DS}}{\partial V_d} + \frac{\partial I_{DS}}{\partial V_t}\frac{dV_t}{dV_d} + \frac{\partial I_{DS}}{\partial T}\frac{d\Delta T}{dV_d};$$

$$\Delta T = R_{th} V_d I_{DS}$$

$$= g_{d0} + \lambda g_{m0} + R_{th}\frac{\partial I_{DS}}{\partial T}\left(I_{DS} + g_d V_d\right); \qquad (A-1)$$

$$\frac{\partial I_{DS}}{\partial T} = \frac{\frac{dI_{DS}}{dT}}{1 + \frac{dI_{DS}}{dT} R_{th} V_d};$$

$$g_m = \frac{\partial I_{DS}}{\partial V_g} + \frac{\partial I_{DS}}{\partial T}\frac{d\Delta T}{dV_g}; \frac{d\Delta T}{dV_g} = R_{th} V_d g_m$$

$$\therefore g_{m0} = \frac{g_m}{1 + \frac{dI_{DS}}{dT} R_{th} V_d}$$

$$g_d = \lambda \frac{g_m}{1 + \frac{dI_{DS}}{dT} R_{th} V_d} + R_{th} \frac{\frac{dI_{DS}}{dT}}{1 + \frac{dI_{DS}}{dT} R_{th} V_d}\left(I_{DS} + g_d V_d\right); g_{d0} = 0 \; at \; saturation$$

$$\therefore \frac{g_d}{g_m} = \lambda + R_{th}\frac{dI_{DS}}{dT}\frac{I_{DS}}{g_m}$$

As we see the g_d/g_m ratio of ref [4] is modified by the self-heating effect. Therefore in the case of deca-nanometer devices we can now take into account both the self-heating along with DIBL in the saturation regime.

Reference

[1] M Saitoh, Ken Uchida "Universal relationship between Low Field Mobility and High Field Carrier velocity in high K and SiO_2 gate dielectric MOSFETs" IEEE IEDM, 10.1109/IEDM.2006.346757, 2006

[2] K Rais, G Ghibaudo, F Balestra and M Dutoit "Study of velocity overshoot in deep submicron silicon MOSFETs from liquid helium up to room temperature" Journal de Physique IV (colloque C6 supplement au Journal de phsique III), vol 4 , p19 1994

[3] W Muller, I Eisele, "Velocity Saturation in Short Channel Field Effect Transistors" Solid State Communications, vol 34, p447 1980

[4] W Firky, G Ghibaudo and M Dutoit, "Temperature dependence of drain-induced barrier lowering in deep sub-micrometer MOSFETs" EL, vol 30, n11, p911, 1994

[5] K Rais, G Ghibaudo, F balestra, and M Dutoit, "Study of saturation velocity overshoot in deep submicron silicon MOSFETs from liquid helium up to room temperature" Journal de Physique IV C6 suppl jour de Phys III, vol 4, p19, 1994

[6] G G Shahidi, D A Antoninadis, H I Smith "Electron Velocity Overshoot at Roon and Liquid Nitrogen Temperatures in Silicon Inversion Layers" IEEE EDL, vol9, n2, p94, 1988

[7] J Jomaah, G Ghibaudo, F Balestra, "Analysis and Modelling of self-heating effects in thin-film SOI MOSFETs as a function of temperature" Solid State Electronics vol 38 n3 p615 1995

[8] C Fenouillet-Beranger et al, "Requirements for ultra thin film devices and new materials for the CMOS roadmap" Solid-State Electronics, vol48, p961, 2004

[9] J P Collinge "Multi-gate SOI MOSFETs" Microelectronic Engineering, vol84, p2071, 2007

[10] Zhi-Hong Liu et.al "Threshold Voltage Model for Deep-Submicrometer MOSFET's" IEEE TED vol40 n1 p86 1993

[11] Ming-C Cheng, FeixiaYu, "SOI Thermal Resistance and its application to thermal modeling of SOI MOSFETs" Silicon-on-Insulator Technology and Devices XI Electrochemical Society Proceedings vol 2003-05 p395

[12] A Raychaudhuri, J Kolk, M J Deen, and M I H King " A Simple method to Extract the Assymetry in Parasitic Source and Drain Resistances from Measurements on a MOS Transistor" IEEE TED vol42 n7 p1388, 1995

[13] R W Coen and R S Muller "Velocity of Surface Carriers in Inversion Layers on Silicon" Solid State Electronics vol23 p35 1980

[14] A Modelli and S Manzini "High Field Drift Velocity of Electrons in Si Inversion Layers" Solid State Electronics vol 31 n1 p99 1988

[15] C G Sodini, P K Ko and J L Moll "The Effect of High Fields on MOS Device and Circuit Performances" IEEE TED vol 31, n10, p1386, 1984

[16] K Hess, "Magneto Resistance of n-Silicon Inversion Layers in the Ohmic and in the Hot Electron Range" Surface Science, vol 58, p235, 1976

6 Results and Discussions

So far from chapter 2-5, we went through linear regime and saturation regime MOSFET operations and extraction methods. In this chapter we discuss advanced MOSFETs devices and extraction results in linear as well as saturation regime. Devices studied here are FDSON Bulk+, FinFet and GAA technologies. FDSON device is studied in both linear and saturation regime under different temperatures. FinFet and GAA are studied in linear regime under different temperatures. The technique using gain factor β vs Temperature discussed in section 4.4 is used to obtain insight into the scattering mechanism.

INTRODUCTION

As per ITRS roadmap, continuous device scaling would bring gate lengths down to 22nm around 2015. But there are several difficulties in using planar bulk/SOI MOSFETs for sub 32nm regime due to its issues related to I_{off}, gate oxide scalability, DIBL and sub threshold slope. Meanwhile the performance limitations of MOSFETs in terms of parasitic and drive current I_{on} lies on the other side of the spectrum of issues that has to be addressed. In the case of planar bulk MOSFETs, the issues related to I_{off}, SCE etc were addressed by pocket implants, ultra shallow S/D, and LDD and worked well down to deep sub-micron technology nodes. The mobility degradation caused by these modifications is addressed by applying strain, by using stressors, Ge implants, SiGe etc. Gate oxide is changed to High K/Metal gate to achieve 1nm range equivalent oxide thickness of SiO_2 while keeping the physical thickness of gate oxide to ~ 3nm and thereby tackling the problem of excessive gate leakage and allowing further gate length scaling below 100nm. FD SOI devices are considered as an alternative for the bulk devices owing to advantages of using thin silicon on SCE and sub-threshold and avoid pocket implants thereby improving mobility. Also SOI gives proper isolation among the devices and thereby avoid the latch up problems (as well as advantageous in RF and high voltage applications). The presence of BOX means low substrate leakage currents and parasitic capacitance. However sub 32 nm node needs even further improvements. These are partially addressed in devices such as UT2B or UTB^2 [21, 22] (ultra-thin body and Box) in the case of FDSOI and FDSON. Thin silicon film multiple gate devices [20] like Double Gate, Triple gate, Pi Gate, Gate All Around GAA, FinFETs etc, that has more than one gate to control the channel region there by immunizing devices against SCE are alternative structures for sub 32nm nodes. Surrounding gate nano-wire is another solution in terms of low leakage and SCE/DIBL immunity [23].

We discussed linear regime and saturation regime MOSFET operation and extraction methods in the previous chapters. We also saw some of the extraction results on FDSOI and Bulk devices. This chapter discusses extraction results on advanced devices namely FinFets, GAA as well as Bulk+ FDSON. The extraction is mainly done in linear regime of the operation at different temperatures. FDSON MOSFETs are characterized in both linear and saturation regimes.

6.1 FDSON –Extraction results

A FDSON device has the advantages of both FDSOI (near ideal sub threshold slope, no floating body effects, etc.) and bulk MOSFET (low series resistance, self-heating) along with compatibility to conventional CMOS process (with a few additional steps). In addition, ultra-thin body silicon channel and buried oxides (UTB²) can be obtained with a thickness control of up to an atomic layer [1-4 and ref there in]. This provides superior resistance to short channel effects (SCE), excellent threshold voltage roll off control and small DIBL, along with high drive capability compared to bulk [4]. Therefore FDSON technology is well suited for gate lengths in deca-nanometer range. The planar Bulk+ [2, 3] technology is based on FDSON technology and features additional sidewall conduction channels, like a tri-gate device (Fig. 6.1) and uses a high-K/metal gate gate-stack to meet the challenges of further aggressive scaling (Narrow channel effect control (using side wall conduction), UTB² requirement (using FDSON), small gate oxide thickness (using High K/metal) etc). The device is shown in fig 6.1. Fig 6.1 (left) shows SEM image of cross section along the channel length and width showing the thin BOX and Si channel and the side walls with source/drain epitaxially grown from the bulk substrate. Fig 6.1 (right) shows the schematic view of the device. Halo implants and ground plane below the BOX are provided for better electrostatic control for DIBL/SCE immunity.

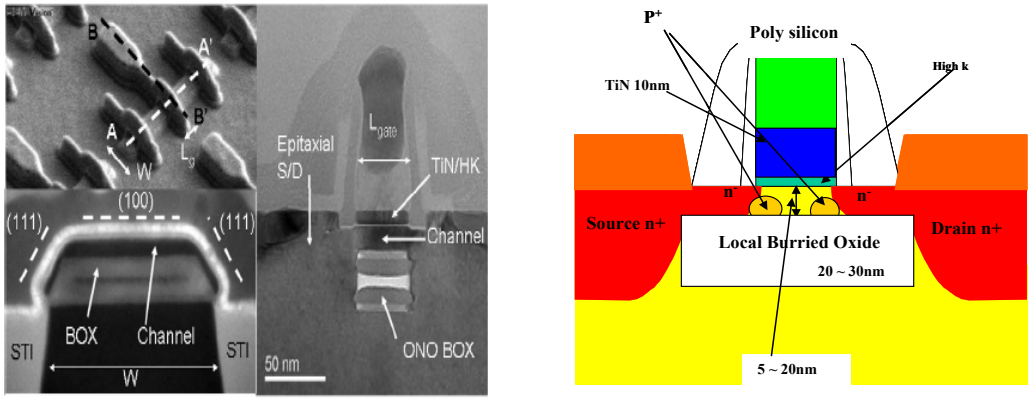


Figure 6.1

FDSON – Bulk+ structure. The SEM Image on left is taken from Ref [2]

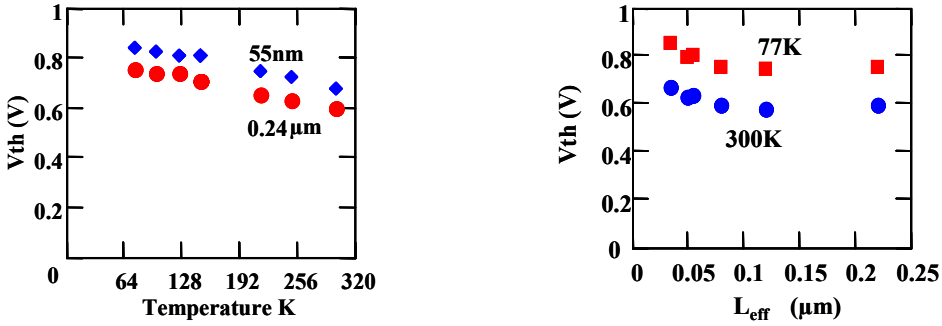


Figure 6.2

Left: Temperature dependence of V_{th} showing expected trend. Right: gate length dependence of V_{th} showing reverse short channel effect (RSCE) due to halo implants.

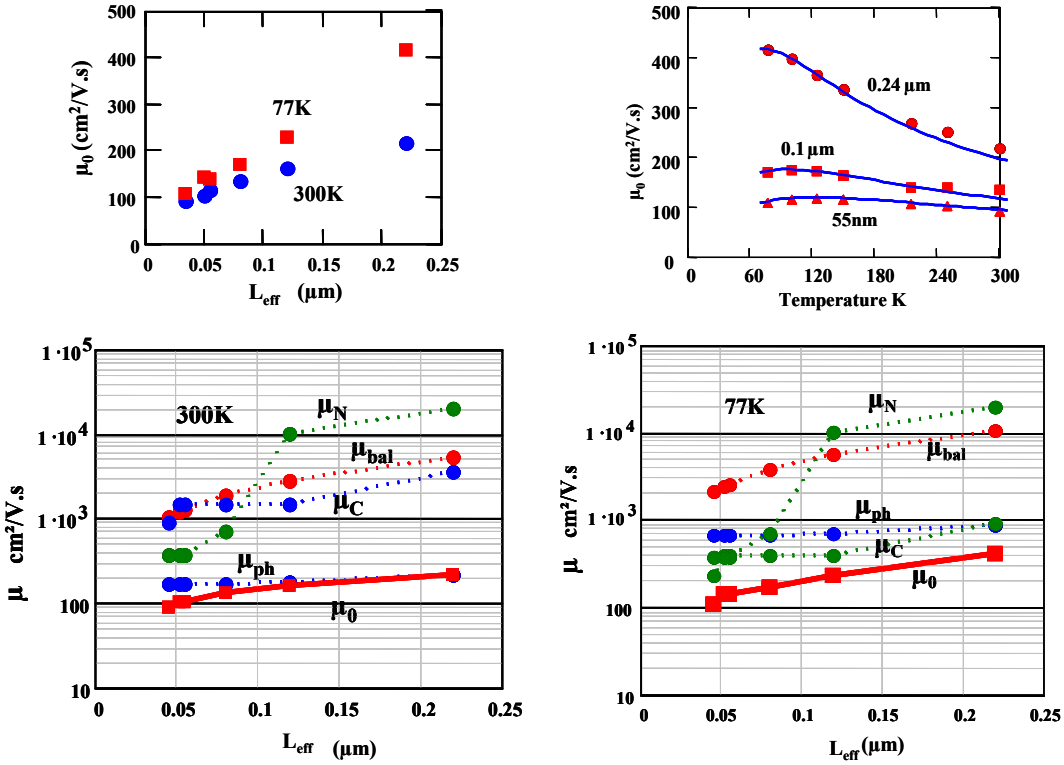


Figure 6.3

Top left: extracted μ_0 vs L_{eff} showing μ_0 roll off

Top right: model fit (lines) to μ_0 extracted values (symbols) vs Temperature for different channel lengths.

Bottom: Gate length dependence of μ_0 and of extracted μ_C, μ_{bal}, μ_N and μ_{ph} contributions at 300K and 77K.

The devices studied have HfO_2/HfSiON high-K / TiN (10nm) metal-gate gate stack, 15nm silicon thickness and 25nm buried oxide (Bulk+ technology). The equivalent gate-oxide thickness (EOT) is 1.84 nm. The channel length ranges from 55 nm to 0.24 µm with large width W = 1 µm. Sidewall conduction could thus be neglected. Both linear region and saturation region parameters were studied at low temperature.

The terms μ_0, R_{sd}, and V_{th} are extracted in the linear regime of the MOSFET using I_{DS} vs V_g model discussed in section 4 and Y function method for simplicity. The effective channel length and EOT are measured using split CV measurement. The parameter μ_0 and V_t are extracted against channel length and Temperature. The result is shown in the fig 6.2 and fig 6.3. V_t vs channel length shows reverse short channel effect RSCE showing the presence of halo implants.

We can also include the term μ_{bal} to take into account the quasi ballistic transport for deca-nanometer MOSFETs. The temperature model discussed in section 4.4 Eq 4.39 is used to extract the individual mobility terms. The value of phonon and neutral defects are fitted with the values already available in the literature [8].The mobility μ_0 vs L_{eff} follows typical roll off for short devices (Fig. 6.3 top left). Fig 6.3 (top right) shows mobility vs Temperature and its model fit. The fit is shown as lines and measured values are shown as symbols. The resulting mobility components and μ_0 (thick line) are plotted against L_{eff} for 300K and 77K in Fig 6.3 (bottom). μ_C roll-off was also observed for short channel lengths. This is possibly due to the effect of halo implants or remote Coulomb scattering with ionized impurities in the S/D. However, the dominant factor for channel length <100nm was found to be neutral defects. Thus, the μ_0 roll off at short channel length (Fig. 6.3 top left) is mostly due to neutral defects near S/D, with halo/remote Coulomb playing some role. At 300K μ_0 is dominated by phonon scattering and therefore shows a smaller roll off with channel length compared to 77K where neutral defects and Coulomb scattering has significant contribution. This makes the mobility more sensitive to channel length for 77K than 300K as seen in the fig 6.3.

DIBL was extracted from $I_{DS} - V_d$ measurement using g_d/g_m ratio for a given V_g (1.5V) against V_d. The plateau at high V_d gives the DIBL value λ [9] (section 5). The value of DIBL can be less than 80mV/V for L = 40nm for a well-optimized Bulk+/FDSON device by ground plane/bulk punch through implantation below the BOX [2, 3]. The relative DIBL (relative to long channel device 0.24µm) is shown in fig 6.4. It did not show any temperature dependence and featured expected behavior against L_{eff}. The value of DIBL for 55nm was found to be 150mV/V, which is much higher than an optimized FDSON/Bulk+ device [2, 3]. This shows that these devices used for measurements are not optimized for DIBL.

As explained in the section 5 for the extraction of v_{sats}, $I_{DS} - V_d$ is corrected first for DIBL [9]. The drift velocity and electric field is extracted using method [10] discussed in section 5. Fig 6.5 (left) shows v_{sat} vs L_{eff} for 77K and 300K. For 77K, we observe an overshoot like behavior in v_{sat} at short channel length, similar to ref [10]. At room temperature, v_{sat} does not show strong overshoot behavior. Fig 6.5 (right) shows the extracted v_{sat} vs temperature for different channel lengths. Fig 6.6 shows the simulated ballistic injection velocity using the Natori model [13] for 55nm device plotted against the temperature along with the measured value. The threshold voltage and series resistance used for

simulation are measured values. The negative slope and smaller values of v_{sat} vs T in fig 6.5 (right) and comparison shown in fig 6.6 indicates that the high-field velocity saturation effect better models the transport even for 55nm gate length. The characterization of the FDSON MOSFETs down to 55nm showed that the transport is still governed by the scattering. The on current is limited by the high field velocity saturation effect rather than ballistic transport.

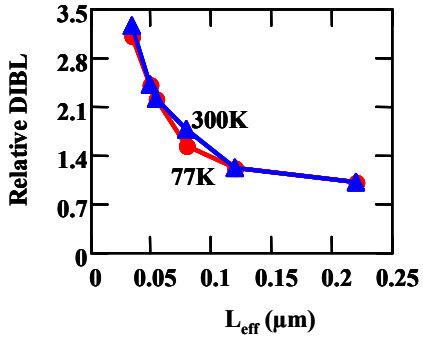


Figure 6.4

Relative DIBL Vs L_{eff} at 300K and 77K, extracted from the plateau of gd/gm.

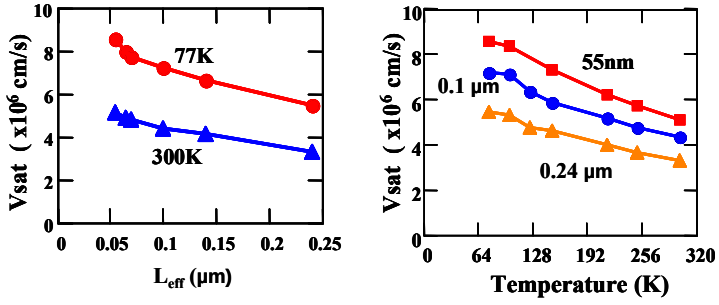


Figure 6.5

Gate length dependence (left) and temperature dependence (right) of the saturation velocity v_{sat}.

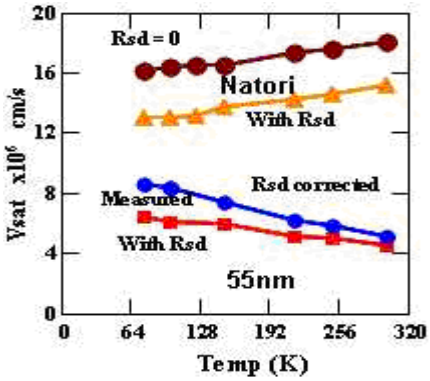


Figure 6.6

Shows ballistic injection velocity v_{bal}. calculated using Natori model [13] for 55nm device against measured value of v_{sat}

6.2 Extraction on FinFet

Multiple gate devices are one of the projected candidates for future technology nodes of MOSFETs beyond 32nm. The gate oxide scaling has its limits, even with High K/ Metal, for technology nodes below 32nm and additional enhancements for the electrostatic channel control is needed to further scale the channel lengths. This is successfully achieved in ultra-thin silicon (<10nm) SOI/UT2B, multi gate structures or surrounded gate nanowires etc.

This section is focused on FinFET structure that has gate surrounding the channel on three sides as shown in fig 6.7. The silicon channel is thin and is of the order of 10nm and less. FinFETs used here are obtained from IMEC and fabrication of the same is given in ref [18], which has reported Si channel thickness down to 4nm. This gives a good SCE/DIBL immunity for these devices.

The impact of sidewall roughness on these devices is given in [17] and it is found that the side wall conduction is degraded by excess sidewall roughness at high fields. This is important because in thin silicon FinFETs, conduction channel is formed on the two sidewall surfaces. Therefore on current is limited by the surface roughness. Also in the case of ultra-thin Si, surface roughness would manifest as thickness fluctuation similar to ultra-thin FDSOI case (T_{si} ~ 5nm and less) discussed in section 2. The effective mass in the side wall which is (110) is also larger reducing the mobility. DIBL reported in [18] is as low as 30mV/V while sub-threshold swing is ~ 70mV/decade for channel length 30nm for nMOS devices. This shows superior SCE control in these devices as expected from these structures.

Device under study has fin height = 65nm, channel length = 35nm and fin width = 10nm. There are 5 parallel fins in one MOSFET and it is assumed that all the fins are uniform. However the dimensions are too small to measure using the CV technique. Low temperature study using the temperature model of linear regime gain factor β discussed in section 4.4 is done for temperature range 100K to 380K for V_d of 5mV. Result is shown in fig 6.8.

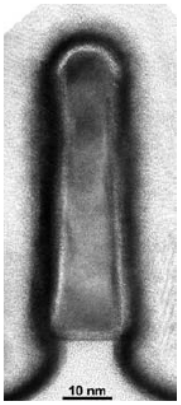


Figure 6.7

Figure taken from ref [18]. The cross section of one of the fins of FinFET is shown with fin height 60nm and width 10nm/ EOT given in [18] ~ 1.9nm for TiN/HfO_2. The oxide thickness is more or less uniform around the structure.

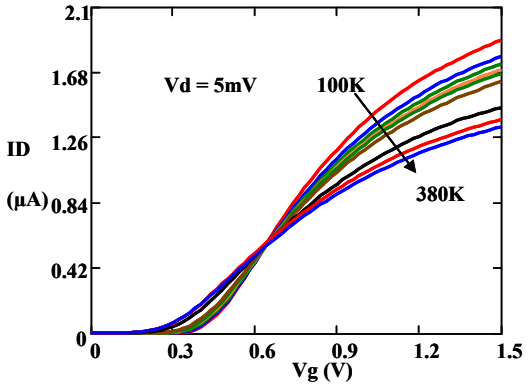


Figure 6.8a

I_{DS}-Vg for different temperatures

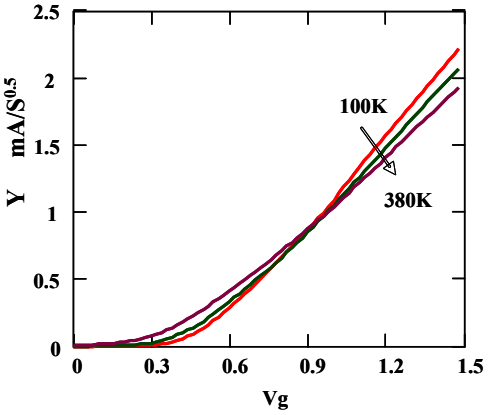


Figure 6.8b

Corresponding Y-Vg for different temperatures

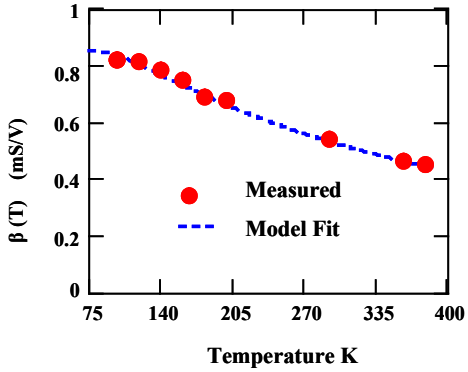


Figure 6.8c

Extracted β vs Temperature and the model fit using the best fitted values

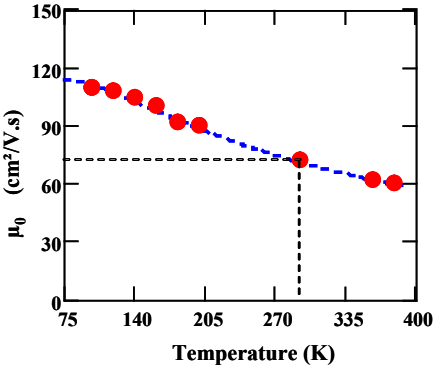


Figure 6.8d

Extracted μ_0 vs Temperature from the model fit is shown

Figure 6.8

Extraction results using β vs Temp model for FinFet 35nm channel length,

Fin width = 10nm, Fin Height = 65nm is shown

Figure 6.8a is the $I_{DS} - V_g$ for V_d = 5mV for different temperatures from 100K to 380K. Fig 6.8b shows the Y function. In this extraction, Y function method is employed as we only have one device. To first order, Y function avoids the impact of series resistance and we can directly measure gain factor β and threshold voltage. Also since device is too small, we cannot measure the capacitance and so we cannot use the split CV. MR on a 10μm wide FinFET is reported in [24], however this measurement is for top surface similar to a planar MOSFET. For the device under consideration, this is not the case and MR cannot be applied. Therefore extraction using Eq 4.40 of section 4 using gain factor obtained from Y function is used for the low temperature characterization. The effective mass for the side wall is $0.283m_0$ (α = 0.283 for (110) surface side wall) using Eq 4.42 of section 4. The model in Eq 4.40 is slightly modified here for the fitting purpose as shown in Eq 6.1. The fig 6.8c shows β vs Temperature plot and the fit.

$$\beta \equiv MF \times \mu_0$$
$$\frac{1}{\mu_0} = \frac{300}{\mu_{C0} T} + \frac{T}{300 \mu_{Ph0}} + \frac{1}{\mu_{const}} + \frac{T^{0.5}}{\mu_{b0}} \qquad (6.1)$$

The fit was made by taking t_{ox} = 1.9nm as given in ref [18] fig 6.7. The phonon component μ_{ph0} is calculated by taking μ_{ph0} given for planar MOSFET (100) of 170 from [19] and using effective mass for FinFET side wall surface (110)/[110]. $\alpha_{(110)}$ = 0.283 and on (100) surface $\alpha_{(100)}$ = 0.19. The ratio $\alpha_{(110)}/\alpha_{(100)} \approx 1.5$. This gives $\mu_{ph0\,(110)} = \mu_{ph0\,[19]}/1.5 \approx 116$ cm^2.V^{-1}.s^{-1}. Therefore a rounded off fit value for μ_{ph0} 110 cm^2/V.s can be justifiably used. Finally we have fit values for the rest of the terms

MF = C_{ox} x W/L = 1.817 10^{-6} x 4.1
μ_{C0} = 1400 cm^2/V.s, μ_{const} =350 cm^2/V.s
μ_{b0} = 1.08 10^4 cm^2 .K$^{0.5}$/V.s, L_{eff} = 32nm

The value of μ_{ph0} =110 cm^2.V^{-1}.s^{-1} can also be justified from β – Temperature plot which shows that temperature dependent mechanism is dominant in the low field mobility. The total width obtained from MF, C_{ox} and L_{eff} (that fits also to ballistic term) is 131nm which is around 2*fin height. Therefore these fit values are most acceptable values. All the values are round off because we always have an uncertainty in this type of fitting. Fig 6.8d shows the plot of μ_0 vs Temperature, μ_0 calculated from β and the value of MF obtained after fitting. The value of μ_0 at 300K is ~ 70cm^2/V.s. The mobility component μ_{ph0} taken from [19] was a good starting point for the fitting procedure.

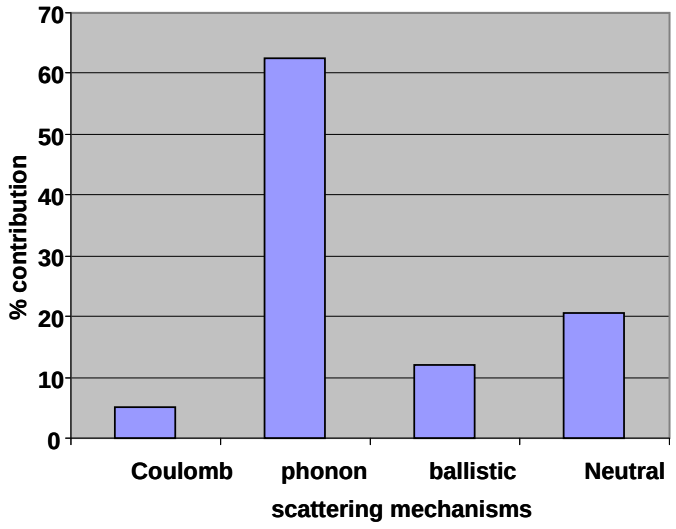


Figure 6.9

Extracted relative contribution of scattering mechanisms and ballistic transport on low field mobility at 300K

Figure shows that phonon scattering is dominating the transport.

Figure 6.9 shows relative contribution of scattering mechanisms governing the transport as per the fit. The figure shows a high level of phonon scattering along with a constant factor μ_{const} which normally is attributed to neutral defect mobility component. The ballistic component μ_{bal} (= $T^{-0.5}$ x μ_{b0}) at 300K is ~ 600cm^2/V.s. The Coulomb component was found to be negligible. The dominant scattering mechanism at room temperature as per the fit obtained is therefore phonon scattering. Previous studies have shown that the FinFET sidewall also has increased SR compared to the top surface [17]. Therefore at high fields current is mainly limited by phonon and SR scattering.

6.3 Extraction on planar GAA

Gate all around devices is another alternative structure of MOSFET developed to tackle the scaling problems in sub 32nm technology nodes. With gate surrounding a thin Si channel we have superior channel control and therefore low leakage, ideal sub threshold swing, low DIBL and good mobility using un-doped channel. Planar GAA is similar to DG with wide top and bottom surfaces relative to the thickness but has a surrounding gate that allows for even superior gate control of the channel. See fig 6.10 for the TEM cross section of the device taken from [15]. Due to effective > 2 channels (depends on the silicon channel thickness T_{si}) per device we get more than twice the drain current per mask width, compared to a planar MOSFET.

GAA devices studied here are fabricated in ST Microelectronics and are based on FDSON process discussed in section 6.1. This allows a good control over channel Si thickness as they are epitaxially grown. Fabrication process is explained in [15, 16]. The device is built on SOI substrate but the transistor rests on the top of Si well as shown in the TEM cross section image. Si well is heavily doped so that parasitic transistor formed below the bottom gate is suppressed (in order to avoid leakage currents). The device has High K/ Metal gate, with HfO_2 = 2.5nm and TiN 10nm. The smallest device channel thickness T_{si} reported in [15] is 4nm which gives excellent sub-threshold and DIBL characteristics. The device dimension in fig 6.10 from [15] is WxL 47nm x 40nm.

As expected from a GAA structure we have excellent sub-threshold swing of 70mV/decade. DIBL of these devices is very small compared to a FDSOI counterpart and is around 30mV/V for 40nm channel length reported in [15]. V_t with channel length [15] showed no roll off and along with the low DIBL this shows good immunity of these devices to SCE. $I_{on} - I_{off}$ characteristics reported in [15] is one of the best among the state of the art devices. I_{on} reported is 1.6mA/μm (mask or foot print width) with I_{off} of 0.8nA/μm given for nMOS devices.

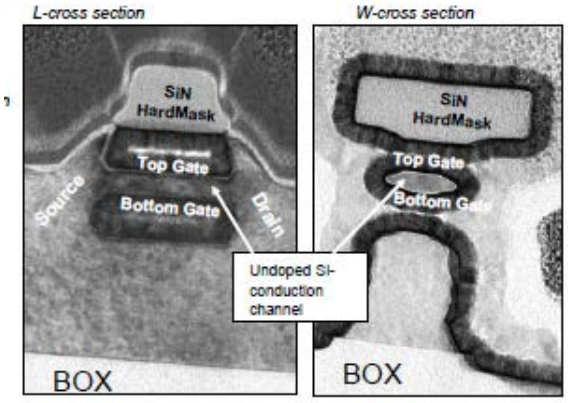


Figure 6.10

TEM cross section of the planar GAA is shown, taken from ref [15]. The device shown here have L = 40nm and W_{mask}= 47nm [15]. T_{si} ~ 5nm. HfO_2 ~ 2.5nm. The device is built on SOI wafer, however we have Si below the bottom gate, above the box. This could create a parasitic transistor. But it is suppressed by heavy well implantation. Normally bottom parasitic transistor therefore will have high threshold voltage ~ 1V [15].

The device subjected for low temperature measurement for transport study has mask dimension WxL 0.11µm x 0.05µm. However we cannot use these values for the extraction as the uncertainty induced by the process is too large. Therefore we use the same extraction method that was used for FinFET in the previous section by analyzing the gain factor β which is a measurable quantity. Silicon channel thickness is around 10nm. Fig 6.11 below gives the temperature measurement results. We use Y-function technique for the gain factor β. We cannot measure the capacitance as devices are too small and therefore split CV measurement is not possible and so is the case of determination of EOT. MR measurement cannot be used in this device either. Therefore the method specified in section 4.4 using β – Temperature is most suitable method for this scenario.

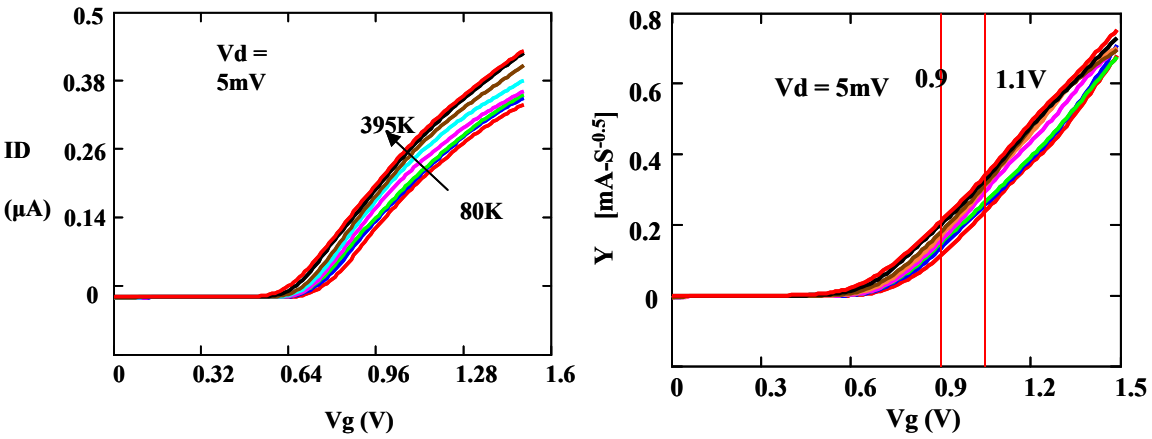


Figure 6.11

Left: Shows measured $I_{DS} - V_g$ for temperatures 80K to 395K. The curve shows typical V_t shift with temperature. Right: Y function for different temperatures is shown. The range of gate voltage from which β is extracted is also shown. This range is chosen to avoid any parasitic effects and high field effects. Negative bias is applied to substrate to suppress parasitic channel.

Figure 6.11 shows the $I_{DS} - V_g$ and Y function for the device measured in the linear regime with V_d = 5mV. The range of gate voltage in Y function for extracting β is chosen such that it is not too close to V_t and no too high where we see effect of high fields. Also there is a lot of measurement noise at high gate voltages as seen in the Y function in fig 6.11 (right). The measurement is made in the temperature range 80K to 395K. β – Temperature model discussed in the section 4.4 is then used to fit the data. This is shown in the fig 6.12. β is extracted for two V_d values 3mV and 5mV here. Both are shown in the figure and the mean value of the two is also shown. The mean value is then used for the fitting. The value for phonon is used from the literature [19] 170 $cm^2/V.s$ @ 300K. β is rewritten as MF x μ_0 as shown in Eq 6.1.

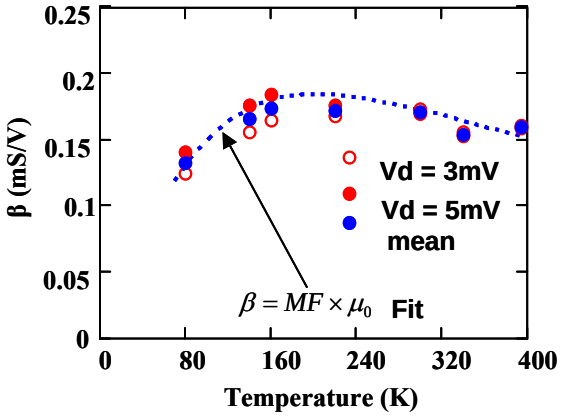


Figure 6.12

Gain factor β (mS/V) vs Temperature (K). Both Vd = 5mV, 3mV is shown and mean values is fitted with the model.

MF = W C_{ox}/L = 1.72 10^{-6},

μ_{C0} = 350 $cm^2/V.s$

μ_{const} = 1500 $cm^2/V.s$,

μ_{b0} = 30000 $cm^2.K^{0.5}/V.s$,

L ~ 75nm taking (100) surface for top and bottom surfaces assuming the device as DG MOSFET (as thickness is small). The channel length obtained is in the range of expected value after the process. Taking a nominal values for EOT = 2nm (HfO_2 = 2.5nm and SiO_2 layer is not known), W ~ 75nm. Therefore the value of mobility μ_0 at room temperature is ~ 100 $cm^2/V.s$. μ_{const} is the mobility component that is insensitive to temperature and is normally attributed to neutral defects in the case of planar MOSFETs due to ion implantation induced defects near S/D [19].

The plot β vs Temperature is a measured quantity and temperature trend observed reveals the presence of Coulomb scattering at low temperature, phonon scattering is dominant at high temperature. The ballistic transport component is very large and so insignificant in these devices. The relative

contribution of different scattering mechanisms and ballistic transport on the mobility is shown in figure 6.13.

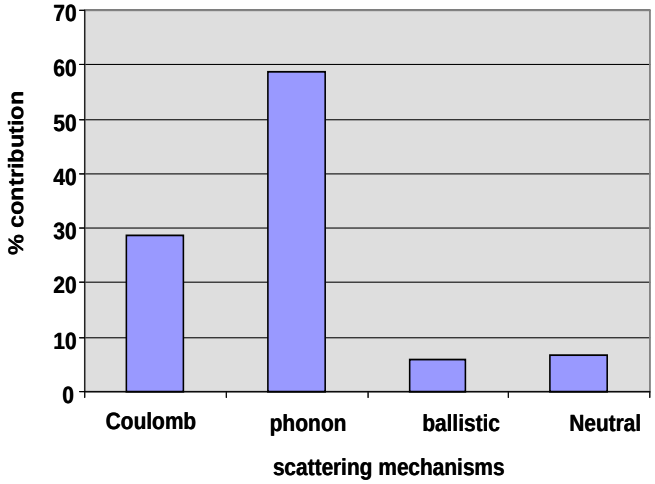


Figure 6.13

Relative contribution of scattering mechanisms and ballistic transport on low field mobility at 300K is shown. Figure shows phonon scattering is dominating the transport while Coulomb scattering also has a significant contribution

The planar GAA device is characterized in linear regime for different temperatures and we can see that devices are well within the scattering dominated transport regime. The dominant scattering mechanism seen is Coulomb and phonon scattering in the room temperature.

Conclusion

In this chapter we have discussed some advanced MOSFET devices and extraction results on the same in linear as well as saturation regime. Devices studied are FDSON Bulk+, FinFet and planar GAA. FDSON device is studied in both linear and saturation regime under different temperatures. FinFet and GAA are studied in linear regime and since their physical dimensions were unknown, the technique using gain factor β vs Temperature discussed in section 4.4 is used to obtain a qualitative insight into the scattering mechanisms. The measurement showed that transport in these devices is still governed by scattering mechanisms. In the case of FinFET device, the scattering is dominated by phonons and surface roughness. GAA devices showed a high level of Coulomb scattering while phonon scattering dominates at room temperature. FDSON devices showed a high level of neutral defects in 55nm devices while phonon scattering playing a dominating role at room temperature. Therefore in all the three advanced structures, transport at room temperature is still dominated by scattering (phonon) even down to 35nm channel length (side wall of FinFET).

Reference

[1] S.Monfray et.al "Emerging silicon-on-nothing (SON) devices technology", Solid State Elec, vol.48 n6, p.887 2004.

[2] G.Bidal et al "Planar Bulk^{+} Technology using TiN/Hf-based gate stack for Low Power Applications ", VLSI Tech. Digest, p146-147 2008

[3] F.Boeuf et al "Optimization of Bulk^{+}/SON Integration for Low Stand-by Power (LstP) Applications ", Proc. SSDM (Ext Abs), p.1030 2009

[4] J. Pretet, S Monfray, S.Cristoloveanu, T.Skotnicki, "Silicon on Nothing MOSFETs: Performance, Short-Channel Effects, and Backgate Coupling", IEEE TED, vol 51 n2, p.240 2004

[5] G.Ghibaudo, Elec Letters, vol 24 n9 , p543 1988

[6] K.Huet et al "Monte Carlo Study of Apparent Mobility Reduction in Nano-MOSFETs", Proc. ESSDERC, p382-385 2007

[7] M. S Shur, "Low Ballistic Mobility in Submicron HEMTs", IEEE EDL, vol. 23 n9 p511 2002

[8] A.Cros et al "Unexpected Mobility degradation for very short devices: A new challenge for CMOS Scaling" IEDM Tech. Digest, DOI:10.1109/IEDM. 2006.346872 2006

[9] W.Firky,G.Ghibaudo, M.Dutoit "Temperature dependence of drain-induced barrier lowering in deep submicrometre MOSFETs" EL Vol30n11p:911 1994

[10] K. Rais et al, "Study of saturation velocity overshoot in deep submicron silicon MOSFETs from liquid helium up to room temperature" Journ. Physique IV, vol. 4, C6-19 1994

[11] D.Fleury, A.Cros, G.Bidal, J.Rosa, G.Ghibaudo "A New technique to Exact the Source/Drain Series Resistance of MOSFETs" IEEE EDL vol.30 n9, p975 2009

[12] J.Kooman, "Investigation of the MOST Channel Conductance in the Weak Inversion" Solid Stat Elec, vol. 16, p.801-810, 1973

[13] K.Natori, "Ballistic metal-oxide-semiconductor field effect transistor" J.Appl.phys, Vol 76, n8, p4879 1994

[14] Tong-Chern Ong, Ping K Ko, and Chenming Hu, "50A° Gate-Oxide MOSFET's at 77K " IEEE TED, vol34, n10, p2129 1987

[15] J.-L. Huguenin et al, "Ultra-Thin (4nm) Gate-All-Around CMOS devices with High-k/Metal for Low Power Multimedia Applications" SSDM 2010

[16] G Bidal et.al, "Gate-All-Around technology: taking advantage of ballistic transport" IEEE ESSDERC, p315, 2009

[17] J W Lee et.al, "Experimental Analysis of Surface Roughness Scattering in FinFET devices" IEEE ESSDERC, p305, 2010

[18] N Collaert et.al, "Tall tripple gate devices with TiN/HfO_2 gate stack " VLSI Symp tech dig, p108, 2005

[19] A.Cros et al. "Unexpected Mobility degradation for very short devices: A new challenge for CMOS scaling", IEDM Tech. Digest, DOI:10.1109/IEDM. 2006.346872 2006

[20] J P Colinge "Multi-gate SOI MOSFETs" Microelectronic Engineering, vol 84, p2071, 2007

[21] F Andrieu et al “Low Leakage and Low Variability Ultra-Thin Body and Burried Oxide UT2B SOI technology for 20nm Low Power CMOS and Beyond” IEEE Symposium on VLSI Technology Digest of Technical Papers, p57, 2010

[22] C Fenouillet-Beranger et al “Requirements for ultra-thin-film devices and new materials for the CMOS roadmaps” Solid State Electronics, vol48 p961 2004

[23] K Tachi et al “Ccomparison of low-temperature electrical charcteristics of gate-all-around nanowire FETS, Fin FETs and fully-depleted SOI FETs ” Microelectronics Reliability vol51 p 885 2011

[24] Carolina D G dos Santos et al, “Magneto Resistance Technique for Mobility extraction in Triple gate FinFETs at low Temperature” EUROSOI, p123, 2010.

[25] K Bennamane et al “DC and low frequency noise characterization of FinFET devices” Solid-State Electronics, vol53, p1263, 2009

7 Conclusion and future prospective

Electron transport is one of the key properties that need to be improved in order to sustain performance improvement for each technological node. Electron transport in MOSFETs in nm scales is governed by out of equilibrium, or even ballistic, phenomena. These various physical phenomena determine the intrinsic MOSFET performance limitations in terms of its drive current and therefore its speed. Therefore along with the advancement in the technology nodes, it becomes necessary to evolve the transport models and parameters to better explain the MOSFET operation. Various parameter extraction methods also have to be evolved to take into account the new model as well as to facilitate accurate parameter extraction for existing models by avoiding various parasitic effects whose impact becomes more and more significant as technology node advances. Extraction of parasitic components like series resistance as well as fringing capacitance becomes important for the short channel MOSFETs.

During the course of this thesis, existing transport models and extraction methods in the linear regime and saturation regime were revisited. Chapter 2 discussed about the concept of mobility in the linear regime of operation. Different scattering mechanisms were discussed and mobility models for each of the scattering mechanisms were studied. Finally ballistic and quasi ballistic models were also discussed for sub 32nm nodes. The models were then used in the extraction in the linear regime in chapter 4 in order to separate the different scattering mechanisms. Chapter 3 discussed about the concept of saturation velocity and related field dependent mobility model for sub μm MOSFET devices operating in the saturation region. In deep sub μm MOSFETs, transport is governed by non-stationary effects and manifests as velocity overshoot. Ballistic and quasi ballistic model and kT layer concept is also discussed for the sake of sub 32nm nodes.

State of the art linear regime empirical models and extraction methods were discussed in the section 4 for the extraction of mobility. Measurement using Y function was modified to take into account the gate voltage dependent series resistance for the sake of short channel MOSFETs. Magnetoresistance measurement is a promising characterization method especially for short channel devices as it does not require any prior knowledge of physical dimensions. However magnetoresistance in the ballistic and quasi ballistic regime does not have an appropriate analytical model. Therefore a new magnetoresistance model was developed for ballistic transport as well as for quasi ballistic transport. A low field mobility vs temperature model was discussed, by making use of the mobility model of different scattering mechanisms as well as ballistic transport, which enables separation of scattering mechanisms in MOSFETS from low temperature characterization. This was extended to gain factor β

in order to apply the same in the short channel devices where it is difficult to measure the physical dimensions. The gain factor can be directly extracted from Y function, McLarty methods etc from $I_{DS} – V_g$ measurement at low V_d. With further improvements, this approach provides a fully experimental extraction method for future technology nodes.

Extraction methods using the velocity saturation models were discussed in chapter 5. Improvement to the electrical extraction in the saturation regime was done by taking into account both DIBL and self-heating effects, by making use of low temperature measurements, for the sake of deep sub μm MOSFETs. Magnetoresistance measurement was made for the first time in the saturation regime of MOSFET operation. It was found that we have magnetoresistance mobility even in the saturation regime. Various aspects of this mobility were studied for bulk MOSFETs from 1μm down to 50nm channel lengths. Advantage of using magnetoresistance measurement in this regime is that we don't need to rely on complex numerical model fitting that is needed in electrical characterization methods and its associated errors in the estimation due to parasitic effects. This makes it attractive for current as well as future short channel MOSFETs. The physics of magnetoresistance in this regime is not fully understood, but with improvements this opens up a new branch of characterization that allows a better understanding of transport in saturation region.

This thesis therefore has successfully covered some important aspects of characterization of electron transport for short channel MOSFETs in the linear and saturation regime. Improvements were suggested on the existing extraction methods in both linear and saturation. Improved extraction methods were successfully applied on real devices. The new linear regime magnetoresistance model for ballistic and quasi ballistic transport will help in extending the applicability of magnetoresistance characterization to future technology nodes. Further improvements are necessary in this model to take into account its short comings. Deeper understanding is necessary for the magnetoresistance observed in the saturation regime down to 50nm and this could be a good alternative for the existing electrical extraction methods in this regime.

Finally to conclude this thesis, low temperature characterization done on advanced MOSFETs like Bulk+ FDSON, GAA, FinFETs were discussed in the chapter 6 and results so far obtained showed that transport is still governed by scattering even for 35nm FinFETs. Phonon scattering has a significant role in the electron transport at room temperature operation of MOSFETs down to deca-nanometer regime with additional scattering like neutral defects and Coulomb contributing in some cases.

8 Résumé en Français

Caractérisation de transport des électrons dans les transistors MOS à canal court

8.1 Introduction

La qualité du transport électronique est l'une des clés permettant de soutenir la progression des performances pour les futures générations de composants. De très nombreux facteurs, comme le choix de l'isolant et du métal de grille, le matériau de canal ou la présence de contraintes mécaniques, affectent de façon négative ou positive ces propriétés de transport. L'épaisseur du canal, qui atteint des dimensions nanométriques joue également un rôle par le biais des interactions associées à la présence des interfaces ou des fluctuations d'épaisseurs, ou des effets de couplage électrostatique ou quantique entre ces interfaces. Il est probable que des mécanismes d'interaction associés à la proximité des zones surdopées de source et de drain puissent également intervenir. A ces dimensions, on s'attend à observer des phénomènes de transport hors d'équilibre, voire balistique, qui peuvent remettre en question la validité des paramètres utilisés pour caractériser le transport. Avec l'avancement de la technologie, il devient donc nécessaire de faire évoluer les modèles de transport et le choix des paramètres qui le décrivent afin de mieux expliquer le fonctionnement du MOSFET. Cette thèse se concentre sur la compréhension des modèles de transport existants et des méthodes d'extraction pour les nœuds technologiques actuels et futurs. Les modèles de transport et les méthodes d'extraction de paramètres en régime linéaire et de saturation ont été explorés au cours de cette thèse. L'impact de la résistance série, qui est une fonction de la tension de grille, dans les MOSFET avancés est pris en compte et une nouvelle méthode d'extraction améliorée a été développée dans le régime linéaire. Des mesures à basse température ont été utilisées en régime linéaire pour l'extraction des mécanismes de diffusion en utilisant le modèle de mobilité. Une nouvelle méthode de correction pour le courant de drain dans le régime de saturation pour les MOSFET canal court est développée en utilisant les mesures à basse température. Cela permet de corriger du DIBL ainsi que des effets de « self heating ». Le modèle de saturation de vitesse et la méthode d'extraction associée sont explorés dans le régime de saturation et sont étudiés en fonction de la température et de la longueur de canal. Mesurer la magnétorésistance offre des perspectives prometteuses pour les dispositifs à canal court et permettant d'extraire directement la mobilité, sans la nécessité de la connaissance des dimensions du canal. Un modèle analytique pour la magnétorésistance est développé pour les modèles de transport balistique et quasi-balistique. La mesure de la magnétorésistance est explorée dans la région de saturation pour la première fois jusqu'à 50 nm sur les MOSFET « bulk » afin de comprendre l'applicabilité de cette méthode d'extraction à ce régime. Enfin les dispositifs bulk+ FDSON, FinFET, et GAA sont caractérisés en fonction de la température et les mécanismes de transport dans ces nouveaux dispositifs sont étudiés jusqu'à 35 nm (FinFET). En outre, le paramètre de champ effectif η est extrait pour les dispositifs sSOI. On trouve qu'il est différent du cas « bulk » comme c'était le cas pour les résultats obtenues sur bulk contraint et FDSOI. Cela est interprété par la rugosité de surface et la diffusion des phonons en raison de l'occupation préférentielle de la sous bande fondamentale.

8.2 Transport électronique dans le transistor MOS – Régime linéaire

Dans le transistor MOS en régime linéaire (faible tension de drain), le transport électronique est classiquement modélisé par l'intermédiaire de la mobilité effective μ_{eff}. Cette mobilité constitue un indicateur de performance intéressant dans la mesure où elle est directement corrélée à la valeur du courant I_{on} du transistor (et donc à sa vitesse). En effet, en régime linéaire, le courant de drain s'exprime comme:

$$I_{DS} = \frac{W}{L} C_{ox} \mu_{eff}(V_g) \left(V_g - V_t - \frac{V_d}{2} \right) V_d \tag{2.1}$$

où W, L et C_{ox} sont respectivement la largeur de canal, la longueur de canal et la capacité d'oxyde. V_t est la tension de seuil tandis que V_d, V_g sont les tensions de drain et de grille. Même pour des longueurs de grille relativement faibles, il existe un corrélation entre I_{on} et la mobilité effective en régime linéaire μ_{eff} [1, 2]. De ce fait, les techniques qui permettent d'accroître la mobilité, comme l'ingénierie de la contrainte mécanique, constituent un moyen prometteur pour augmenter I_{on}.

La mobilité électronique utilisée dans l'équation 2.1 est différente de la mobilité dans le silicium massif. Ceci du fait que dans le transistor MOS, le transport se fait dans un canal d'inversion, sur une épaisseur de l'ordre de quelques nanomètres, le long d'une interface Si/ SiO_2. Aux interactions présentes en volume, telles que l'interaction avec les phonons et avec les charges Coulombiennes, s'ajoute donc une interaction avec la surface. Dans certaines conditions [10, 11] et en particulier dans la mesure où ces interactions peuvent être considérées comme indépendantes, la mobilité résultante peut être obtenue à partir des composantes associées à chacun de ces types d'interaction selon la règle de Mathiessen [12 and others]:

$$\frac{1}{\mu_{eff}} = \frac{1}{\mu_C} + \frac{1}{\mu_{ph}} + \frac{1}{\mu_{SR}} \tag{2.4}$$

où μ_C, μ_{ph}, et μ_{SR} représentent les mobilités associées à l'interaction Coulombienne avec les charges ponctuelles, à l'interaction avec les phonons et à l'interaction avec la rugosité d'interface (SR pour "surface roughness" en anglais). Même si elle n'est pas rigoureuse dans toutes les conditions, cette règle fournit une méthode analytique simple pour analyser la mobilité effective, identifier les interactions dominantes et séparer les contributions des différents types d'interaction à condition qu'elle soit couplée à une analyse approfondie en fonction de la température, dans une large plage de fonctionnement allant de faible à forte inversion.

La Figure 2.1 montre comment la mobilité effective varie en fonction du champ effectif E_{eff}., de la température et du dopage du canal [6, 7]. Rappelons que le champ E_{eff}, défini comme le champ transverse moyen dans le canal d'inversion, est fréquemment utilisé à la place de V_g-V_t ou de la charge d'inversion Q_{inv}, pour caractériser l'état d'inversion du canal dans des composants différents,

indépendamment de leurs caractéristiques précises, telles qu'épaisseur de diélectrique, capacité d'oxyde, ou dopage de canal. L'interaction avec la rugosité de surface est le mécanisme dominant en forte inversion, dans la zone à fort champ E_{eff} où toutes les courbes se rejoignent. Pour les faibles dopages, la mobilité à 300K est dominée par l'interaction avec les phonons et avec la rugosité de surface. Le dopage fait dévier μ_{eff} de cette courbe dite universelle du fait de la présence des interactions Coulombiennes avec les impuretés ionisées. Cette mobilité, qui est d'autant plus faible que le dopage est fort, est surtout présente en faible inversion, pour les faibles valeurs du champ E_{eff}. Notons en outre qu'à Q_{inv} donné, E_{eff} est d'autant plus élevé que le dopage est fort, ce qui participe également à la dégradation de mobilité. A basse température, l'interaction avec les phonons est réduite du fait du gel des phonons et la mobilité est dominée par les interactions Coulombiennes et avec la rugosité de surface. Les différentes composantes de la mobilité ont des comportements différents en fonction du champ effectif - ou de la charge d'inversion - et de la température [7, 8]. Pour analyser en détail la mobilité effective, on l'étudie donc en fonction de E_{eff} (Q_{inv}), la température, le dopage, la densité de charges dans l'oxyde et la densité de pièges d'interface, et on utilise l'équation 2.4 pour séparer les différentes composantes.

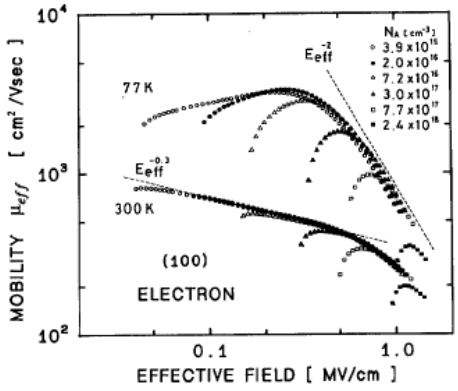


Figure 2.1

Mobilité effective en fonction du champ effectif à 300K et 77K, pour différents niveaux de dopage dans le canal, d'après Takagi et al [6].

Dans les transistors MOS à canal court, d'autres mécanismes d'interaction sont présents. Certains son liés à la proximité des contacts de source et drain et sont affectés par leur technologie de fabrication. C'est le cas par exemple de l'interaction avec les défauts neutres et chargés injectés pendant les recuits à partir des zones implantées de source et de drain [3, 4]. Ces mécanismes qui dégradent le transport à proximité des contacts provoquent une réduction de la mobilité effective globale pour le transistor au fur et à mesure de la réduction de la longueur de canal. Pour les transistors MOSFET FDSOI sur film silicium ultra mince, la mobilité est en outre dégradée par les fluctuations d'épaisseur. Enfin, particulièrement pour les longueurs de canal sub-32nm, le transport balistique lui-même se traduit par une réduction de la mobilité apparente définie par l'équation 2.1. Dans le régime intermédiaire, dit

quasi-balistique, où le transport dans le canal est non stationnaire, on peut malgré tout, ainsi que l'a proposé Shur [5], définir une mobilité apparente vérifiant:

$$\mu_{eff} = (1-r)\mu_{bal} = \left[\frac{1}{\mu_{bal}} + \frac{1}{\mu_{effdd}}\right]^{-1} ; \ \mu_{bal} = v_{th}\frac{q}{2kT}L_{eff} \quad (2.5)$$

Le terme r est un coefficient de rétro diffusion [9], v_{th} est la vitesse d'injection thermique, L_{eff} est la longueur de canal, k la constante de Boltzmann, T la température et q la charge de l'électron.

Références

[1] M Saitoh, Ken Uchida "Universal relationship between Low Field Mobility and High Field Carrier velocity in high K and SiO_2 gate dielectric MOSFETs" IEDM, 10.1109/IEDM.2006.346757, 2006

[2] A. Lochtefield, D. A. Antoniadis "Investigating the relationship between electron mobility and velocity in deeply scaled NMOS via mechanical stress". IEEE EDL, vol 22, n 12, p591, 2001

[3] G.Ghibaudo et.al, "Electrical Transport Characterization of nano CMOS devices with ultra thin silicon film", p58, IEEE 9th intl workshop on Junction Technology 2009.

[4] A Cros et.al , "Unexpected Mobility degradation for very short devices: A new challenge for CMOS Scaling", IEDM Tech. Digest, DOI:10.1109/IEDM. 2006.346872 (2006)

[5] M S Shur, "Low Ballistic Mobility in Submicron HEMTs" IEEE EDL vol23 n9 p511 2002

[6] Shin-ichi Takagi et.al "On the universality of Inversion layer Mobility in Si MOSFET's: Part I – Effects of Substrate Impurity Concentration", IEEE TED vol41, n12, p2357, 1994

[7] T.Ando, A B Fowler, F Stern, "Electronic Properties of 2D systems", review of modern physics vol54, n2, p437, 1982.

[8] D.S.Jeon, D.E.Burk, "MOSFET Electron Inversion Layer Mobilities-A Physically based Semi-Empirical Model for a Wide Temperature Range" IEEE TED vol 36, n8, p1456, 1989

[9] V Barral et al "Experimental Investigation on the quasi ballistic Transport: Part I – Determination of a New Backscattering Coefficient Extraction Methodology" IEEE TED vol56 n3 p408 2009

[10] F Stern. "Calculated Temperature Dependence of Mobility in Silicon Inversion Layers" PRL vol44 n22 p1469 1980

[11] D Esseni, F Driussi, "A Quantitative Error Analysis of the Mobility Extraction According to the Matthiessen Rule in Advanced MOS Transistors" IEEE TED vol58 n8 p2415 2011

[12] A.Mathiessen and C Vogt, "On the influence of Temperature on Electric Conducting-Power of Thallium and Iron", Phil.Trans.R.Soc.London, vol153, p369-383, 1863

8.3 Transport électronique dans le transistor MOS – Régime de saturation

Dans le transistor MOS long classique, la saturation du courant est provoquée par le pincement du canal près du drain. Le pincement se produit lorsque la tension de drain excède V_g-V_t près du drain. Le courant atteint alors sa valeur maximale pour cette tension de grille:

$$I_{DS} = \frac{W}{2L} C_{ox} \mu_{eff}(V_g)(V_g - V_t)^2 \tag{3.1}$$

Pour les longueurs de canal sub-micrométrique, le transport est toutefois affecté par des effets de champ longitudinal fort. Ceux-ci peuvent se manifester pour des longueurs de grille inférieures à quelques micromètres [3 – 6]. En particulier, la distribution énergétique des porteurs le long du canal s'éloigne de la distribution d'équilibre pour des tensions de drain élevées. L'efficacité des interactions en est modifiée. En particulier la probabilité d'émission de phonons optiques augmente. Ceci se traduit par une saturation de la vitesse des porteurs à une valeur v_{sat} qui pour le silicium massif est autour de 10^7cm/s. [15]

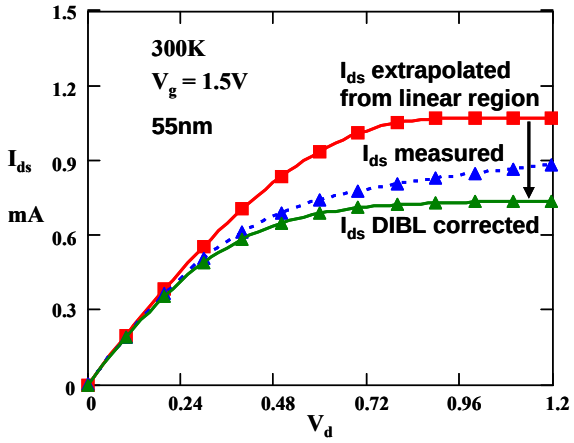


Figure 3.1

Courbes $I_{DS} - V_D$ mesurée (en pointillés bleus), corrigée du DIBL (ligne continue et losanges verts) et extrapolée à partir de la mobilité effective en régime linéaire (ligne continue et carrés rouges). Ces courbes ont été obtenues à température ambiante pour un dispositif FDSON, décrit dans le chapitre 6, de 55nm de longueur de canal et de 1µm de large.

La saturation de vitesse explique par exemple, dans la Figure 3.1, pourquoi le courant mesuré, même corrigé de l'abaissement de la barrière source/canal sous l'effet de la tension de drain (DIBL pour Drain Induced Barrier Lowering en anglais), est plus faible que le courant extrapolé à partir de la mobilité effective extraite en régime linéaire. Le courant en régime de saturation est donné par l'équation 3.2.

$$I_{DS} = WC_{ox}(V_g - V_t)v_{sats} \tag{3.2}$$

Le courant de drain I_{DS} est calculé à proximité de la source où la densité de charge peut s'exprimer comme $C_{ox}(V_g - V_t)$ et où v_{sats} représente la vitesse de saturation des porteurs près de la source [3]. A tension d'alimentation donnée, le courant I_{on} est indépendant de la longueur de grille. Toutefois, la

réduction des dimensions s'accompagne également d'une réduction de la charge capacitive en sortie, de sorte qu'au-delà de l'augmentation en densité d'intégration, la miniaturisation apporte également un gain en performance. Il est possible en outre d'augmenter I_{on} en appliquant une contrainte mécanique [1, 2]. La vitesse de dérive v_d peut s'exprimer en fonction des composantes longitudinale $E_{//}$ et transverse $E_{\perp}$ du champ électrique sous la forme donnée par Thornber [10] :

$$v_d(E_{II}, E_{\perp}) = \frac{\mu(E_{\perp})E_{II}}{\left(1 + (\mu(E_{\perp})E_{II} / v_{sats})^{\beta}\right)^{1/\beta}} \tag{3.3}$$

Dans cette expression $\mu(E\perp)$ représente la variation de la mobilité avec le champ transverse (noté E_{eff} dans le chapitre précédent). La vitesse de saturation est plus faible que pour le matériau massif et décroît normalement quand la température augmente [4, 5, 6, 15]. La vitesse de saturation des trous est inférieure à celle des électrons. Il s'agit en réalité d'une vitesse de saturation "apparente", liée à un blocage de la valeur du champ électrique au niveau de l'injection source/canal et non pas à un mécanisme de saturation de vitesse analogue à ce qui se passe dans le matériau massif [7].

Lorsque la longueur de canal décroît jusqu'à des dimensions fortement sub-microniques (~100nm), on voit apparaître dans le canal des effets de transport hors équilibre ou non stationnaire, correspondant à des situations où l'énergie et la vitesse des porteurs ont une distribution différente de leur distribution d'équilibre pour le champ électrique considéré. Le régime de survitesse s'observe lorsque le temps de transit à travers le canal devient de l'ordre de grandeur du temps de relaxation de l'énergie τ_w [8, 9, 11]. Pour les longueurs de canal de quelques dizaines de nanomètres, il peut être nécessaire d'utiliser un modèle quasi-balistique. Si l'on appelle r_{sat} le coefficient de rétrodiffusion en saturation et $\tilde{v}_{th}$ la vitesse thermique d'injection dégénérée [12, 13], le courant de saturation s'exprime alors sous la forme:

$$I_{DSsat} = WC_{ox}(V_g - V_t)\left[\frac{1 - r_{sat}}{1 + r_{sat}}\right]\tilde{v}_{th} \tag{3.4}$$

On voit que régime balistique, pour lequel $r_{sat} = 0$ impose une limite maximum au courant de drain. Le coefficient r_{sat} a été corrélé à la distance caractéristique L_{kT} de la source où le potentiel électrostatique a chuté de kT/q par rapport au sommet de la barrière de potentiel source/canal [12]. Selon les hypothèses faites, il existe plusieurs méthodes pour extraire ce coefficient de rétrodiffusion. Dans cette thèse, nous ne nous sommes pas mis dans cette hypothèse de régime quasi-balistique. Nous verrons que les dispositifs étudiés paraissent encore assez éloignés du régime balistique. Les analyses du régime de saturation que nous avons faites pour les transistors FD-SOI, jusqu'à des longueurs de grille de 55nm et à différentes températures, montrent que les performances peuvent encore être expliquées par un modèle de saturation de vitesse et par des effets de survitesse, ainsi que nous le verrons dans le chapitre 6.

Références

[1] M Saitoh, Ken Uchida "Universal relationship between Low Field Mobility and High Field Carrier velocity in high K and SiO_2 gate dielectric MOSFETs" IEEE IEDM, 10.1109/IEDM.2006.346757, 2006

[2] A. Lochtefield, D. A. Antoniadis "Investigating the relationship between electron mobility and velocity in deeply scaled NMOS via mechanical stress". IEEE EDL vol 22 n 12 p591 2001

[3] C G Sodini, P K Ko and J L Moll "The Effect of High Fields on MOS Device and Circuit Performances" IEEE TED vol 31, n10, p1386, 1984

[4] R W Coen and R S Muller "Velocity of Surface Carriers in Inversion Layers on Silicon" Solid State Electronics vol23 p35 1980

[5] A Modelli and S Manzini "High Field Drift Velocity of Electrons in Si Inversion Layers" Solid State Electronics vol 31 n1 p99 1988

[6] F F Fang, A B Fowler "Hot Electron Effects and Saturation Velocities in Si Inversion Layers" Journal of Applied Physics, vol 41, n4, p1825 1970

[7] W Muller, I Eisele, "Velocity Saturation in Short Channel Field Effect Transistors" Solid State Communications, vol 34, p447 1980

[8] G.G Shahidi, D A Antoniadis, H I Smith, "Electron Velocity Overshoot at Room and Liquid Nitrogen Temperature in Silicon Inversion Layers" IEEE EDL vol9 n2 p94 1988

[9] T Kobayashi, K Saito, "Two-Dimensional Analysis of Velocity Overshoot Effects in Ultrashort-Channel Si MOSFET's" IEEE TED vol32 n4 p788 1985

[10] K K Thornber, "Relation of drift velocity to low field mobility and high field saturation velocity", Journal of applied physics vol 51 n4 p2127 1980

[11] G Baccarani, S Reggiani "A Compact Double-Gate MOSFET Model Comprising Quantum-Mechanical and Nonstatic Effects" IEEE TED vol 46 n8 p1656 1999

[12] M Lundstrom, Z Ren "Essential Physics of Carrier Transport in Nanoscale MOSFETs" IEEE TED vol49 n1 p133, 2002

[13] V Barral et al "Experimental Investigation on the quasi ballistic Transport: Part I – Determination of a New Backscattering Coefficient Extraction Methodology" IEEE TED vol56 n3 p408 2009

[14] M Chen et.al "Temperature Dependent Channel Backscattering Coefficients in Nanoscale MOSFETs" IEDM p39 2002

[15] C.Jacoboni et.al "A review of some charge transport properties of silicon", Solid-State Elec, vol20, p77-89,1977

8.4 Extraction de paramètres en régime linéaire

En régime linéaire, le courant de drain du MOSFET et la transconductance gm peuvent s'exprimer sous la forme générale donnée par l'équation 4.1:

$$I_{DS} = \frac{W}{L}\mu_{eff}(V_g)Q_{inv}(V_g)V_d$$
$$gm = \frac{dI_{DS}}{dV_g} = \frac{W}{L}V_d\left[\mu_{eff}\frac{dQ_{inv}}{dV_g} + Q_{inv}\frac{d\mu_{eff}}{dV_g}\right] \quad (4.1)$$

On peut donc extraire la mobilité à condition de connaître la charge d'inversion Q_{inv} (C/cm^2), la largeur du canal W, et sa longueur L. La charge d'inversion peut être extraite à partir de la mesure de la capacité grille-canal par la méthode dite "split CV". La charge d'inversion est donnée par l'équation 4.2.

$$Q_{inv}(V_g) = \int_{Vin}^{Vg} C_{gc}(V_g)dV_g \quad (4.2)$$

En forte inversion elle peut être exprimée sous la forme approchée $C_{ox}(V_g\text{-}V_t)$, où C_{ox} est la capacité d'oxyde par unité de surface, également extraite des mesures CV, et V_t est la tension de seuil pour la charge. Le chapitre 4 détaillera l'état de l'art de la méthode split CV et des techniques d'extraction de l'épaisseur équivalente d'oxyde t_{ox} et de la longueur effective de canal L_{eff}. Nous discuterons également ses limites [22].

La mobilité effective peut quant à elle être modélisée sous la forme [5, 6]:

$$\mu_{eff}(V_g) = \frac{\mu_0}{1 + \Theta_1(V_g - V_t) + \Theta_2(V_g - V_t)^2} \quad (4.3)$$

Θ_1 et Θ_2 sont les paramètres d'atténuation d'ordre 1 et 2. Les résistances séries de source et de drain sont incluses dans le paramètre Θ_1 et l'expression du courant de drain peut être réécrite de façon très générale sous la forme:

$$I_{DS} = \frac{\beta V_{gt}V_d}{1 + \Theta_1 V_{gt} + \Theta_2 V_{gt}^2}; \beta = \frac{W_{eff}}{L_{eff}}C_{ox}\mu_0; \Theta_1 = \Theta_{10} + \beta R_{SD}(V_g)$$
$$V_{gt} = V_g - V_t; R_{SD}(V_g) = R_S(V_g) + R_D(V_g) \quad (4.4)$$

$R_{SD}(V_g)$ modélise la résistance série source/drain comme une fonction quelconque de la tension de grille. R_{SD} est en ohms. L'expression empirique de l'équation 4.3 peut être reliée à la mobilité physique [7]. Ceci est intéressant pour exploiter les valeurs de mobilité effective obtenues par les diverses méthodes d'extraction basées sur ce modèle empirique. La signification physique des différents paramètres utilisés dans ce modèle empirique permet d'analyser les mécanismes d'interaction. Le chapitre 6 présentera et discutera ainsi des résultats d'extraction obtenus sur des transistors avancés, comme des MOSFETs FD-SOI, GAA ou des FinFETs.

Certaines des techniques qui sont utilisées pour extraire les paramètres de transport en régime linéaire dans les transistors avancés sont discutées en détail dans la thèse, avec un accent particulier sur l'utilisation de la fonction Y et de la méthode de McLarthy [1-3]. Ces méthodes fournissent une procédure complètement analytique pour reproduire les résultats expérimentaux sans procédure complexe d'ajustement numérique. Un autre avantage de ces méthodes est qu'elles sont valides même si la mobilité et la tension de seuil varient avec la longueur de grille. Cependant les techniques actuelles ne prennent pas en compte la dépendance de la résistance d'accès R_{SD} avec la tension de grille, qui devient nécessaire dans les composants avancés, principalement du fait du rôle plus important que prend le recouvrement de la grille au-dessus des zones LDD dans les MOSFETs fortement sub-microniques [8, 9]. Pour prendre en compte cette dépendance de la résistance S/D avec la tension de grille, nous avons modifié la technique de la fonction Y et mis en place une nouvelle méthodologie d'extraction présentée ci-dessous.

8.4.1 Nouvelle méthodologie d'extraction - Fonction Y modifiée incorporant une dépendance $R_{SD}(V_g)$ pour les CMOS avancés

A partir de l'équation 4.4, il est aisé de montrer que la fonction Y peut être exprimé sous la forme de l'équation 4.5 [4]:

$$Y_{\max} = \frac{I_{DS}}{\sqrt{g_m}} = \frac{\sqrt{\beta V_d}\, V_{gt}}{\sqrt{1-(\Theta_2 + R'_{SD}\,\beta)V_{gt}^2}} \tag{4.5}$$

Dans cette expression ainsi que dans les suivantes, l'apostrophe désigne la dérivée par rapport à V_g. Le terme Θ_{eff} et sa dérivée Θ'_{eff} s'expriment alors selon les équations 4.6 et 4.7:

$$\Theta_{eff}(V_g) \equiv \frac{\beta V_d}{I_{DS}} - \frac{1}{V_g - V_t} = \Theta_{10} + R_{SD}(V_g)\beta + \Theta_2 V_{gt} \tag{4.6}$$

$$\Theta'_{eff}(V_g) = R'_{SD}(V_g)\beta + \Theta_2 \tag{4.7}$$

On en déduit:

$$\Theta_{eff}(V_g) - \Theta_2 V_{gt} = \Theta_{10} + R_{SD}(V_g)\beta \tag{4.8}$$

La procédure d'extraction est illustrée sur la figure 4.1 (à gauche). Les valeurs initiales de β et V_t sont extraites de la fonction Y_{mes} mesurée à faible V_{gt}, où l'effet de champ transverse fort peut être négligé. La procédure d'itération décrite fournit les valeurs estimées de β et V_t après quelques itérations (Figure 4.1, à droite). Cette procédure doit être appliquée à une série de transistors de même largeur de canal mais de longueur de canal variable. On peut alors en déduire, pour toutes les tensions de grille, les valeurs de Θ_{eff} et Θ'_{eff} et utiliser les équations 4.7 et 4.8 pour en déduire Θ_2, Θ_{10} et R_{SD} par interpolation linéaire des courbes $\Theta_{eff}(V_g)$ et $\Theta'_{eff}(V_g)$ en fonction de β, de manière analogue à ce qui est proposé dans [4].

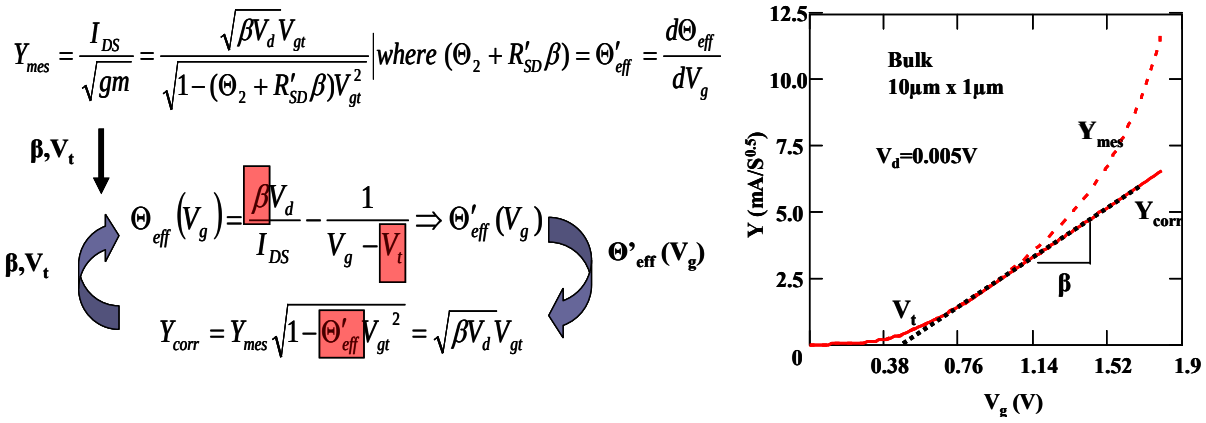


Figure 4.1

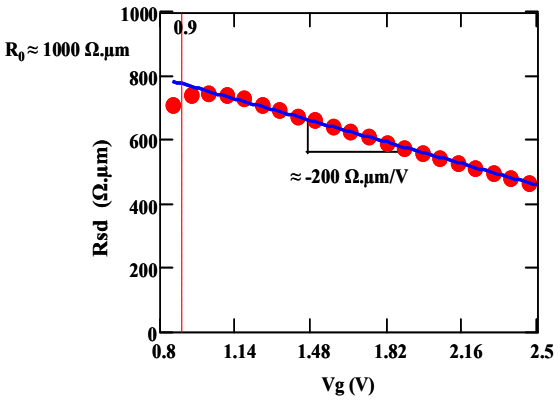
A gauche: procédure d'extraction de β et V_t permettant de s'affranchir de l'effet de la résistance série S/D. A droite: exemple de résultat d'extraction pour un transistor MOSFET sur substrat massif, de largeur de canal 10µm et de laongueur de canal 1µm. On voit sur Y_{mes} l'effet de Θ_2 et de R_{SD}. Y_{corr} est quant à elle linéarisée après quelques itérations.

La figure 4.2 montre le résultat de cette méthodologie d'extraction pour un transistor FDSOI à canal long. Les composants étudiés provenaient du LETI et avaient été fabriqués en utilisant une technologie à faible bilan thermique utilisable pour réaliser l'intégration séquentielle de plusieurs couches de transistors [10]. On constate que R_{SD} varie linéairement avec V_g pour ces transistors (figure 4.2-a). La figure 4.2-b montre que Θ_{10} et Θ_2 présentent un plateau à tension de grille élevée. La figure 4.2-d montre la mobilité μ_0 et la tension de seuil V_t extraites en fonction de la longueur de canal. La figure 4.2-c montre quant à elle la courbe I_d-V_g reconstruite à partir des paramètres extraits. On vérifie qu'elle fournit a posteriori un très bon ajustement des données expérimentales. Nous discuterons plus en détail ces techniques d'extraction, ainsi que les effets d'ordre supérieur comme la variation de Θ avec la longueur de canal dans le chapitre 4 de cette thèse.

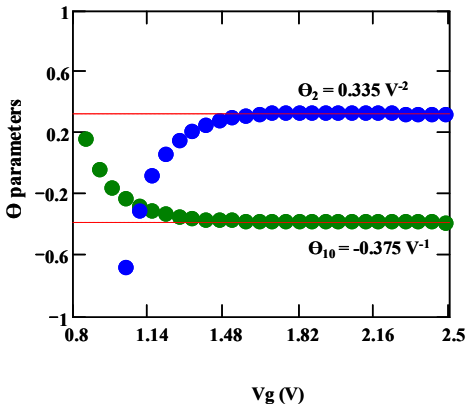
8.4.2 Paramètre η - Extraction pour des transistors sur SOI contraint (sSOI)

Nous avons extrait le paramètre η pour la première fois pour des transistors sur substrat SOI contraint. Pour ce faire, nous avons utilisé la méthode décrite dans [12]. Les composants utilisés pour cette étude provenaient du LETI. Le substrat sSOI utilisé comportait un BOX de 145nm et un film de silicium contraint de 10nm, avec une contrainte bi-axiale en tension de 20%. Les transistors intégraient une grille HfO_2 /TiN présentant une EOT d'environ 2nm. La figure 4.3 montre les valeurs de η obtenues pour des transistors de 10µm et de 2µm de long. On obtient une valeur de l'ordre de η ~ 0.8 qui diffère de manière significative de la valeur conventionnelle de 0.5 obtenue dans les composants sur substrat massif, et qui est même plus élevée que les résultats trouvés dans [11] pour du silicium contraint. Au premier ordre, on voit que la valeur de η vaut respectivement 0.5, 0.6, 0.7 et 0.8 pour des transistors sur bulk, FDSOI, sSi et sSOI. Ceci résulte de l'augmentation de l'espacement entre les niveaux énergétiques des sous bandes et à une redistribution des porteurs entre ces sous bandes qui se traduit

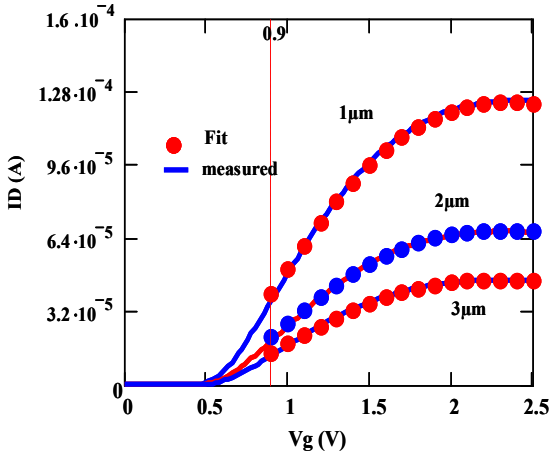
par un confinement des porteurs plus près de l'interface. Il s'en suit un accroissement des interactions avec les phonons et avec la rugosité de surface à champ transverse identique par rapport à un transistor sur substrat massif. Le champ effectif et le paramètre extraits sont de ce fait surestimés. Une discussion plus détaillée pourra être trouvée dans la thèse.



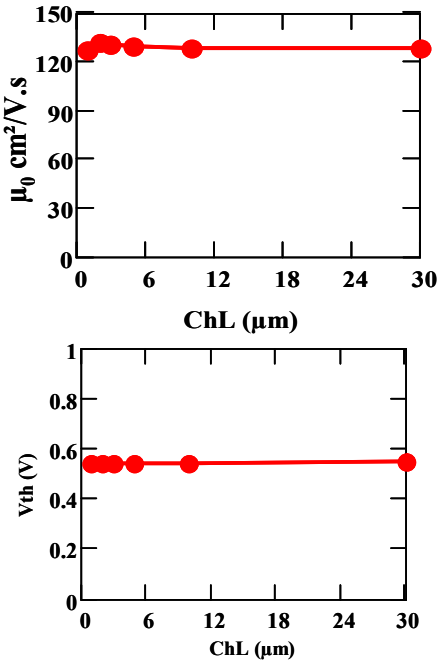
(a) Valeur de R_{SD} (en Ω.µm) extraite en fonction de Vg. Pour cette technologie, on trouve que R_{SD} varie linéairement en fonction de V_g.



(b) Valeurs extraites pour Θ_{10} et Θ_2



(c) Courbes Id-V_g mesurées (lignes) et reconstruites à partir des paramètres extraits (symboles).



(d) Paramètres μ_0 et V_t extraits

Figure 4.2: Résultats de l'extraction effectuée à l'aide de la fonction Y modifiée pour des PMOS à canal long du LETI [10]

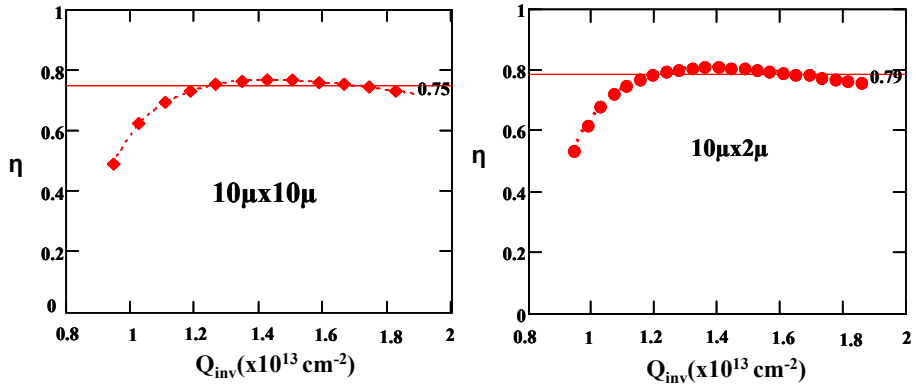


Fig 4.3

Paramètre η extrait pour un transistor de 10µm x 10µm (à gauche) et de 10µm x 2µm (à droite).

8.4.3 Etude de la mobilité à basse température.

Les différents mécanismes d'interaction ont des dépendances en température spécifiques qui permettent de séparer leurs contributions respectives à la mobilité. Dans cette étude, nous avons tenu compte de la composante balistique du transport grâce au modèle de Shur. En supposant valide la décomposition de Mathiessen, la mobilité en champ faible peut s'exprimer comme:

$$\frac{1}{\mu_{0dd}} = \frac{1}{\mu_0} - \frac{1}{\mu_{bal}}$$
$$\frac{1}{\mu_{0dd}(T)} = \frac{1}{\mu_N} + \frac{T}{300\mu_{ph0}} + \frac{300}{T\mu_{C0}} \qquad (4.9)$$

Le terme représente la mobilité en champ faible qui serait observée en l'absence de composante balistique, sous l'effet des seules interactions. Ses différentes composantes peuvent être extraites par ajustement de sa variation en température [7, 13, 14]. Le chapitre 6 en donne un exemple, avec la caractérisation en température de la mobilité dans des transistors FD-SON.

Nous avons généralisé cette méthodologie en introduisant le facteur de gain (Equation 4.10).

$$\frac{1}{\beta(T)} = \left(\frac{300}{\beta_C T} + \frac{T}{300\beta_{Ph}} + \frac{1}{\beta_N} + \frac{T^{0.5}}{\beta_{Bal0}} \right) \qquad (4.10)$$

Dans cette équation, les termes β_c, β_{ph} et β_N représentent respectivement les contributions des interactions avec les charges Coulombiennes, avec les phonons et avec les défauts neutres [14], tandis que le terme β_{Bal0} représente la contribution des porteurs balistiques qui peut s'exprimer sous la forme:

$$\beta_{Bal0} = \frac{qL_{eff}}{\sqrt{2k\pi m^*}} \frac{W_{eff} C_{ox}}{L_{eff}} \qquad (4.11)$$

L'intérêt de cette généralisation est que le facteur de gain est une quantité directement mesurable, indépendamment du détail des dimensions des composants. Elle permet donc d'étendre le domaine d'application de cette méthode aux générations technologiques futures, avec des composants de plus petite dimension où l'incertitude sur les dimensions réelles augmente. Toutefois, cette méthode est

délicate à utiliser et doit être utilisée avec précaution et esprit critique si l'on veut obtenir des résultats sensés. Elle est utilisée dans le chapitre 6 pour caractériser des FinFET et des transistors GAA.

8.4.4 Magnetorésistance en régime balistique et quasi-balistique

La mesure de la mobilité de magnétorésistance (MR) est un outil puissant pour extraire la mobilité dans un transistor MOS. En particulier, il n'est pas nécessaire de connaître les dimensions exactes du composant pourvu que sa largeur soit nettement plus grande que la longueur de canal (W/L > 5). La mobilité de magéntorésistance peut être mesurée directement en régime linéaire où le transistor se comporte en résistance contrôlée. On mesure pour ce faire la variation relative du conrant de drain en fonction du champ magnétique transverse. La mobilité s'en extrait selon la relation [15-19]:

$$\left.\frac{I_{D0}-I_{DB}}{I_{DB}}\right|_{Vg,Vd} = \mu_{MR}^2 B_Z^2 \tag{4.12}$$

Le terme I_{D0} est le courant de drain en l'ansence de champ magnétique, tandis que I_{DB} est sa valeur en présence d'un champ magnétique B_z. L'indice z indique que le champ magnétique est appliqué dans la direction z, perpendiculaire au plan du canal (plan x,y).

Pour les générations sub-32nm, il est nécessaire de reconsidérer la MR pour tenir compte du transport balistique. Dans cette thèse, nous proposons une modélisation analytique de la MR. Nous la généralisons ensuite pour introduire un faible degré d'interaction pour modéliser des composants plus réalistes fonctionnant dans le régime quasi balistique. La mobilité de magnétorésistance balistique est donnée par l'équation 4.13, sous les conditions restrictives que $r_c > L_{ch}$ et $E_x/B_z << v_T$ où r_c est le rayon cyclotronique et v_T est la vitesse thermique, de l'ordre de 10^7cm/s.

$$\mu_{balMR} = \frac{1}{\sqrt{2\pi}} \frac{qL_{ch}}{m^* v_T} \left(\frac{\Im_{-1/2}(\eta_{F1})}{\Im_{1/2}(\eta_{F1})} \right)^{1/2} \tag{4.13}$$

Comme la mobilité balistique, la mobilité de magnétorésistance balistique est une fonction linéaire de la longueur de canal pour des valeurs faibles du champ électrique et du champ magnétique.

En régime de transport quasi balistique, la méthode des flux [21] permet de réduire l'expression du courant de drain et d'obtenir une formulation similaire à celle proposée par Shur pour la mobilité effective:

$$\mu_{QBMR} = \left[\frac{1}{\mu_{balMR}} + \frac{1}{\mu_{constMR}} \right]^{-1} ; \mu_{balMR} \propto L_{ch} \tag{4.14}$$

La figure 4.4 montre comment varient les mobilités MR balistique et quasi balistique, ainsi que la mobilité effective balistique. La figure 4.5 rappelle des résultats de simulation Monte-Carlo publiés récemment [17, 20] qui montrent des tendances similaires. Pour effectuer une comparaison plus

quantitative, il faudrait refaire des simulations avec les hypothèses utilisées pour le modèle analytique. Dans la thèse nous fournissons le détail des calculs permettant d'exprimer la mobilité de magnétorésistance en régime balistique et quasi-balistique. Nous discutons également l'impact des non idéalités et d'un certain nombre d'effets d'ordre supérieur.

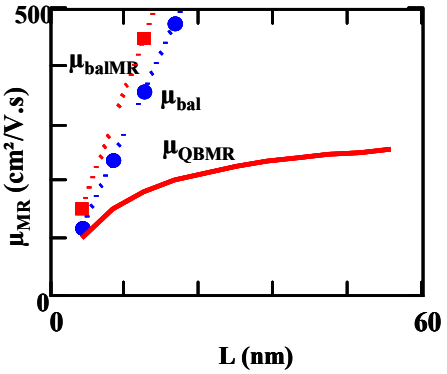


Figure 4.4

Variation théorique des mobilités de magnétorésistance en régime balistique et quasi balistique avec la longueur de canal, en comparaison de la mobilité effective balistique pour un gaz non dégénéré.

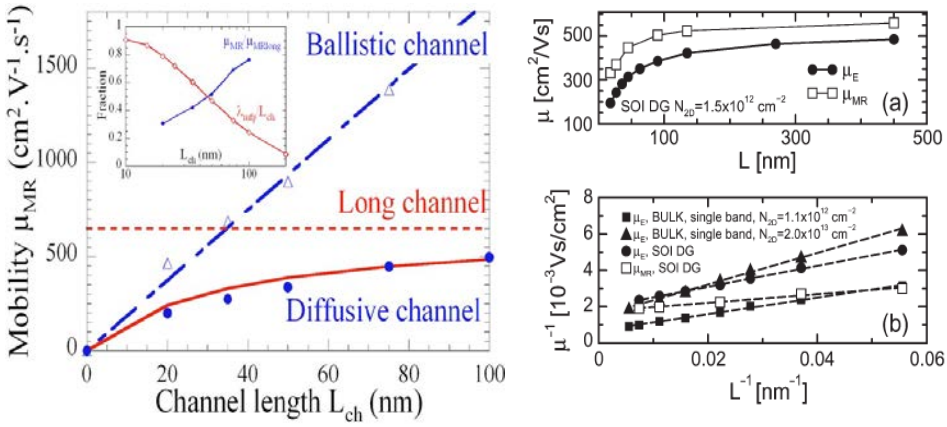


Figure 4.5

A gauche: figure extraite de [17] montrant la dépendance linéaire de avec la longueur de canal , en accord avec la formulation alaytique obtenue ici, ainsi que le cas d'un canal diffusif qui peut être approximé de façon relativement précise par le modèle de Shur avec une mobilité balistique légèrement différente de sa valeur théorique. A droite: figures extraites de [20] montrant en haut les mobilités extraite d'une simulation MSMC, avec là aussi une évolution similaire à celle prédite par notre modèle et en bas la variation linéaire de μ^{-1} avec L^{-1} cohérente avec la décomposition proposée par Shur.

Références

[1] P K McLarty, S. Cristoloveanu, O Faynot, V Misra, J R Hauser and J J Wortman "A Simple Parameter Extraction Method for Ultra Thin Oxide MOSFETs" Sol Stat Elec, vol 38, n6, p1175, 1995

[2] C.Mourrain, B.Cretu, G.Ghibaudo and P.Cottin "New Method For Parameter Extraction in Deep Submicrometer MOSFETs " ICMTS, p181-186, 2000

[3] G.Ghibaudo, "New Method for the extraction of MOSFET Parameters" Elec Letters, vol 24 n9, p543 (1988)

[4] N. Subramanian, G. Ghibaudo, M. Mouis " Parameter extraction of nano-scale MOSFETs using modified Y-function method " IEEE ESSDERC, p309-312, 2010

[5] Tong-Chern Ong, Ping K Ko, and Chenming Hu, "50 A° Gate-Oxide MOSFET's at 77K" IEEE TED, vol34, n10, p2129, 1987

[6] D. Fleury, A. Cros, H. Brut and G. Ghibaudo " New Y-Function Based Methodology for Accurate Extraction of electrical Parameters on Nano-Scaled MOSFETs" ICMTS, p160-165, 2008

[7] G.Reichert, T.Ouisse, "Relationship Between Empirical and Theoretical Mobility Models in Silicon Inversion Layers", IEEE TED, vol 43, n9, p1394, 1996

[8] S D Kim, Jason C S Woo, "Advanced Model and Analysis of Series Resistance for CMOS Scaling Into Nanometer Regime – Part I-Theoretical Derivation", IEEE TED vol 49 n3 p457 2002

[9] S D Kim, Jason C S Woo, "Advanced Model and Analysis of Series Resistance for CMOS Scaling Into Nanometer Regime – Part II- Quantitative Analysis", IEEE TED vol 49 n3 p467 2002

[10] P.Batude et.al "Advances in 3D CMOS Sequential Integration" IEDM, 10.1109/IEDM.2009.5424352, 2009

[11] K Bennamane, M DeMichielis, G Ghibaudo and D Esseni, "Extraction of η parmeter characterizing μ_{eff} against E_{eff} curves in strained Si nMOS " Electronics Letters, vol44, n20, p1219, 2008

[12] K Bennamane, G Ghibaudo and A Benfdila, "Method for extraction of η parameter characterizing μ_{eff} against E_{eff} curves in FD-SOI Si MOS devices " Electronics Letters vol 45 n12 p655 2009

[13] D.S.Jeon, D.E.Burk, "MOSFET Electron Inversion Layer Mobilities-A Physically based Semi-Empirical Model for a Wide Temperature Range" IEEE TED vol 36, n8, p1456, 1989

[14] A.Cros et al. "Unexpected Mobility degradation for very short devices: A new challenge for CMOS scaling", IEDM Tech. Digest, DOI:10.1109/IEDM. 2006.346872 2006

[15] Y M Meziani et al, "Magnetoresistance Mobility Measurements in Sub 0.1μm Si MOSFETs " IEEE ESSDERC, p157, 2004

[16] G Ghibaudo "Transport in the inversion layer of a MOS transistor: use of Kubo-Greenwood formalism" J Phys C: Solid State Physics, vol 19, p769-780, 1986

[17] K Huet et al, "Monte Carlo study of apparent magnetoresistance mobility in the nanometer scale metal oxide semiconductor field effect transistors" Journal of Applied Physics vol 104 – 044504, 2008

[18] Wipa.Chaisantikulwat et al, "Magnetoresistance Technique for Mobility Extraction in Short Channel FDSOI Transistors", ESSDERC, p569, 2005

[19] L.Thevenod et.at "Magnetoresistance mobility extraction on $TiN/HfO_2/SiO_2$ metal-oxide-semoconductor field effect transistors" APL vol 90 – 152111, 2007

[20] M Zilli, D Essni, P Palestri and L Selmi "On the Apparent Mobility in Nanometric n-MOSFETs" IEEE EDL vol28 n11 p1036 2007

[21] A Rahman and M S Lundstrom " A compact scattering model for the nanoscale Double-Gate MOSFET " IEEE TED vol 49, n3 p481 2002

[22] K Romanjek et al, "Improved Split CV Method for Effective Mobility Extraction in Sub-0.1 μm Si MOSFETs" IEEE EDL vol25 n8 p583 2004

8.5 Extraction en régime de saturation

Dans ce chapitre nous discutons l'état de l'art des techniques d'extraction de paramètres en régime de saturation. Les paramètres extraits sont essentiellement la vitesse de saturation v_{sats}, la réduction de la barrière source/canal sous l'effet de la tension de drain (DIBL) et les effets d'auto-échauffement. En outre nous montrons pour la première fois qu'il est possible d'extraire une mobilité de magnétorésistance en régime de saturation du courant. Nous discutons en détail l'état de l'art de l'extraction de v_{sats} [4,5]. Nous montrerons les résultats d'une telle extraction pour des transistors FDSON dans le chapitre 6.

8.5.1 Effets combines de DIBL et d'auto-échauffement

En suivant la même procédure que dans [1] pour l'auto-échauffement et en incluant le DIBL [2], nous avons développé une nouvelle méthodologie, adaptée aux transistors submicroniques, pour extraire la résistance thermique et le DIBL corrigé des effets thermiques, représenté par le paramètre λ (V/V). On peut montrer que le rapport g_d/g_m mesuré en saturation peut s'exprimer comme:

$$\frac{g_{dsat}}{g_m} = \lambda + R_{th}\frac{dI_{DSsat}}{dT}\frac{I_{DSsat}}{g_m} \tag{5.1}$$

Il est possible de tracer la fonction g_d/g_m en fonction du terme $\frac{dI_{DSsat}}{dT}\frac{I_{DSsat}}{g_m}$ pour des valeurs données de V_g et T et pour une valeur variable de V_d. Par interpolation linéaire, on en déduit les valeurs de R_{th} (K/W), à partir de la pente, et de λ (V/V), à partir de l'ordonnée à l'origine (Figure 5.1). Dans la figure 5.1, les caractéristiques montrées en haut à gauche représentent les courbes I_d-V_d mesurées. En haut à droite, on a reporté le rapport g_d/g_m en fonction de V_d. La courbe du bas montre g_d/g_m en fonction de $\frac{dI_{DSsat}}{dT}\frac{I_{DSsat}}{g_m}$. On vérifie bien que cette courbe est linéaire à fort V_d. Les valeurs extraites sont $4.2x10^5$K/W pour R_{th} et 0.2V/V pour λ. La valeur obtenue pour la résistance thermique est en bon accord avec la référence [1] compte tenu des dimensions du composant et de l'épaisseur du BOX. On constate également sur la figure 5.1 (en haut à droite) que le DIBL extrait directement à partir du plateau de g_d/g_m [2] est sous-estimé du fait de l'auto-échauffement. La nouvelle méthode donne une valeur un peu plus élevée. Nous détaillons le déroulement des calculs et discutons ces effets en détail dans la thèse.

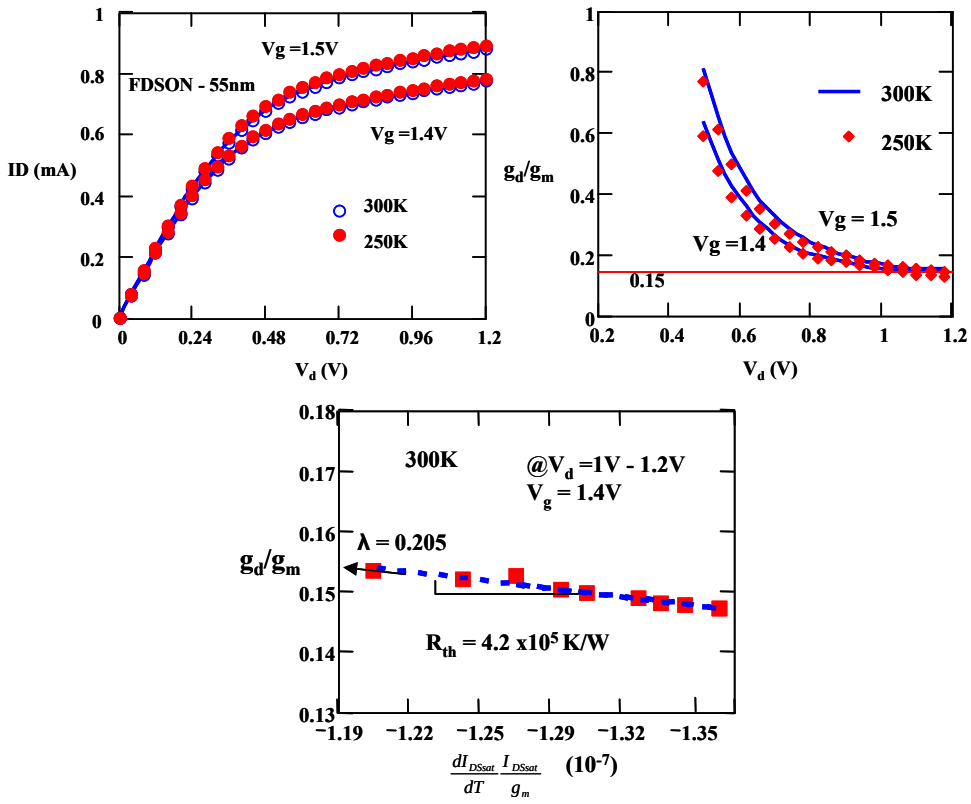


Figure 5.1

Extraction du DIBL et de la résistance thermique par la nouvelle méthode proposée dans la thèse, pour des transistors FD SON de 55nm. En haut, à gauche: caractéristiques I_d-V_d pour deux valeurs de V_g (1.4V et 1.5V) et pour des températures de 300K et 250K. En haut, à droite: variation de g_d/g_m avec V_d. Le DIBL extrait directement du plateau à fort V_d est sous-estimé parce qu'on néglige les effets d'auto-échauffement. En bas: variation de g_d/g_m avec $\frac{dI_{DSsat}}{dT}\frac{I_{DSsat}}{g_m}$ montrant bien une variation liénaire à fort. La pente permet d'extraire la résitance thermique R_{th} (K/W) et l'ordonnée à l'origine fourni le DIBL λ (V/V) corrigé des effets d'auto échauffement.

8.5.2 Magnétorésistance en régime de saturation

L'intérêt des mesures de magnétorésistance a déjà été discuté dans le chapitre 4. Nous avons investigué pour la première fois l'effet d'un champ magnétique transverse en régime de saturation du courant. Nous montrons qu'on peut définir une mobilité de magnétorésistance en saturation, μ_{MRsat}, qui dépend de la longueur de canal et de la température de façon analogue à la mobilité de magnétorésistance extraite en régime linéaire, μ_{MRlin}. Les mesures ont été faites au LNCMI de Grenoble, pour des champs statiques variant de 0T à 11T, à une température proche de l'ambiante (270K) et à basse température (200K et 100K). Pour ces expériences, nous avons utilisé des MOSFETs sur substrat massif à grille high-κ/métal (EOT de 2.4nm). La largeur de grille était fixée à 10µm, tandis que la longueur de canal variant de 1µm à 50nm, respectant la condition W/L > 5. La figure 5.2 montre les valeurs de μ_{MR}

extraites en fonction de V_d, pour différentes longueurs de canal, et pour une tension de grille (V_g=1V) et une température (270K) données.

On constate qu'il est effectivement possible d'extraire μ_{MR} en fonction de la tension de drain depuis le régime linéaire, avec une valeur μ_{MRlin}, jusqu'au régime de saturation du courant, avec une valeur μ_{MRsat}. On vérifie que μ_{MRsat} est pratiquement indépendant de V_d. La figure 5.3 montre la variation de μ_{MRsat} et μ_{MRlin} avec la longueur de canal, ainsi que la relation entre μ_{MRsat} et μ_{MRlin}. On constate que les valeurs extraites en régime linéaire et en régime de saturation sont corrélées.

Par ailleurs, en utilisant un modèle de courant de drain en saturation dépendant de la vitesse de saturation à la source, v_{sats}, nous avons extrait la valeur de v_{sats} pour différentes longueurs de canal à champ magnétique nul à partir du plateau de $g_m/C_{ox}W$ à fort V_d, où g_m est la transconductance, C_{ox} est la capacité de grille et W est la largeur de canal. Les valeurs de v_{sats} et de μ_{MRsat} extraites à V_d=1.1V et V_g=1.1V sont tracées en fonction de la longueur de canal sur la figure 5.4 (à gauche). Noter que la valeur de v_{sats} extraite par cette méthode est très inférieure à la vitesse de saturation dans le silicium massif à cette température (~ 10^7cm/s), même pour une longueur de canal de 50nm. Nous pouvons voir que μ_{MRsat} et v_{sats} sont corrélés et que μ_{MRsat} décroît quand v_{sats} croît (figure 5.4 à droite). On peut interpréter ce résultat en utilisant le modèle de blocage du champ électrique à la source [3] et détaillé en détail dans la thèse.

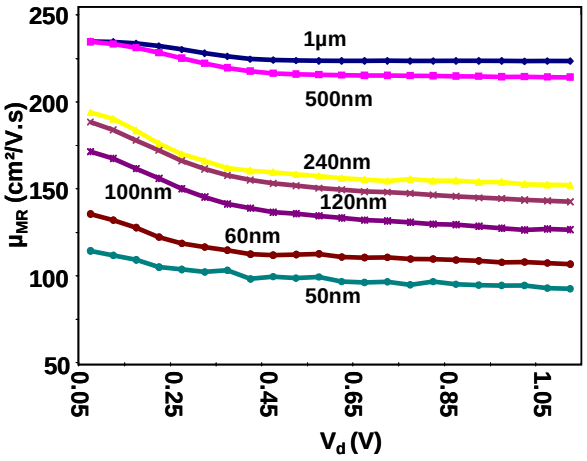


Figure 5.2

Variation en fonction V_d de de la mobilité μ_{MR} extraite à 270K pour V_g = 1V et pour différentes longueurs de canal.

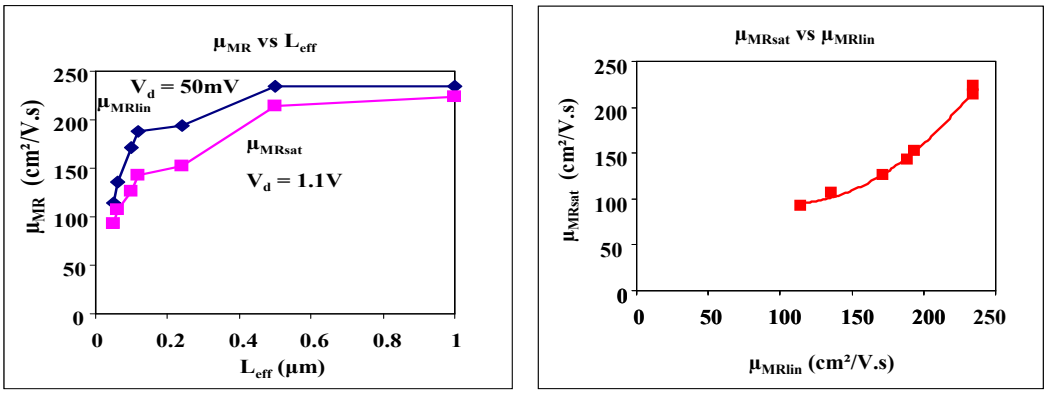


Figure 5.3

A gauche: Variation en fonction de la longueur de canal L_{eff} des mobilités de magnétorésistance μ_{MRlin} , extraite en régime linéaire (à V_d=50mV) et μ_{MRrsat} , extraite en saturation (à V_d = 1.1V) à 270K. V_g=1V. A droite: relation entre μ_{MRsat} et μ_{MRlin}

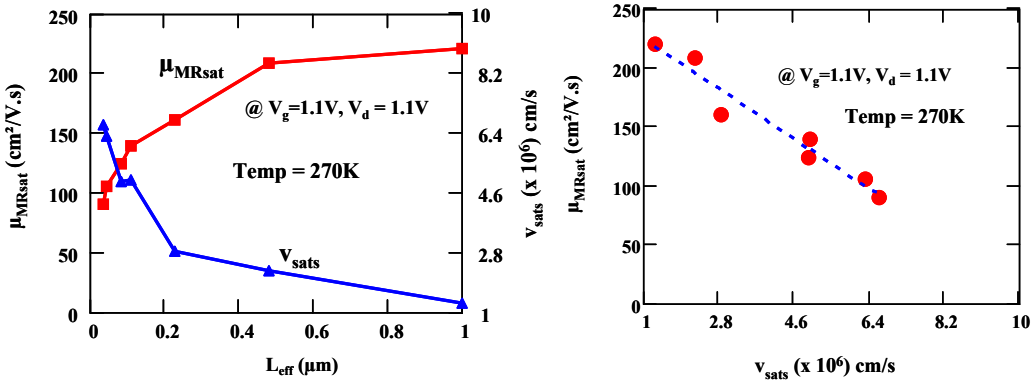


Figure 5.4

A gauche: Tracé de μ_{MRsat} et v_{sats} en fonction de la longueur de canal L_{eff}. v_{sats} est extraite du plateau de $g_m/C_{ox}W$ à fort V_d et champ magnétique nul. A droite: Relation entre les valeurs extraites pour μ_{MRrsat} et les valeurs v_{sats} extraites à champ magnétique nul. T=270K et V_g=1.1V.

Références

[1] J Jomaah, G Ghibaudo, F Balestra, “Analysis and Modelling of self-heating effects in thin-film SOI MOSFETs as a function of temperature” Solid State Electronics vol 38 n3 p615 1995

[2] W Firky, G Ghibaudo and M Dutoit, “Temperature dependence of drain-induced barrier lowering in deep sub-micrometer MOSFETs” EL, vol 30, n11, p911, 1994

[3] W Muller, I Eisele, “Velocity Saturation in Short Channel Field Effect Transistors” Solid State Communications, vol 34, p447 1980

[4] K Rais, G Ghibaudo, F balestra, and M Dutoit, “Study of saturation velocity overshoot in deep submicron silicon MOSFETs from liquid helium up to room temperature” Journal de Physique IV C6 suppl jour de Phys III, vol 4, p19, 1994

[5] G G Shahidi, D A Antoninadis, H I Smith “Electron Velocity Overshoot at Roon and Liquid Nitrogen Temperatures in Silicon Inversion Layers” IEEE EDL, vol9, n2, p94, 1988

8.6 Résultats et discussion

Nous donnons ici une description rapide des résultats de caractérisation obtenus pour des transistors FD-SON, GAA et FinFET. On trouvera le détail des discussions dans la thèse.

8.6.1 FDSON – Résultats d'extraction

Les composants étudiés ont une grille high-κ/métal HfO_2/HfSiON/TiN(10nm), une épaisseur de silicium de 15nm et un oxyde enterré de 25nm (technologie Bulk+ de STMicroelectronics) [1, 2]. L'épaisseur équivalente d'oxyde est de 1.84 nm. La longueur de canal varie de 55 nm à 0.24 µm avec une largeur W = 1 µm. Nous en avons fait une étude en température, à la fois en régime linéaire et en saturation. La tension de seuil V_t et la mobilité μ_0 ont été extraites en fonction de la longueur de canal et de la température (Figures 6.1 et 6.2). Un effet de canal court inverse (RSCE) se manifeste sur la variation de V_t avec la longueur de canal (Figure 6.1). Il est lié à la présence de zones halos implantées.

Nous avons utilisé le modèle de mobilité basse température discuté au chapitre 4 pour séparer les contributions des différents processus d'interaction (Figure 6.2, en haut). Nous avons utilisé la référence [3] pour l'interaction avec les phonons. A température ambiante, l'interaction avec les phonons est le processus dominant pour les canaux de longueur supérieure à 100nm, tandis que l'interaction avec les défauts neutres apporte une contribution non négligeable pour les canaux plus courts, dégradant ainsi la mobilité de canal court. A 77K, l'interaction Coulombienne joue un rôle plus important tandis que l'interaction avec les phonons perd de l'importance (Figure 6.2, en bas). Les résultats d'extraction en régime de saturation sont reportés sur les figures 6.3, 6.4 et 6.5. Le DIBL varie de façon typique avec la température et la longueur de canal (Figure 6.3). Sur la figure 6.4, la variation de v_{sats} semble indiquer une présence de survitesse à 77K. La figure 6.5 quant à elle compare la valeur de v_{sats} extraite pour les composants les plus courts (55nm) avec la valeur limite théorique correspondant à une injection balistique dans le modèle de Natori [4]. On en conclut que le transport est encore fortement influencé par les mécanismes d'interaction et que ces composants sont encore éloignés de la limite balistique.

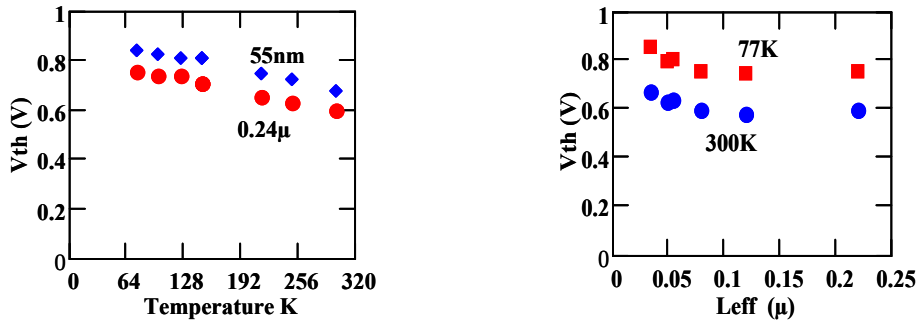


Figure 6.1

Dépendance de la tension de seuil V_t en fonction de la temperature (à gauche) de de la longueur de canal Leff (à droite, montrant la présence d'un effet de canal court inverse - RSCE).

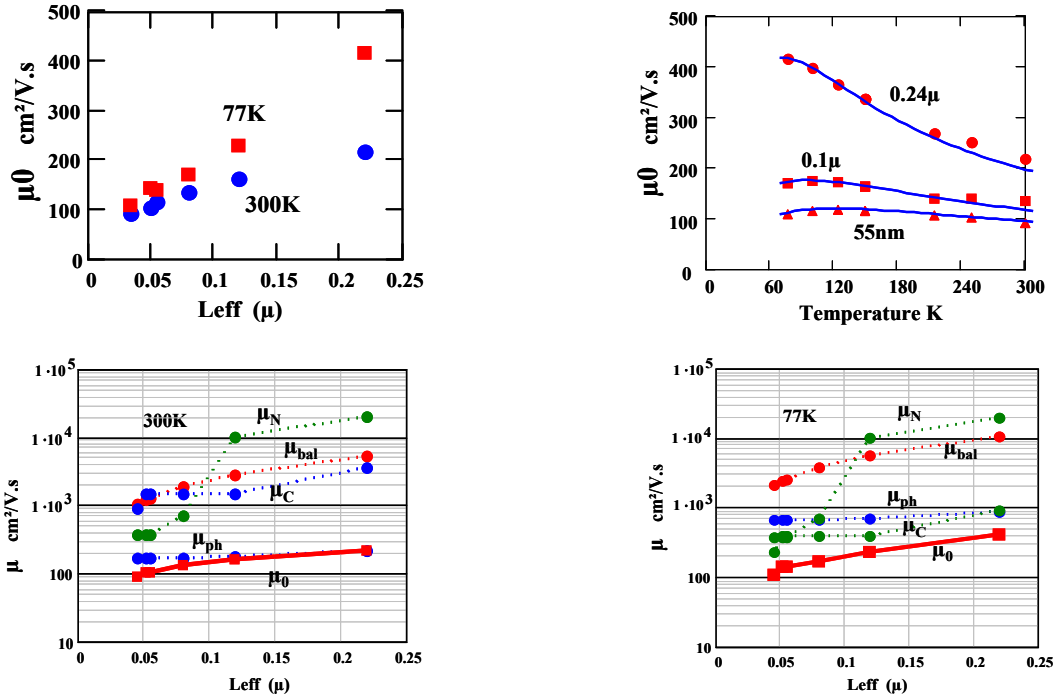


Figure 6.2

En haut, à gauche: Variation de μ_0 en fonction de L_{eff}. Valeurs extraites, montrant une dégradtion en canal court. En haut, à droite: Résultat de l'ajustement du modèle de μ_0 (lignes) sur les valeurs extraites (symboles) en fonction de la température et de la longueur de grille. En bas: Variation de μ_0 avec L_{eff} et valeurs extraites pour les contributions μ_C, μ_{bal}, μ_N et μ_{ph} à 300K et 77K.

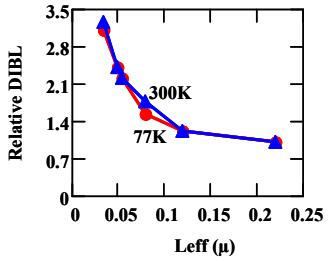


Figure 6.3

DIBL (relatif à la valeur en canal long) en fonction de L_{eff} à 300K et 77K

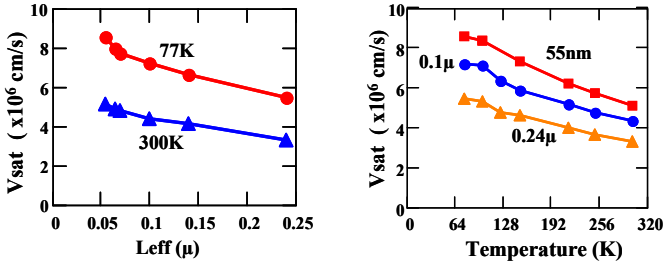


Figure 6.4

Variation de v_{sats} avec L_{eff} (à gauche) et avec la température (à droite)

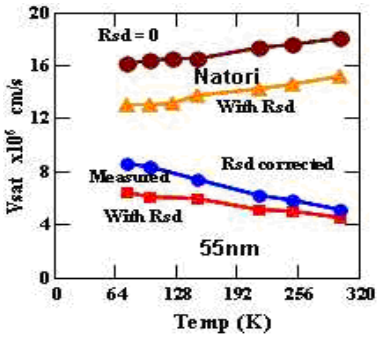


Figure 6.5

Comparaison entre la vitesse d'injection balistique calculée à l'aide du modèle de Natori [4] et les valeurs de v_{sats} extraites pour les composants de 55nm.

8.6.2 Extraction de paramètres pour des transistors FinFET

Les FinFETs étudiés ici ont été fournis par l'IMEC. Leur fabrication a été décrite dans la référence [8]. Les composants étudiés avaient une hauteur d'aileron de 65nm, une longueur de canal de 35nm et une largeur d'aileron de 10nm. L'étude en température a été effectuée entre 100K et 380K en utilisant le facteur de gain β, comme indiqué dans le chapitre 4. Nous avons utilisé la référence [3] pour l'interaction avec les phonons. Nous avons tenu compte de l'orientation différente des flancs des ailerons, orientés selon le plan (110), avec une mobilité $\mu_{ph0\ (110)}=110cm^2/V.s$. Pour les besoins de la procédure d'ajustement, le modèle de l'équation 4.10 a été légèrement modifié sous la forme:

$$\beta \equiv MF \times \mu_0$$
$$\frac{1}{\mu_0} = \frac{300}{\mu_{C0} T} + \frac{T}{300\mu_{Ph0}} + \frac{1}{\mu_{const}} + \frac{T^{0.5}}{\mu_{b0}} \qquad (6.1)$$

L'ajustement a été effectué en prenant une valeur $t_{ox}=1.9nm$ comme dans la référence [8]. La figure 6.6 montre les contributions relatives des différents mécanismes d'interaction, telles que nous les avons extraites de l'ajustement de ce modèle avec les résultats expérimentaux. Cette figure montre un haut niveau d'interaction avec les phonons et la présence d'une contribution constante, μ_{const}, qui est en général attribuée à l'interaction avec les défauts neutres. On extrait une mobilité balistique à 300K ($\mu_{bal}=T^{-0.5}\times\mu_{b0}$) de ~$600cm^2/V.s$ et une mobilité μ_0~$75cm^2/V.s$. Suivant l'ajustement que nous avons effectué, l'interaction dominante à température ambiante est celle avec les phonons pour la mobilité à

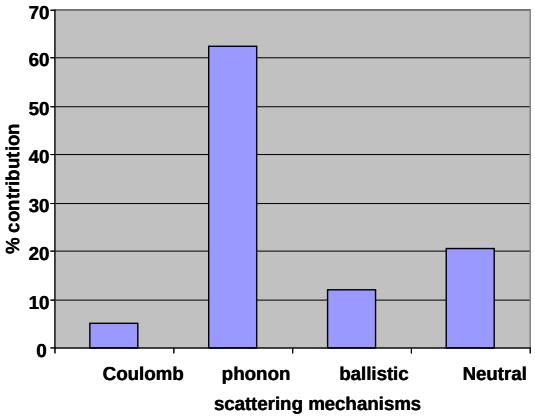


Figure 6.6

Contributions relatives extraites pour les différents mécanismes d'interaction et pour la mobilité balistique dans la mobilité faible champ à 300K. On constate que l'interaction avec les phonons est prépondérante.

champ faible. Des études antérieures ont montré également que l'interaction avec la rugosité d'interface était plus importante pour les flancs verticaux que pour la face supérieure des ailerons, du fait de leur rugosité supérieure [7]. En forte inversion (champ fort), le courant est donc limité principalement par l'interaction avec les phonons et la rugosité de surface.

8.6.3 Extraction pour des transistors GAA planaires

Les transistors à grille enrobante (GAA) sont une alternative au MOSFET classique pour les technologies les plus miniaturisées. Les transistors planaires GAA ont des performances similaires aux transistors DG lorsque la largeur des grilles supérieure et inférieure est supérieure à l'épaisseur du canal de conduction, mais le contrôle de grille y est encore amélioré par un contrôle latéral et ils permettent de créer un nombre effectif de canaux par unité de largeur de masque supérieur à 2. Les composants GAA étudiés ici ont été fabriqués à STMicroelectronics sur la base de la technologie FDSON présentée dans le paragraphe 6.1 [5, 6]. Les composants ont une grille high-κ/métal HfO_2/HfSiON/TiN, avec 2.5nm de HfO_2 et 10nm de TiN. L'épaisseur de silicium est de 10nm. \

Les dispositifs dont nous étudions ici les propriétés de transport, grâce à une étude en température (entre 80K et 395K), avaient des dimensions de grille W×L de 0.11µm×0.05µm sur masque. Du fait de l'incertitude importante sur les dimensions réelles, nous avons utilisé la même technique d'extraction que pour les FinFETs (équation 6.1). La mobilité de phonons, extraite de la référence [3] a été fixée à 170 cm^2/V.s à 300K. La contribution relative des différents mécanismes d'interaction et du transport balistique, telle qu'extraite de notre ajustement sur les résultats expérimentaux, est montrée sur la figure 6.7. La mobilité μ_0 à température ambiante valait ici ~ 100 cm^2/V.s. On constate que, là encore, les composants sont dans un régime de transport encore largement dominé par les interactions. A faible champ transverse, les interactions dominantes sont celles avec les phonons et avec les charges Coulombiennes.

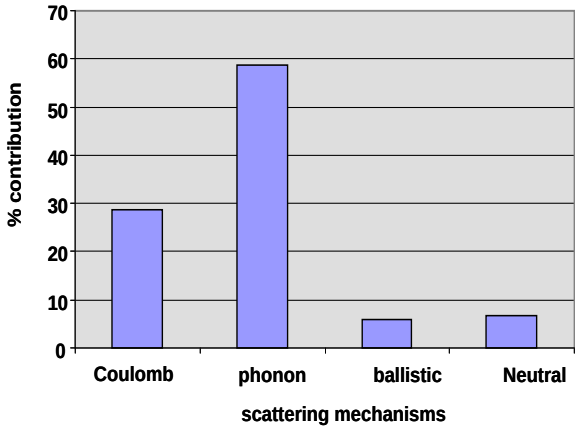


Figure 6.7

Contributions relatives des différents processus d'interaction et du transport balistique à la mobilité en champ faible à 300K. On observe que l'interaction avec les phonons est dominante et que les interactions Coulombiennes ont également une contribution significative.

Références

[1] G.Bidal et al "Planar Bulk^{+} Technology using TiN/Hf-based gate stack for Low Power Applications ", VLSI Tech. Digest, p146-147 2008

[2] J. Pretet, S Monfray, S.Cristoloveanu, T.Skotnicki, "Silicon on Nothing MOSFETs: Performance, Short-Channel Effects, and Backgate Coupling", IEEE TED, vol 51 n2, p.240 2004

[3] A.Cros et al "Unexpected Mobility degradation for very short devices: A new challenge for CMOS Scaling" IEDM Tech. Digest, DOI:10.1109/IEDM. 2006.346872 2006

[4] K.Natori, "Ballistic metal-oxide-semiconductor field effect transistor" J.Appl.phys, Vol 76, n8, p4879 1994

[5] J.-L. Huguenin et al, "Ultra-Thin (4nm) Gate-All-Around CMOS devices with High-k/Metal for Low Power Multimedia Applications" SSDM 2010

[6] G Bidal et.al, "Gate-All-Around technology: taking advantage of ballistic transport" IEEE ESSDERC, p315, 2009

[7] J W Lee et.al, "Experimental Analysis of Surface Roughness Scattering in FinFET devices" IEEE ESSDERC, p305, 2010

[8] N Collaert et.al, "Tall tripple gate devices with TiN/HfO_2 gate stack " VLSI Symp tech dig, p108, 2005