Status of thesis

Title of thesis	Downscaling of 0.35 μm to 0.25 μm CMOS Transistor by Simulation

I NORAINI BT OTHMAN

hereby allow my thesis to be placed at the Information Resource Center (IRC) of Universiti Teknologi Petronas (UTP) with the following conditions:

1. The thesis becomes the property of UTP
2. The IRC of UTP may make copies of the thesis for academic purposes only.
3. This thesis is classified as

☐ Confidential

☒ Non-confidential

If this thesis is confidential, please state the reason:

The contents of the thesis will remain confidential for ______ years.

Remarks on disclosure:

Endorsed by

Signature of Author

Permanent address :
1234, Jln. Bunga Raya,
Kg. Sg. Tapah,
30020, Ipoh, Perak

Date : 31/07/06

Signature of Supervisor

Name of supervisor
AP Dr Norani Muti Mohamed

Date : 31/07/06

Approval Page

UNIVERSITI TEKNOLOGI PETRONAS

Approval by Supervisor (s)

The undersigned certify that they have read, and recommend to the Postgraduate Studies Programme for acceptance, a thesis entitled **"Downscaling of 0.35 µm to 0.25 µm CMOS Transistor by Simulation"** submitted by **Noraini bt Othman** for the fulfillment of the requirements for the degree of Master of Science in Electrical and Electronics Engineering.

Date 31st July 2006

Main Supervisor

Assoc Prof Dr Norani Muti Mohamed
Senior Lecturer
Electrical & Electronic Engineering
Universiti Teknologi Petronas
31750 Tronoh
Perak Darul Ridzuan, MALAYSIA

31st July 2006

Co - Supervisor I

Mohd Awan

Dr. Mohammad Awan
Associate Professor
Electrical & Electronic Engineering
31750 Tronoh Perak
Tel: 05-368 7834 Fax: 365 7443

31st July 2006

Title Page

UNIVERSITI TEKNOLOGI PETRONAS

Downscaling of 0.35 µm to 0.25 µm CMOS Transistor by Simulation

By

Noraini bt Othman

A THESIS

SUBMITTED TO THE POSTGRADUTE STUDIES PROGRAMME

AS A REQUIREMENT FOR THE

DEGREE OF MASTER OF SCIENCE IN ELECTRICAL AND ELECTRONICS ENGINEERING

ELECTRICAL AND ELECTRONICS ENGINEERING

BANDAR SERI ISKANDAR

PERAK

JULY 2006

Declaration

I hereby declare that the thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at UTP or other institutions.

Signature :

Name : NORAINI BINTI OTHMAN

Date : 31/07/02

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisor, Associate Professor Dr. Norani Muti Mohamed, for her continuous guidance and support given for the past two years. The advice and motivation given from time to time is most welcome whenever I'm facing difficulties to proceed with my research work. Even with her significant workload in the Department of Electrical and Electronics Engineering, she has always guided me in the right direction with insight and commitment. My deepest appreciation also goes to my co-supervisor, Associate Professor Dr. Mohd Awan for his valuable assistance.

I also wish to thank the Postgraduate Office and its staff, namely the Director of Postgraduate Office, Associate Professor Ir Dr Ahmad Fadzil Mohamad Hani and his staffs - Mrs. Haslina and Mrs. Norma for the supports and assistance given throughout my research period in UTP. Special credit also goes to the Technical Staff of Virtual Fabrication Lab, Mr. Rosli and Mr. Megat, both who have been incredibly helpful and most efficient in providing technical support when needed.

I would also like to thank Mr. Asim Ali and Mr. Shahab A. Najmi at IC Microsystems, for entertaining all those technical queries and clearing out uncertainties when it comes to the Silvaco TCAD Tools Software utilization.

Last but not least, I would like to thank all persons who have contributed to this project but have been inadvertently not mentioned.

ABSTRACT

Silicon (Si) based integrated circuit (IC) has become the backbone of today's semiconductor world with MOS transistors as its fundamental building blocks. The integrated circuit complexity has moved from the early small-scale integration (SSI) to ultra-large-scale integration (ULSI) that can accommodate millions of transistors on a single chip. This evolution is primarily attributed to the concept of device miniaturization. The resulting scale-down devices do not only improve the packing density but also exhibit enhanced performance in terms of faster switching speed and lower power dissipation. The objective of this work is to perform downscaling of 0.35 μm to 0.25 μm CMOS transistor using Silvaco 2-D ATHENA and ATLAS simulation tool. A "two-step design" approach is proposed in this work to study the feasibility of miniaturization process by scaling method. A scaling factor, κ of 1.4 (derived from direct division of 0.35 with 0.25) is adopted for selected parameters. The first design step involves a conversion of the physical data of 0.35 μm CMOS technology to the simulated environment, where process recipe acquired from UC Berkeley Microfabrication Lab serves as the design basis. The electrical data for the simulated structure of 0.35 μm CMOS was extracted with the use of the device simulator. Using the simulated, optimized 0.35 μm structure, downscaling to a smaller geometry of 0.25 μm CMOS transistor was carried out and subsequent electrical characterization was performed in order to evaluate its performance. Parameters that are monitored to evaluate the performance of the designed 0.25 μm CMOS transistor include threshold voltage (V_{th}), saturation current (I_{dsat}), off-state leakage current (I_{off}) and subthreshold swing (S_t). From the simulation, the V_{th} obtained is of 0.51 V and -0.4 V for NMOS and PMOS respectively, with a difference of 15%-33% as compared to other reported work. However, for results of I_{dsat}, the values obtained which is of 296 μA/μm for NMOS and 181 μA/μm for PMOS is much lower than other reported work by 28%-50%. This is believed to be due to direct scaling of 0.25 μm transistor from the 0.35 μm geometry without alterations on the existing structure. For I_{off} and S_t, both results show a much better value as compared to other work. I_{off} obtained which is of <10 pA/μm is about 80%-96% lower than the maximum allowable specification. As for S_t, the values obtained which is <90 mV/dec is only within 5% differences as compared to specification. In overall, these results (except for I_{dsat}) are found to be within the range of accepted values for the particular 0.25 μm technology. From this work, the capability to perform device miniaturization from 0.35 μm to 0.25 μm has been developed. This is achieved by acquiring the technical know-how on the important aspects of simulation required for successful simulation of 0.35 μm technology. Ultimately, the outcome of this work which is a simulated 0.25 μm CMOS transistor can be used as a basis for scaling down to a much smaller device, namely towards 90-nm geometry.

ABSTRAK

Litar bersepadu berasaskan silikon menjadi tulang belakang kepada dunia semikonduktor masa kini dengan transistor MOS sebagai blok binaannya yang terpenting. Kompleksiti litar bersepadu telah mengalami evolusi bermula dari integrasi-berskala-rendah (SSI) ke integrasi-berskala-ultra (ULSI) di mana kini, jutaan transistor dapat dimuatkan di atas sekeping cip. Pendorong utama kepada evolusi ini adalah pengenalan kepada konsep penskalaan peranti. Peranti yang terhasil daripada kaedah penskalaan ini bukan sahaja menunjukkan peningkatan dari segi kepadatan pembungkusan, tetapi turut menunjukkan peningkatan keupayaan dari segi kelajuan pensuisan yang lebih tinggi dan kehilangan kuasa pelesapan yang lebih rendah. Objektif kerja ini adalah untuk merekabentuk peranti CMOS 0.25 μm melalui kaedah penskalaan daripada data fizikal teknologi 0.35 μm dengan menggunakan simulator Silvaco 2-D ATHENA dan ATLAS. Di dalam kerja ini, pendekatan "dua peringkat" dicadangkan dalam mengkaji keupayaan penghasilan pengecilan peranti melalui kaedah penskalaan. Faktor penskalaan, κ iaitu 1.4 (diperolehi melalui pembahagian secara terus 0.35 μm dengan 0.25 μm) digunakan ke atas beberapa parameter terpilih. Peringkat pertama rekabentuk melibatkan penukaran data fizikal dari teknologi CMOS 0.35 μm ke persekitaran simulasi, di mana resepi proses yang diperolehi dari Makmal Mikrofabrikasi UC Berkeley dijadikan sebagai asas rujukan. Data elektrik untuk peranti CMOS 0.35 μm seterusnya diperolehi dari simulator peranti. Dengan menggunakan struktur 0.35 μm yang telah disimulasi dan dioptimakan, penskalaan kepada geometri yang lebih kecil iaitu transistor CMOS 0.25 μm dilakukan dan pencirian elektrik dijalankan untuk menilai prestasinya. Parameter yang menjadi petunjuk akan prestasi transistor CMOS 0.25 μm adalah voltan ambang (V_{th}), arus ketepuan (I_{dsat}), arus bocor (I_{off}) dan ayunan sub-ambang (S_t). Daripada simulasi, nilai V_{th} yang diperolehi adalah 0.51 V dan -0.4 V untuk NMOS dan PMOS, iaitu perbezaan di antara 15%-33% jika dibandingkan dengan hasil yang dilaporkan di dalam kertas kerja yang lain. Walaubagaimanapun, bagi nilai untuk I_{dsat}, keputusan yang diperolehi iaitu 296 μA/μm untuk NMOS dan 181 μA/μm untuk PMOS adalah jauh lebih rendah jika dibandingkan dengan nilai yang dilaporkan di dalam kertas kerja lain iaitu sebanyak 28%-50%. Adalah dipercayai bahawa penskalaan secara terus yang dilakukan terhadap transistor asal yang bergeometri 0.35 μm kepada transistor 0.25 μm tanpa sebarang perubahan pada struktur sedia ada,

menyebabkan I_{dsat} yang diperolehi adalah lebih rendah. Untuk I_{off} dan S_t, nilai yang diperolehi untuk kedua-dua transistor NMOS dan PMOS menunjukkan nilai yang lebih baik berbanding dengan lain-lain nilai yang dilaporkan. I_{off} yang diperolehi iaitu <10 pA/µm adalah 80%-96% lebih rendah jika dibandingkan dengan spesifikasi maksimum yang dibenarkan. Untuk S_t, nilai yang diperolehi iaitu <90 mV/dec menunjukkan perbezaan kurang daripada 5% jika dibandingkan dengan spesifikasi yang ditetapkan. Secara keseluruhannya, didapati bahawa nilai yang diperolehi dari simulasi (kecuali nilai bagi I_{dsat}) adalah di dalam julat nilai yang diterima bagi teknologi 0.25 µm. Daripada kerja ini, keupayaan untuk melaksanakan penskalaan peranti telah dapat dibangunkan. Ini dapat dicapai melalui perolehan pengetahuan mengenai perkara penting yang terlibat dalam simulasi yang diperlukan untuk menjayakan simulasi peranti teknologi 0.35 µm. Seterusnya, keputusan simulasi transistor CMOS 0.25 µm boleh dijadikan asas dalam merekabentuk peranti yang lebih kecil dimensinya iaitu ke arah peranti bergeometri 90-nm.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

This chapter starts off with the background of the research and the motivation for this work. This is followed by the descriptions of the objective, scope of the research and the outline of the thesis.

1.1 Background of Research

1.1.1 Metal-Oxide-Semiconductor Field-Effect-Transistor (MOSFET)

Ever since it was fabricated in 1960, metal-oxide-semiconductor field-effect-transistor (MOSFET) had become the basic building block of very-large-scale-integrated (VLSI) circuits, making it the most important microelectronic device. MOSFET is a four-terminal device with gate, source, drain and substrate (body). There are two types of MOSFET, namely the n-channel MOS (NMOS) and p-channel MOS (PMOS). For n-channel device, it consist of a p-type semiconductor substrate into which two n^+ regions which is the source and drain are formed. A p-channel device may be considered by exchanging p for n and reversing the polarity of the voltage. The metal contact on the insulator is called gate; heavily doped polysilicon or a combination of silicide and polysilicon can also be used as the gate electrode. The basic device parameters are the channel length *(L)*, which is the distance between the two metallurgical n^+-p junctions; the channel width *(W)*; the gate oxide thickness (t_{ox}) and the junction depth (x_j). Figure 1.1 illustrates the basic structure of a MOSFET.

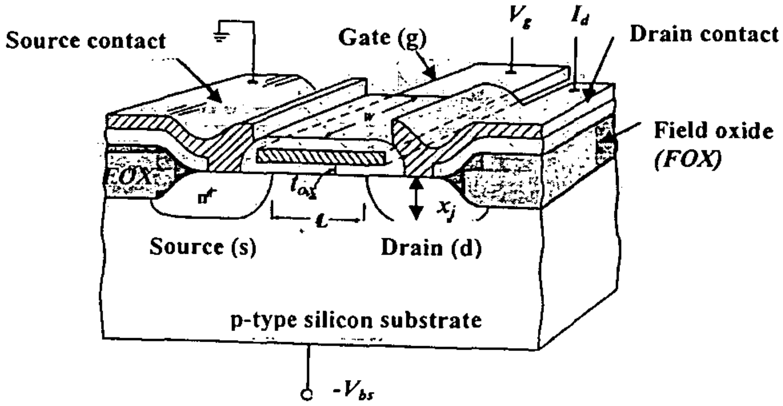


Figure 1.1: Basic structure of a MOSFET (Reproduced from Arora, 1993)

When no voltage is applied to the gate, the source-to-drain electrodes correspond to two p-n junctions connected back-to-back. The only current that can flow from source to drain is the reverse leakage current (this is the n-channel enhancement type MOSFET-which is normally off). When a sufficiently large positive bias is applied to the gate so that a surface inversion layer (or channel) is formed between the two n^+ regions, the source and the drain are then connected by a conducting-surface n-channel through which a large current can flow. The conductance of this channel can be modulated by varying the gate voltage. The back-surface contact (or substrate contact) can have the reference voltage or be reverse-biased; the back-surface voltage will also affect the channel conductance.

For a PMOS, the substrate is an n-type semiconductor, and the source/drain are heavily p-type doped. Negative gate bias is used to generate negative charges (electrons) on a metal-oxide surface, which in turn drives away electrons (majority carriers of substrate) and attracts holes (minority carriers) to the silicon-oxide surface to form a hole-channel underneath the gate. Holes from source and drain can then flow through the channel and conduct electric current between source and drain. The PMOS transistor is turns off when a positive bias is applied.

In normal operation of a NMOS, a positive voltage is applied between source and drain (V_d). No current flows between source and drain $(I_d=0)$ with gate voltage, $V_g=0$ because of back to back p-n junctions. When $V_g>V_{th}$, electric field attracts electrons, thus creating a channel. Channel is a p-type silicon which is inverted to n-type by the electrons attracted by the electric field. When V_d is non-zero, the channel becomes smaller closer to the drain.

As illustrated in Figure 1.2, with a small V_d where $V_d \leq V_g - V_{th}$, current flows from drain to source. This is termed as a *linear region*. I_d is a function of both V_g and V_d and the transistor acts as a voltage-controlled resistor.

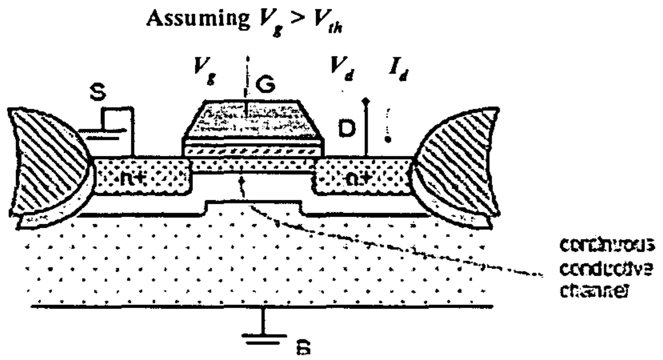


Figure 1.2: NMOS transistor in a linear region (Reproduced from Brown University, 2005)

When $V_d > V_g - V_{th}$, the channel is pinched off close to the drain since $V_g > V_{th}$. This mode of operation is shown clearly in Figure 1.3. Number of carriers arriving at pinch-off point from source remains the same which means that current flowing from drain to source remains constant. This is termed as a *saturated region*. I_d is a function of V_g, almost independent of V_d and the transistor now acts a voltage-controlled current source.

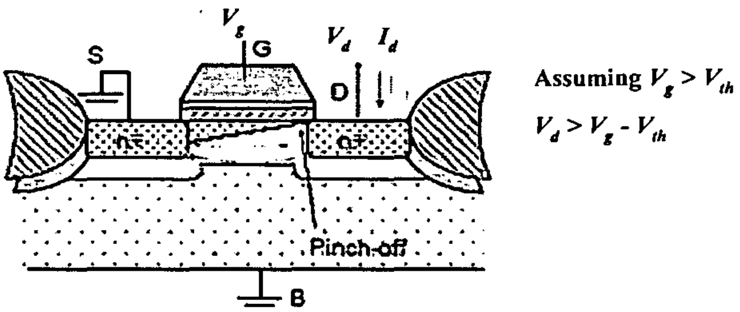


Figure 1.3: NMOS transistor in a saturation region (Reproduced from Brown University, 2005)

Figure 1.4 illustrates the basic output characteristics of an idealized MOSFET. For a given V_g, the drain current first increases linearly with drain voltage (the linear region), then gradually levels off, approaching a saturated value (the saturation region). The dashed line indicates the locus of the saturation drain voltage ($V_{d(sat)}$). For $V_d > V_{d(sat)}$, the drain current remains constant.

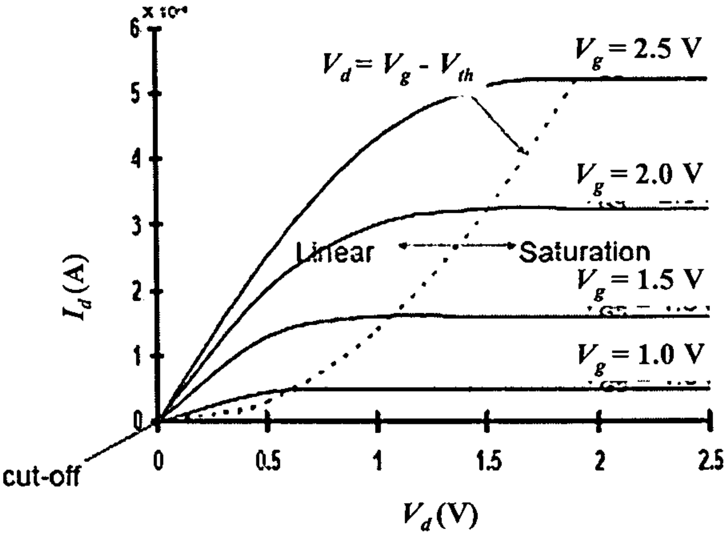


Figure 1.4: Ideal characteristics of I_d vs. V_d of a MOSFET (Reproduced from Brown University, 2005)

1.1.2 Complementary-Metal-Oxide-Semiconductor (CMOS)

With the evolution of the VLSI era, power consumption in NMOS circuits began to exceed tolerable limits. This leads to the introduction of CMOS technology which makes use of both n-type and p-type MOS transistors in its structure. Figure 1.5 illustrates an example of a CMOS structure fabricated using twin-well technology.

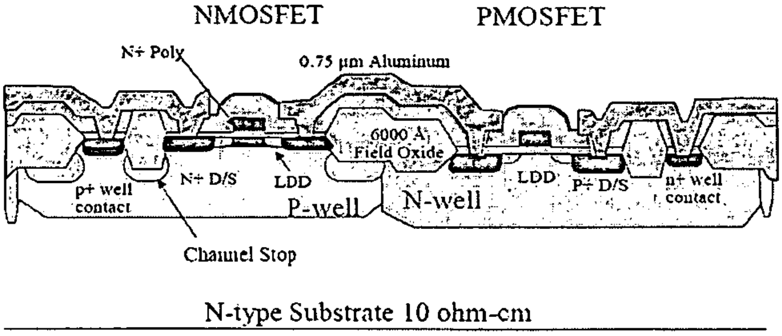


Figure 1.5: Structure of CMOS (Reproduced from RIT, 2005)

CMOS circuits which has a low power dissipation characteristics offered a solution to the power dissipation problem. In a CMOS circuits, only one of the two transistors is driven at any one time. When a CMOS inverter, as shown in Figure 1.6 is not switching from one state to the other, virtually no current flows and almost no DC (direct current) power is dissipated as a high-impedance path exists from the supply voltage to ground, regardless of the state the inverter is in. Besides the favorable low-power dissipation characteristics, CMOS circuits also switches reasonably fast and has a high packing density on a silicon real estate, making it the dominant Integrated Circuit (IC) technology at present (Asai & Wada, 1997).

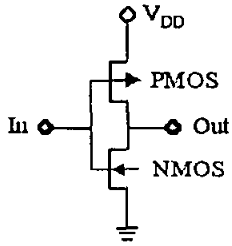


Figure 1.6: Symbol of CMOS inverter

1.1.3 Scaled-down Devices

During the last four decade, the integrated circuit complexity has moved from the early Small-Scale Integration (SSI), to Medium-Scale Integration (MSI), to Large-Scale Integration (LSI), to Very-Large-Scale Integration (VLSI) and finally to Ultra-Large-Scale Integration (ULSI) that can accommodate million of transistors on a single chip. It is projected that by year 2010, there will be about 10 billion transistor components on an Integrated System (Stanford Univ., 2005). This increase in the integrated circuit density has been primarily attributed to the reduction in the size of elementary features of the underlying structures which is the transistor itself. Enhanced performance resulting from the scaling-down/downsizing of MOSFET's includes faster switching speed and lower power dissipation. Nowadays, aggressive downscaling is to be performed on certain MOS technology parameter such as gate oxide thickness which is to be less than 10 angstrom, channel length reduced to about ~20 nm and reduction of junction depth to about ~1-2 nm (Stanford Univ., 2005). However, when scaling the gate oxide thickness below 20 angstrom, problems such as gate leakage, dielectric breakdown and dopant penetration through gate oxide will start to arise as depicted in Figure 1.7 (Stanford Univ., 2005).

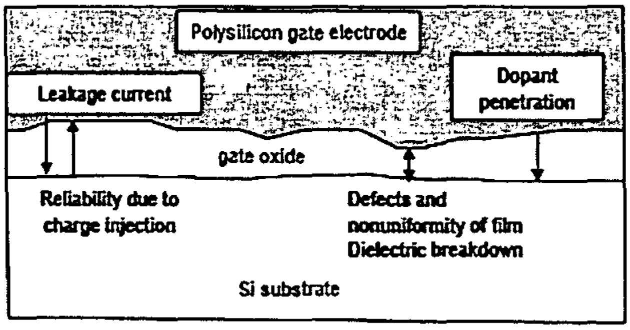


Figure 1.7: Illustration on gate leakage, dielectric breakdown and dopant penetration through gate oxide (Reproduced from Stanford Univ., 2005)

According to Bohr (2003), the limits to device scaling will ultimately be the leakage current. As the transistor gate length, L_g is reduced, there will be quantum mechanical tunneling through barrier. The total leakage trend is depicted in Figure 1.8.

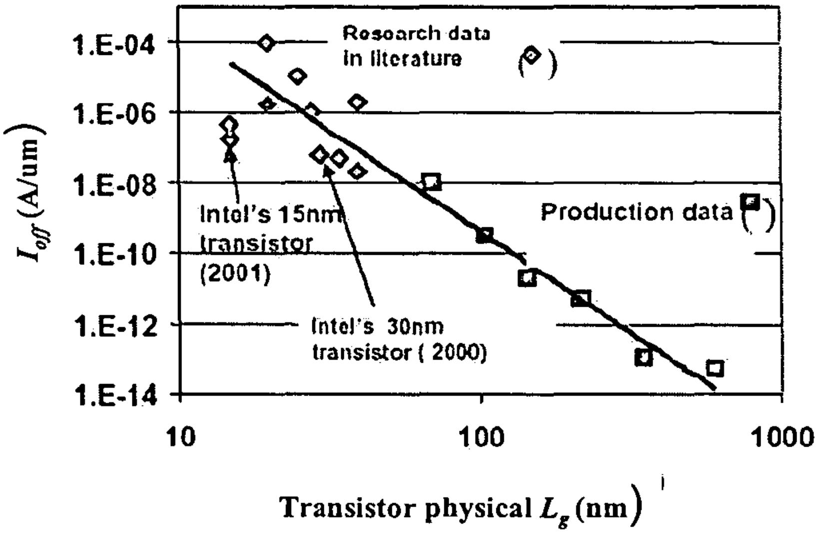


Figure 1.8: Total leakage trend (Reproduced from Bohr, 2003)

However, it is expected that the problems of leakage current will not be the limit in this work as the scale of the final device design is of 0.25 μm. This is as according to Taur (2002), increasing leakage currents will only be apparent for CMOS devices below 0.1 μm.

1.2 Objective

The importance of MOS transistors which are the backbone of modern microelectronics has led to the need of a thorough understanding of how these transistors are being materialized. An in-depth knowledge on the MOS fabrication processes and its electrical characteristics is needed as currently, this knowledge which is strictly treated as Intellectual Properties (IP) is only available to the industries. Confidentiality has made it difficult for academic institutions to obtain this knowledge, and thus this research work serves as a platform to explore this area. The steady downscaling of CMOS device dimensions has also been known as the main stimulus to the growth of the microelectronics and the computer industry over the past two decades. Coupling the need of an in-depth knowledge of CMOS technology at academic level together with miniaturization trend at industrial level, has driven the research effort towards achieving the following objective.

The objective of this work is to design a downscale device of 0.25 μm CMOS from physical data of 0.35 μm technology and characterize its electrical properties by means of simulation tools. The work focused on understanding the fabrication processes involved, integrating the physical data to simulation environment and studying the feasibility of using a 0.35 μm CMOS technology to produce a scale down device of 0.25 μm.

1.3 Scope of Research

In the absence of industrial-scale fabrication facilities, this study is made possible through the use of numerical simulators (commonly known as Technology Computer Aided Design (TCAD) Tools) which will enable users to have better understanding of the deep sub-micrometer structures by providing the internal view.

In this work, the 0.25 μm CMOS transistor design is confined to a 2D-process and device simulations using *Silvaco ATHENA* and *ATLAS* simulation software version. Both simulators are the leading TCAD tools used in the microelectronic industry. *ATHENA* is a process simulator used to simulate various semiconductor fabrication processes while *ATLAS* is used to perform device characterization. In the process of downscaling the CMOS transistor from 0.35 μm to 0.25 μm, a "two-step design" approach is proposed which consist of the following. The first design step involves conversion of the physical data of 0.35 μm CMOS technology to simulation environment. Process recipe acquired from UC Berkeley Microfabrication Lab (Horvath *et al.*, 2005) serves as the design basis. Upon successful conversion, the electrical parameters of the 0.35 μm technology are extracted. The second design step involves miniaturization of the 0.35 μm technology to 0.25 μm through downscaling. This is followed by the extraction of the electrical parameters to evaluate its performance.

1.4 Organization of the Thesis

This thesis is organized into six chapters. Chapter 1 provides a brief introduction on the background of the research, which includes the objective, scope and the outline of the thesis. Chapter 2 introduced the concept of device miniaturization which covers the guidelines for transistor scaling and the short-channel effects that arises as the adverse results. Besides descriptions on the MOS key features such as Local-Oxidation-of-Silicon (LOCOS), Lightly-doped drain (LDD) and Self-aligned silicide (SALICIDE) structure, processes such as oxidation, diffusion, ion implantation and rapid thermal annealing (RTA) which made up the basis of MOS transistor formation are also included in this chapter. Chapter 3 describes the important features of the simulation software used in this research work. Guidelines on the determinations and selections of the physical models to be used in each stage of simulation are also reported. Chapter 4 provides an overview of the design approach adopted in the process of downscaling of the 0.35 μm to 0.25 μm CMOS transistor. Included are the summaries of models and parameters used throughout the process and device simulation work. Chapter 5 discusses the results obtained from the simulations of both the 0.35 μm and 0.25 μm CMOS transistor. Evaluation on the electrical performances are made by referring to the results extracted for values of threshold voltage (V_{th}), saturation current (I_{dsat}), off-state leakage current (I_{off}) and subthreshold swing (S_t). The final chapter which is Chapter 6 presents the summary and conclusion of the research work, together with recommendations for future studies.

CHAPTER 2

DEVICE PROCESS TECHNOLOGY AND TRENDS

This chapter will first describe the concept of scaling-down of devices and guidelines used for scaling purposes. Adverse effects known as short-channel effects, resulting from the miniaturization process are also highlighted. Descriptions on processes such as oxidation, diffusion, ion implantation and rapid thermal annealing (RTA) are also explained, followed by descriptions on formation of certain key CMOS technology such as Local-Oxidation-of-Silicon (LOCOS), Lightly-doped drain (LDD) and Self-aligned silicide (SALICIDE) structure.

2.1 Deep-submicron Scaling

2.1.1 Introduction

To allow more devices on a chip, technologies are scaled every few years (SIA, 1999). Since 1965, integrated circuit technology has followed Moore's Law as shown in Figure 2.1, which states that the number of devices integrated double every 18 months. This growth is made possible by continuous miniaturization in feature size of components devices which are integrated, a concept known as device scaling.

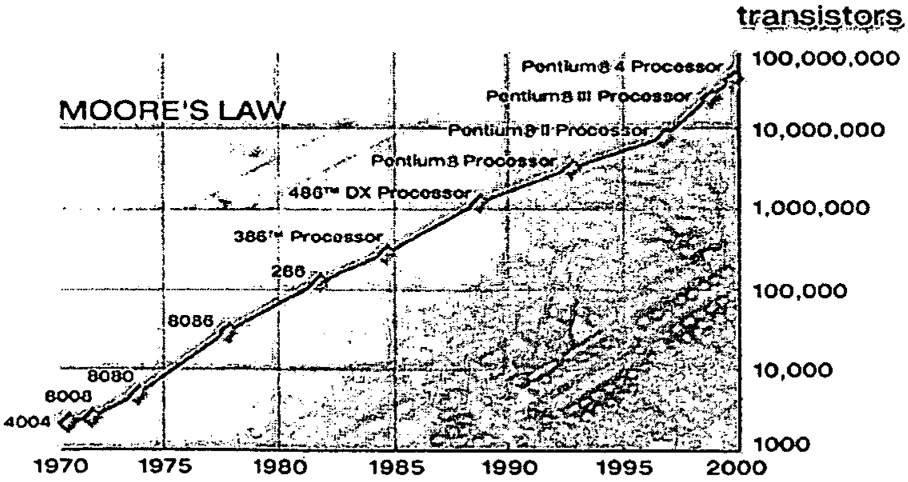
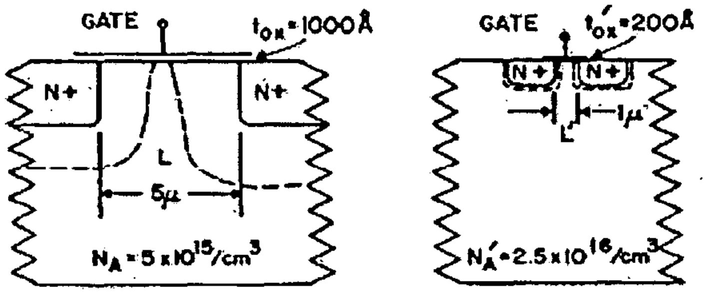


Figure 2.1: Moore's Law (Reproduced from Intel Corporation)

In detail, device scaling refers to systematic reduction of various structural parameters of a MOSFET to ensure the device continue to function properly. These include lateral as well as vertical dimensions such as the channel length, the width, the source/drain junction depth x_j and the gate oxide thickness. For proper device scaling, power supply voltages should also be reduced to keep the internal field constant. Depending on the variable, the parameter could be multiplied, or divided by κ which is a unitless scaling factor. Reduction of feature size tends to bring advantages of increased speed and an improved density (smaller areas for devices and circuits). Figure 2.2 offers a conceptually simple picture of device miniaturization.



(a) (b)

Figure 2.2: Illustration of device scaling principles with $\kappa = 5$ (a) Conventionally available device structure (b) Scaled-down device (Reproduced from Dennard *et al.*, 1974)

2.1.2 Scaling Guidelines

The first complete scaling scheme was introduced by Dennard *et al.* in 1974 and is regarded as the seminal reference in scaling theory for MOSFET integrated circuits. There are three major scaling approaches: *constant-field scaling (also known as full scaling)*: reduces both sizes and voltages to maintain constant relative electric fields, *constant-voltage scaling*: keeps voltages fixed as dimensions are scaled and *general scaling*: a mixture of constant-field and constant-voltage scaling trends. All scaling methods are attempts to reproduce long channel behavior in a short channel device by keeping the proportions between physical and electrical characteristics of the device constant, thereby avoiding short channel and non-ideal effects.

2.1.2 (a) Constant-field scaling (Full scaling)

The principle of first order scaling of MOSFET dimensions which maintains the maximum field strength criterion is the "constant-field scaling" theory from Dennard *et al.* (1974). This scaling options attempts to preserve the magnitude of internal electric fields in the MOSFET, while the dimensions are scaled down by a factor of κ. The primed parameters refer to the new scaled-down device. Electric field across gate oxide, E_{ox} is therefore given by :

$$E_{ox} = \frac{V_g}{t_{ox}} \tag{2.1}$$

$$E_{ox}' = \frac{\left(V_g/k\right)}{\left(t_{ox}/k\right)} \tag{2.2}$$

Divide Equation (2.2) with (2.1)

$$\frac{E_{ox}'}{E_{ox}} = \frac{\left(V_g/k\right)}{\left(t_{ox}/k\right)} \div \frac{V_g}{t_{ox}} \tag{2.3}$$

$$\frac{E_{ox}'}{E_{ox}} = 1 \tag{2.4}$$

$$E_{ox}' = E_{ox} \tag{2.5}$$

In addition to scaling of dimensions, all potentials must also be scaled down proportionally by the same scaling factor to maintain a constant electric field which will therefore affect the V_{th}. Table 2.1 summarizes the concepts of constant-field scaling (Critchlow, 1999, Dennard *et al.*, 1974 and Slisher *et al.*, 1999).

Table 2.1: Constant-field scaling (Full scaling)

Parameter	Before Scaling	After Scaling
Gate length, L_g	L_g	$L_g' = L_g / \kappa$
Gate width, W	W	$W' = W / \kappa$
Gate oxide thickness, t_{ox}	t_{ox}	$t_{ox}' = t_{ox} / \kappa$
Junction depth, x_j	x_j	$x_j' = x_j / \kappa$
Power supply voltage, V_{dd}	V_{dd}	$V_{dd}' = V_{dd} / \kappa$
Threshold voltage, V_{th}	V_{th}	$V_{th}' = V_{th} / \kappa$
Field, E	E	E' = Remain unchanged
Doping densities, N_A N_D	N_A N_D	$N_A' = \kappa.N_A$ $N_D' = \kappa.N_D$
Gate oxide capacitance/area, C_{ox}	C_{ox}	$C_{ox}' = \kappa.\ C_{ox}$
Current, I_{dsat}	I_{dsat}	$I_{dsat}' = I_{dsat}/\kappa$
Gate delay, $\tau = CV/I$	τ	$\tau' = \tau / \kappa$
Power dissipation/circuit, $P=VI$	P	$P' = P / \kappa^2$
Power density, P/A	$P / Area$	$P' / Area'$ = Remain unchanged

One of the most attractive features of constant-field scaling is the significant reduction of the power dissipation. However, the full scaling strategy may not be very practical in many cases when it comes to the scaling of the power supply voltage and all terminal voltages in proportion with the device dimensions. This is as the peripheral and interface circuitry may require certain voltage levels for all input and output voltages. The need to scale down these voltages in turn would necessitate multiple power supply voltages and complicated level-shifter arrangements. Thus, this method has not been adopted in its original form. While actual technologies have not adhered strictly to constant-field scaling (Davari, 1996), it is illustrative of the general trends and problems associated with scaling. Different scaling scheme were soon to follow, such as *constant-voltage scaling*.

2.1.2 (b) Constant-voltage scaling

In constant-voltage scaling, all dimensions of the MOSFET are reduced by a factor of κ as in full scaling. The power supply voltage and the terminal voltages, on the other hand, remain unchanged. Electric field across gate oxide E_{ox} is therefore given by:

$$E_{ox} = \frac{V_g}{t_{ox}} \tag{2.6}$$

$$E_{ox}' = \frac{V_g}{\left(t_{ox}/k\right)} \tag{2.7}$$

Divide Equation (2.7) with (2.6)

$$\frac{E_{ox}'}{E_{ox}} = \frac{V_g}{\left(t_{ox}/k\right)} \div \frac{V_g}{t_{ox}} \tag{2.8}$$

$$\frac{E_{ox}'}{E_{ox}} = k \tag{2.9}$$

$$E_{ox}' = k.E_{ox} \tag{2.10}$$

Table 2.2 summarizes the concept of constant-voltage scaling (Critchlow, 1999, Dennard *et al.*, 1974 and Slisher *et al.*, 1999) where the primed parameters refer to the new scaled-down device. If constant-voltage scaling is performed, L_g and W is reduced by the same scaling factor κ. As the drain-source voltage remains unchanged, the lateral field increases. It increases approximately by a factor κ since $E \propto V_d/L_g$. While this solves the circuit incompatibility problems, it leads to higher electric fields inside the device which will cause dielectric breakdown and other reliability problems. This leads us to another type of scaling which is of *general scaling.*

Table 2.2: Constant-voltage scaling

Parameter	Before Scaling	After Scaling
Gate length, L_g	L_g	$L_g' = L_g / \kappa$
Gate width, W	W	$W' = W / \kappa$
Gate oxide thickness, t_{ox}	t_{ox}	$t_{ox}' = t_{ox} / \kappa$
Junction depth, x_j	x_j	$x_j' = x_j / \kappa$
Power supply voltage, V_{dd}	V_{dd}	V_{dd}'= Remain unchanged
Threshold voltage, V_{th}	V_{th}	V_{th}' = Remain unchanged
Field, E	E	$E' = \kappa.E$
Doping densities, N_A N_D	N_A N_D	$N_A' = \kappa^2.N_A$ $N_D' = \kappa^2.N_D$
Gate oxide capacitance/area, C_{ox}	C_{ox}	$C_{ox}' = \kappa.C_{ox}$
Current, I_{dsat}	I_{dsat}	$I_{dsat}' = \kappa\ I_{dsat}$
Gate delay, $\tau = CV/I$	τ	$\tau' = \tau / \kappa^2$
Power dissipation/circuit, $P=VI$	P	$P' = \kappa.\ P$
Power density, P/A	$P / Area$	$P' / Area' = \kappa^3$

2.1.2 (c) General scaling

In this method, voltages and dimensions are scaled with different factors (suggest a mixture of constant-field and constant-voltage scaling trends). Dimensions are scaled by a scaling factor κ while voltages are scaled by a scaling factor U. Supply voltage is being scaled but at a slower rate than feature size. Table 2.3 summarizes the concept of *general scaling* (UMBC, 2004) where the primed parameters refer to the new scaled down device.

Table 2.3: General scaling

Parameter	Before Scaling	After Scaling
Gate length, L_g	L_g	$L_g' = L_g / \kappa$
Gate width, W	W	$W' = W / \kappa$
Gate oxide thickness, t_{ox}	t_{ox}	$t_{ox}' = t_{ox} / \kappa$
Power supply voltage, V_{dd}	V_{dd}	$V_{dd}' = V_{dd} / U$
Threshold voltage, V_{th}	V_{th}	$V_{th}' = V_{th} / \mathrm{U}$
Doping densities, N_A N_D	N_A N_D	$N_A' = (\kappa^2 / U).N_A$ $N_D' = (\kappa^2 / U).N_D$
Gate oxide capacitance/area, C_{ox}	C_{ox}	$C_{ox}' = \kappa.C_{ox}$

This method is a compromise between the two scaling methods described previously and is the most realistic scaling method to be used in today's situation.

2.1.3 Short-Channel Effects (SCE)

As the gate length is reduced, characteristics of a MOSFET change due to short-channel effects. These effects arise as a result of having very short gate (and channel) lengths. The departure of the characteristics of a MOSFET from long-channel behavior arises as a result of a two-dimensional potential distribution and high electric fields in the channel region. For a given channel doping concentrations, as the channel length is reduced, the depletion-layer widths of the source and drain junctions become comparable to the channel length. The potential distribution in the channel now depends on both the transverse field, E_x (controlled by the gate voltage and the back-surface bias) and the longitudinal field, E_y (controlled by the drain bias). A number of physical effects that are unimportant in larger MOSFET can now become relevant in smaller transistor due to this two-dimensional potential result such as degradation of the subthreshold behavior, failure of current saturation due to punch-through and dependence of the threshold voltage on channel length and biasing voltages (Sze, 1981).

2.1.3 (a) Degradation of the subthreshold behavior

In long-channel devices, the subthreshold drain current is independent of drain voltage. In small-geometry MOSFETs however, the potential barrier is now controlled by both the gate voltage, V_g and the drain voltage, V_d. If the V_d is increased, the potential barrier in the channel decreases, leading to Drain-induced barrier lowering *(DIBL)*. This reduction of potential barrier will eventually allow electron to flow between the source and the drain, even if the gate-to-source voltage is lower than the V_{th}. This in turn causes excess injection of charge carriers into the channel which leads to lower V_{th} and an increase in the off-currents (Chua & Liu, 1993). The channel current that flows under these conditions $(V_g < V_{th})$ is called the subthreshold current. Figure 2.3 shows that for long-channel, I_d is independent of V_d. For short-channel (where L_g>180 nm), there is a parallel shift of I_d while for ultra-short-channel (where L_g>90 nm)) the subthreshold slope increased and eventually punch-through will occur.

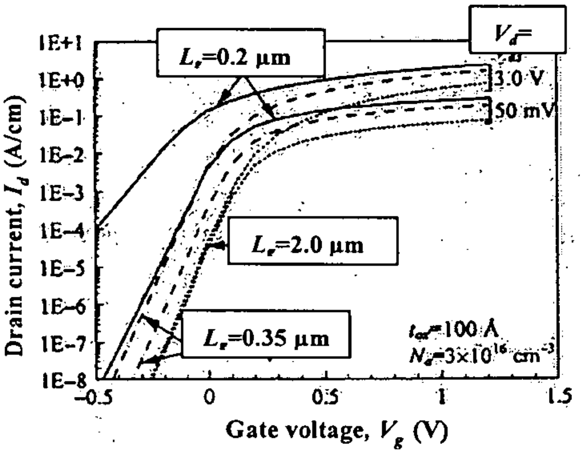


Figure 2.3: Graph of drain current, I_d versus gate voltage, V_g
(Reproduced from KNU, 2002)

2.1.3 (b) Failure of current saturation due to punch through

In short-channel devices, a separation between depletion regions at the drain-substrate and the source-substrate junctions is relatively small. For large drain-bias voltages, the depletion region surrounding the drain can extend farther toward the source, and the two-depletion regions can eventually merge as shown in Figure 2.4. This condition is termed punch-through where majority carriers in the source region can be injected into the depleted channel region, where they will be swept by the field and collected at the drain. The V_g loses its control upon the I_d, and the current rises sharply once punch-through occurs (Sze, 1981). Figure 2.5 (a) shows the drain characteristics of a short-channel MOSFET operated in punch through conditions, where we can see that there is no true saturation while Figure 2.5 (b) shows the drain characteristics of a long channel MOSFET

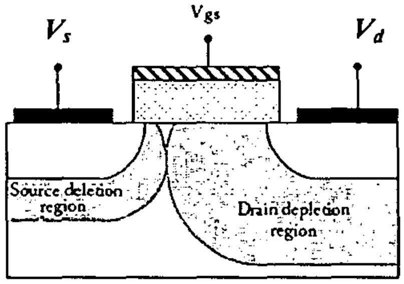
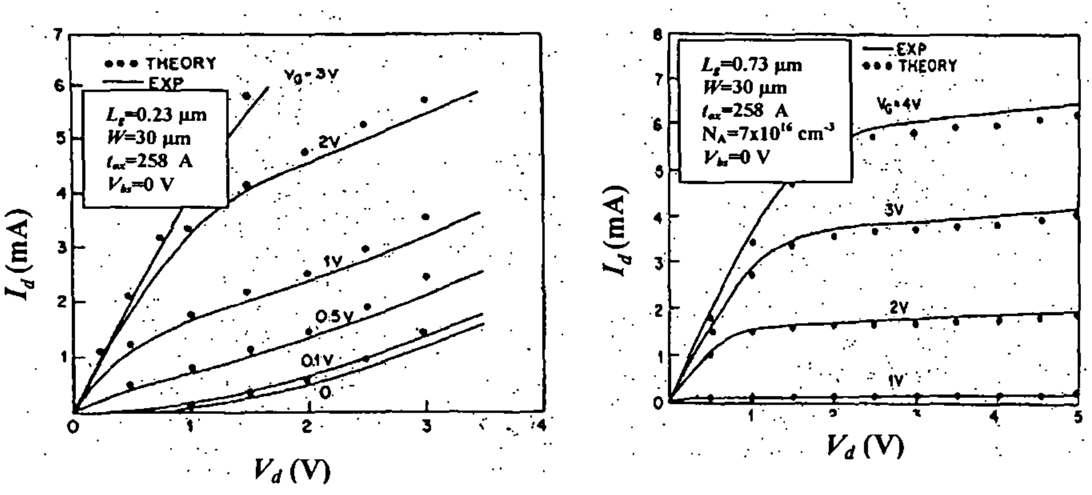


Figure 2.4: Mechanism of punch through. For increasing V_d, L_{eff} decreases causing an increase in I_d (Reproduced from Slisher *et al.*, 1999)



(a) (b)

Figure 2.5: Illustrate the I_d-V_d curve for (a) short-channel devices with 0.23 µm gate length versus (b) long channel devices with 0.73 µm gate length (Reproduced from Sze, 1981)

2.1.3 (c) Dependence of the threshold voltage on channel length and biasing voltages

The most important parameter to model the operation of a MOSFET is its V_{th}. It is defined as the V_g at which significant current starts to flow from the source to the drain and is a measure of control of the gate electrode to the channel region of MOSFET. The lower the V_{th}, the higher the control capability that the gate can provide and hence higher I_d (NUS, 2003).

There are several definitions of V_{th} and many methods have been developed over time to represent this parameter (Liou *et al.*, 1997). The majority of procedures used to determine the V_{th} of MOSFETs carry out the extraction from the strong inversion operation characteristics. The most common method used are *constant-current-V_{th} method* - defining V_{th} as the gate voltage corresponding to a certain predefined practical constant drain current and *linear extrapolation method* - finding the gate voltage axis intercept of the linear extrapolation of the I_d-V_g characteristics at its maximum first derivative (slope) point.

The expression of V_{th} is given as (Lim & Xing, 1998):

$$V_{th} = 2\phi_b + \frac{\sqrt{2\varepsilon_{si} q N_A 2\phi_b}}{C_{ox}} + V_{fb} \tag{2.11}$$

where $\phi_b = \frac{kT}{q}\ln\left(\frac{N_A}{N_i}\right)$

N_A - density of the carriers in the doped semiconductor substrate

N_i - the carrier concentration of intrinsic (undoped) silicon

$N_i = 1.45 \times 10^{10} cm^{-3}$ (at 300 K)

k is the Boltzman's constant, T is temperature, q is electronic charge of carriers

$\frac{kT}{q} = 25mV$ (at 300 K)

ε_{si} - permittivity of silicon

$\varepsilon_{si} = 1.06 \times 10^{-12}$ Farads/cm

C_{ox} - gate-oxide capacitance

$$C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$$

where ε_{ox} is the permittivity of gate oxide layer and t_{ox} is the gate oxide thickness

$$V_{fb} = \phi_{ms} - \frac{Q_{fc}}{C_{ox}}$$

Q_{fc} represent the fixed charge due to imperfections in silicon-oxide interface and doping; and

ϕ_{ms} is work function difference between gate material, ϕ_{gate} and silicon substrate, ϕ_{Si}

For submicron device, its V_{th} varies with the effective channel length. This is because as the channel length becomes shorter, the depletion charge Q_B under the gate is not fully controlled by the gate bias alone, but also modulated by the source and drain bias. Hence, for short channel MOSFET, the modulation effect for Q_B is not only from the vertical direction (one-dimensional effect) but from both lateral and vertical directions. This effect is also known as two-dimensional effect (Lim & Xing, 1998). The V_{th} generally decreased as the gate length is decreased (known as V_{th} roll-off). An increase in V_d will also cause a decrease in V_{th} (Slisher *et al.*, 1997).

In order to predict the V_{th} of a short-channel device, the shift in the threshold voltage, ΔV_{th} must be approximated (Wolf, 1995). The short-channel effects V_{th} $((V_{th0})_{SCE})$ is given by the following formula where V_{th0} is the long channel V_{th} (Uyemura, 1992).

$$(V_{th0})_{SCE} = V_{th0} - (\Delta V_{th}) \tag{2.12}$$

A method of calculating ΔV_{th} is by using the charge-sharing model. This model assumes that the charge under the gate is shared between the source/drain depletion regions and gate inversion region. Therefore, less voltage is required to invert the channel. ΔV_{th} increases as the depletion region's length approaches the gate length. A simple model to predict ΔV_{th} is proposed by Yau (1974). The model used the assumption that the charge caused from the gate is simply the trapezoidal region under the gate. The following analytical formula to calculate ΔV_{th} can be used for uniformly doped channels (Wolf, 1995).

$$\Delta V_{th} = \frac{q.N_{sub}.d_{\max}.r_j(\sqrt{\left\{1+\frac{2d_{\max}}{r_j}\right\}}-1)}{C_{ox}.L} \tag{2.13}$$

where d_{max} is the maximum width of the depletion region under the gate, and r_j is the length of the depletion region of the source/drain. The model shows that decreasing the gate oxide thickness as shown in Equation 2.11 and decreasing the depletion regions of the source/drain will decrease ΔV_{th}. The charge-sharing model can be used to understand the concept of decreasing ΔV_{th}, but does not predict the change in the V_{th} accurately against experiment data, especially for narrow gate lengths and high V_d. The model does provide a first order approximation (Wolf, 1995).

It is noted that as L_g gets smaller for the scaled technology generations, the short-channel effects become more acute. As short-channel effects degrade device performance, it should therefore be eliminated or minimized so that a physical short-channel device can still exhibits the "electrical" long-channel behavior. In order to control short-channel effects, state-of-the-art MOSFET devices usually utilize source/drain-engineering and channel-engineering structures in the device design. In source/drain-engineering, a source-drain extension (SDE) which refers to shallow diffusion that connects the channel with the deep source and drain is usually employed. The lightly-doped drain (LDD) structure is an example of SDE which is formed by first etching the gate, followed by ion implantation to form the shallow extension. After the SDE implant, a spacer will be added attached to the gate. The spacer's purpose is to block the higher dose source/drain implants. At deep-submicron dimensions (≤ 0.35 μm), care should be taken on the n^- drain region depth and its corresponding profile. This is because as compensation is being made to the p-type doping in the short-channel region by the n^- region profile, serious charge-sharing problems could arise unless the n^- drain region is shallow and steeply profiled (Tasch *et al.*, 1990). SDE should be made relatively shallow compared to the source/drain implants to improve device short-channel characteristics as the amount of channel charge controlled by the drain is now reduced (Thompson *et al.*, 1998).

Another technique to improve short-channel characteristics is through channel-engineering. The purpose of channel engineering is to optimize the channel profile in order to minimize the off-state leakage while maximizing the linear and saturated drive currents. This is done by changing the doping profile in the channel region which will cause changes in the distribution of the electric field and potential contours. Super Steep Retrograde Wells (SSRW) and halo implants are examples of channel engineering. Retrograde well engineering changes the characteristics of the well profile by creating a retrograde profile toward the Si/SiO_2 surface. Such a design is achieved by using high energy ion implantation to place dopants at a desired substrate depth and then annealing at a low temperature to activate the implants. This leads to better control of punch through since the doping concentration at the bottom of the well is higher (Wolf, 1995). Halo architecture which refers to halo implants performed in vertical or angled after gate patterning on the other hand, creates a localized dopant distribution near the source-drain extension regions as shown in Figure 2.6. They add additional well dopant around the source and drain regions providing an increased source-to-drain barrier for current flow. Even though halo implant can help reduce short-channel effects such as punch through and V_{th} roll-off, they also tend to degrade the transistor drive current. Thus, trade-off between drive current and the reduction of short-channel effects must be considered to maximize performance of the transistor (Hwang *et al.*, 1996)

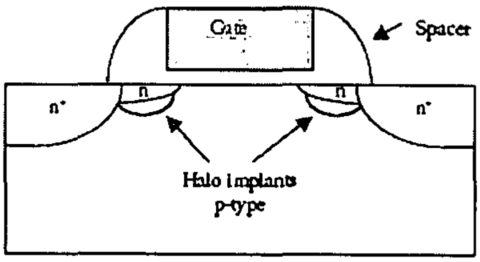


Figure 2.6: Halo implant for an n-channel MOSFET
(Reproduced from Slisher *et al.*, 1999)

2.2 Theory of Processes

2.2.1 Oxidation

The most frequently used method to form silicon dioxide films is by thermal oxidation of silicon through the chemical reaction. For short reaction times, the oxide thickness increases linearly with time, and for prolonged oxidation, the thickness varies as the square root of time. While the silicon dioxide is growing, it consumes the substrate silicon and the film grows into the silicon substrate. For silicon dioxide film of thickness *d*, a layer of silicon of thickness 0.45*d* is consumed as shown in Figure 2.7.

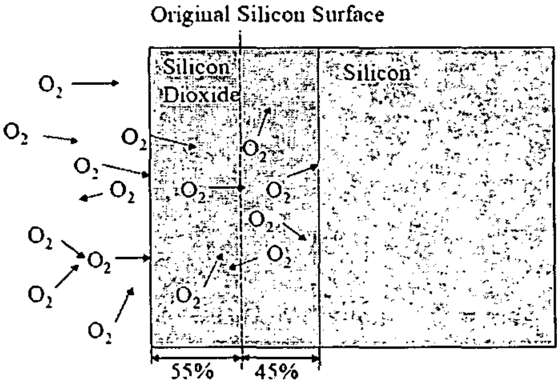


Figure 2.7: Silicon oxidation process (Reproduced from Xiao, 2001)

Oxide growth rate is determined by many factors such as (Xiao, 2001):

(i) oxidation temperature

Oxide growth rate is very sensitive to temperature because the oxygen diffusion rate in silicon dioxide is exponentially related to temperature, $D \propto exp\ (E_a / kT)$. Here, D is the diffusion coefficient, E_a is the activation energy, $k = 2.38 \times 10^{-23}\ J/K$ is the Boltzmann constant and T is the temperature.

(ii) oxygen source (O_2 or H_2O)

Oxide growth rate is also related to the oxygen source. Dry oxidation with O_2 has a lower oxide growth rate than wet oxidation with H_2O. This is because the diffusion rate of the oxygen molecule, O_2 in silicon dioxide is lower than that of the hydroxide, HO generated from the disassociation of H_2O molecules at high temperature.

(iii) silicon-crystal orientation

Oxidation growth rate is also related to the single-crystal silicon orientation. Normally, <111> orientation silicon has a higher oxide growth rate than <100> orientation silicon. This is because the <111> silicon surface has higher silicon atoms density than that of the <100> silicon surface, and can provide more silicon atoms to react with oxygen and form a thicker silicon dioxide layer.

(iv) dopant type and concentration

Oxidation rate is also related to dopant and dopant concentration. Generally, heavily doped silicon oxidizes faster than lightly doped silicon. During oxidation, boron in the silicon tends to be drawn up to the silicon dioxide and cause depletion of the boron concentration at the silicon-silicon dioxide interface. N-type dopants such as phosphorous, arsenic and antimony have the opposite effect (while oxide grows into the silicon, these dopants are driven deeper into the silicon). The n-type dopant concentration in the silicon-silicon dioxide interface could be significantly higher than its original value. Figure 2.8 illustrates the pile-up effect of n-type dopant and the depletion effect of p-type dopant.

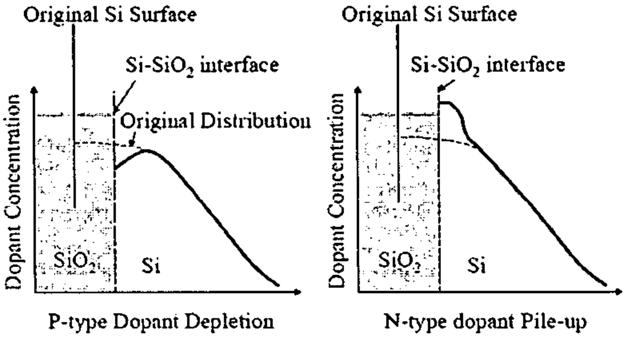


Figure 2.8: Oxidation-caused dopant depletion and pile-up effects (Reproduced from Xiao, 2001)

(v) gas

Oxidation rate is also related to additional gas such as HCl, which is commonly used in the gate oxidation process to suppress mobile ions. With the appearance of HCl, the oxidation rate can increase by about 10%.

2.2.2 Diffusion

Diffusion is the physical phenomenon of material moving from high concentration regions to low concentration regions, driven by thermal motion of the molecules. Since the diffusion rate in solid materials is exponentially related to temperature, $D \propto exp(-E_a/kT)$, the diffusion process can be accelerated significantly at a high temperature.

The idea of particle movement created by non-uniformities can be expressed mathematically in the form of Fick's First Law, which relates particle flux J to concentration gradients:

$$J = -D\frac{\partial C(x,t)}{\partial x} \tag{2.14}$$

where C is the concentration of the solute, D is the diffusion coefficient, x is the distance into the substrate and t is the diffusion time. The negative sign indicates that the diffusing mass flows in the direction of decreasing concentration.

From the continuity relation, we know that:

$$\frac{\partial C}{\partial t} = -\frac{\partial J}{\partial x} \tag{2.15}$$

By inserting Fick's first equation into the continuity relation, we will obtain Fick's Second Law of Diffusion which states:

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \tag{2.16}$$

In order to solve Fick's Law, one initial condition and two boundary conditions are required. Two solutions to Fick's Law are generally encountered in Integrated circuit fabrication: constant-source and constant-dose diffusion. Both solutions are described below.

(a) Constant-source diffusion (Pre-deposition)

In constant-source diffusion, a source of the dopant is applied to the surface and we assume that this has the effect of holding the surface concentration at some high, constant value. This generally corresponds to the process step known as "pre-deposition". In this case, the initial condition and boundary conditions are:

$$\begin{aligned} &C(x,0) = 0 \\ &C(0,t) = C_s \\ &C(\infty,t) = 0 \end{aligned} \qquad (2.17)$$

where C_s is the surface concentration. The solution of Fick's Law under these conditions is:

$$C(x,t) = C_s erfc\left[\frac{x}{2\sqrt{Dt}}\right] \qquad (2.18)$$

where "*erfc*" is the complementary error function and $\sqrt{Dt}$ is the characteristic length.

(b) Constant-dose diffusion / Gaussian type diffusion (Drive-in)

A second important diffusion method is the constant-dose diffusion. As the name implies, a constant dose of dopant is introduced into the semiconductor from some source. The source is removed and then diffusion proceeds with a fixed amount of dopant available. Generally, we assume that initial dose is located exactly at the surface of the wafer. This corresponds to the process step known as "drive-in" or any subsequent heat cycles. In this case, the initial condition and boundary conditions are:

$$\begin{aligned} &C(x,0) = 0 \\ &C(\infty,t) = 0 \\ &\int C(x,t)dx = S \end{aligned} \qquad (2.19)$$

where S is called the "dose". The solution to Fick's Law under these conditions is:

$$C(x,t) = \frac{S}{\sqrt{\pi Dt}} \exp\left[\frac{-x^2}{4Dt}\right] \qquad (2.20)$$

In silicon, dopant diffusion arises from interactions with native point defects – vacancies (V) or interstitials (I) atoms. Apart from a minority of dopants (notably arsenic and antimony) which diffuse partly or fully via interactions with vacancies, most dopants migrate as interstitials or impurity-interstitial complexes (Cowern *et al.*, 1999).

2.2.3 Ion Implantation

Ion implantation is an adding process by which dopant atoms are forcefully added into the semiconductor in the form of energetic ion beam injection. The purpose of ion implantation is to change the electrical, metallurgical, or chemical properties of a substrate. The typical ion energies are between 10 and 400 keV, and typical ion doses vary from 10^{11} to 10^{16} ions/cm^2. The main advantages of ion implantation include precise control over total dose, depth profile, and area uniformity, low-temperature processing and the ability for the implanted junctions to be self-aligned to the edge of the mask.

In an ion implantation process, ions lose their energy gradually by colliding with lattice atoms, transferring their energy to these atoms in the process (Xiao, 2001). The transferred energy is high enough for these atoms to break free from the lattice-binding energy, typically about 25 eV. These freed atoms collide with other lattice atoms while traveling inside the substrate, and knock them free from the lattice by transferring enough energy to them. These processes continue until none of the freed atoms have enough energy to free other lattice atoms. One energetic ion can cause thousands of displacements of lattice atoms. Damage caused by an energetic implanted ion is shown in Figure 2.9.

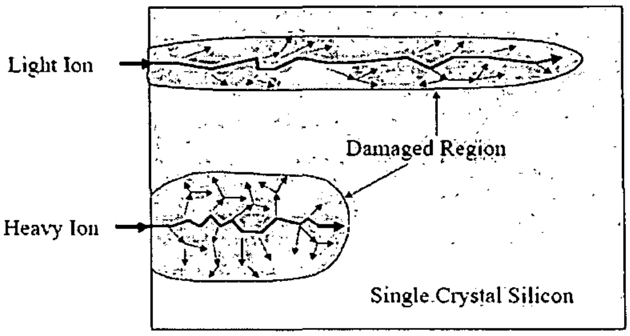


Figure 2.9: Damage caused by one ion (Reproduced from Xiao, 2001)

The damage created by one ion can be repaired quickly through self-annealing by the thermal movement of the atoms inside the substrate at the room temperature. However, if the amount of ions is so large (which is usually the case in an ion implantation process), there will be substantial lattice damage in the single-crystal substrate. The layer close to the surface becomes amorphous, and the crystal damage cannot be repaired by self-annealing process in a short time. The effect of damage is related to the dose, energy and mass of the ion species. The higher the dose and ion energy, the higher the damage induced. The substrate crystal structure can be completely destroyed and become amorphous near the substrate surface within the ion range, if the implantation dose is high enough. In order to repair the damage and activate the dopant after ion implantation process, a high temperature (> 1000 °C) annealing process is usually required. Figure 2.10 illustrates the crystal recovery and dopant activation in the thermal annealing process.

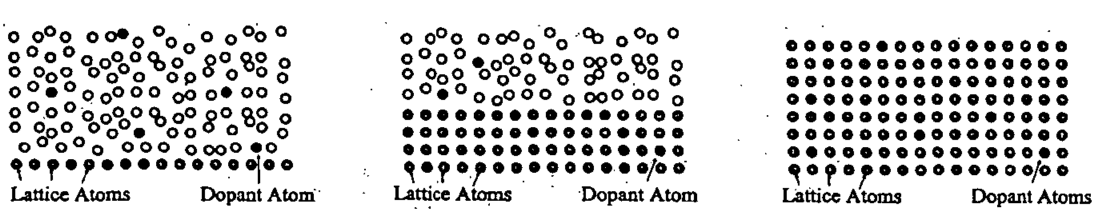


Figure 2.10: Post-implantation annealing process (Reproduced from Xiao, 2001)

The projected range of an ion in an amorphous material always follows Gaussian distribution, also called normal distribution. In single-crystal silicon, the lattice atoms have orderly arrangements, and lots of channels can be seen at certain angles. If an ion with the right implantation angle enters the channel, it can travel a long distance with very little energy loss as illustrated in Figure 2.11. This is called channeling effect (Xiao, 2001).

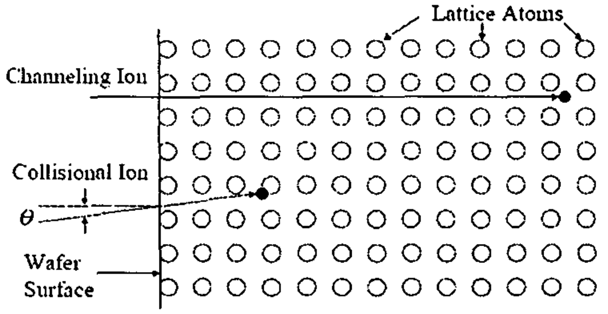


Figure 2.11: Channeling effect (Reproduced from Xiao, 2001)

The channeling effect causes some ions to penetrate deeply into the single-crystal substrate, forming a "tail" on the normal dopant distribution curve as show in Figure 2.12. It is an undesired dopant profile which could affect the performance of a device and thus, several techniques have been developed to minimize it.

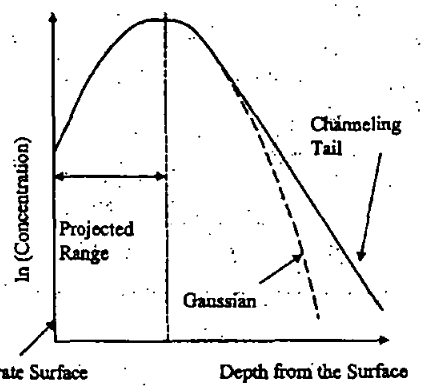


Figure 2.12: Dopant distribution with channeling effect
(Reproduced from Xiao, 2001)

One way to minimize the channeling effect is ion implantation on a tilted wafer, typically with a tilt angle of $\theta = 7°$ (Xiao, 2001). By tilting the wafer, the ions impact with the wafer at an angle and cannot reach the channel. The incident ions will have nuclear collisions instantaneously and effectively reduce the channeling effect. Most ion implantation process uses this technique to minimize the channeling effect, and therefore most wafer holders of ion implanters have the capability to adjust tilting angle of the wafer.

2.2.4 Rapid Thermal Anneal

With the continuous down-scaling of all relevant device parameters in silicon ULSI technology and the increasing circuit complexity, long-time furnace annealing steps have gradually been replaced by Rapid Thermal Processing (RTP) processes in the critical section of the device fabrication process. The most common application of RTP is Rapid Thermal Anneal (RTA) performed after the post-implantation process. After ion implantation process which involves energetic ion bombardment, the crystal structure of the silicon near the surface is usually heavily damaged. In order to anneal the damage and to activate the dopant, a high-temperature process is needed. At lower temperatures, the diffusion process outpaces the annealing process. However, at high temperature (>1000°C), the annealing process is faster due to higher annealing activation energy (~5 eV) as compared to diffusion activation energy (3 to 4 eV). Furnace process usually takes a long time with low anneal temperature and thus, it cannot minimize the dopant diffusion. In comparison, a RTP can perform the ramp up and cool down of wafer temperature very quickly at a temperature ramp-rate up to 250°C/sec and good temperature uniformity control at about 1100°C. The implantation damage can then be annealed in less than 20 seconds (Xiao, 2001).

Figure 2.13 shows the comparison between dopant diffusion in furnace and in RTP annealing process. Furnace annealing is typically being used in device with large feature size and for processes such as well implantation, post-implantation annealing and dopant drive-in, where the dopant atom diffusion is not a critical parameter. However, for devices with small feature size and for critical doping processes such as source/drain implantation annealing, RTP is being used to replace the furnace annealing in order to minimize the dopant diffusion and to produce shallow junctions.

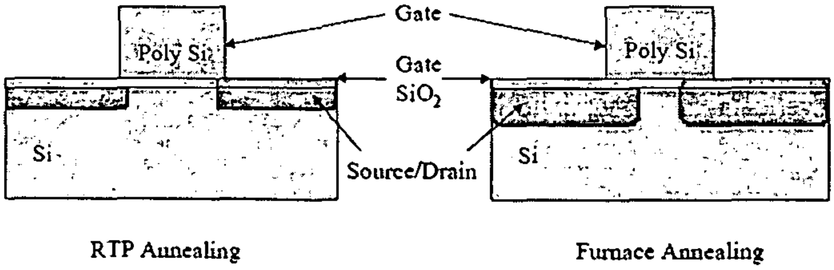


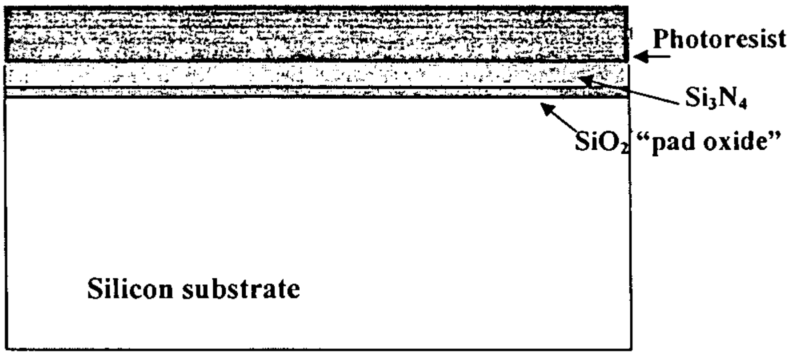
Figure 2.13: Dopant diffusion in furnace and RTP annealing process (Reproduced from Xiao, 2001)

2.3 Device Process Technology

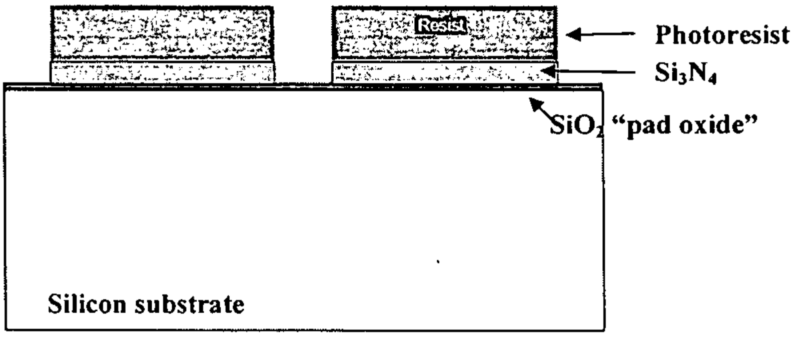
Basic fabrication processes described in the previous sections have been utilized to form certain key CMOS technology such as Local-Oxidation-of-Silicon (LOCOS), Lightly-doped drain (LDD) and Self-aligned silicide (SALICIDE) structures. These structures are required in order to address certain issues arises as a result of the reduction of the lateral and vertical dimensions of MOSFET, which is carried out in order to increase the density and speed of modem ICs.

2.3.1 Local-Oxidation-of-Silicon (LOCOS) Process

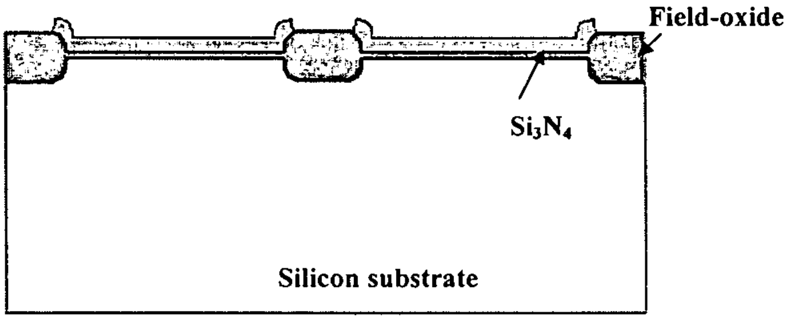
Several device isolation techniques have been developed in the need for a reduction of the spacing between devices in integrated circuits. One of the basic process used to provide electrical isolation for MOS ICs is the LOCOS isolation. As the name implies, the goal is to oxidize silicon only wherever a field oxide is needed (local oxidation). This can be done by using silicon nitride (Si_3N_4) to protect the silicon areas that are not to be oxidized. This is as Si_3N_4 does not allow oxygen diffusion at the typical oxidation temperatures (1000°C-1100°C). However, the selection of Si_3N_4 poses its own problems. The as-deposited Si_3N_4 films are typically in a state of high tensile stress (of the order of 1 GPa) and can induce large edge forces on the substrate when patterned. To address this problem, a thin thermal oxide (pad oxide) is usually grown in order to act as a stress-relief layer before depositing Si_3N_4. Figure 2.14 below shows the LOCOS structure formation. A thin pad oxide is initially grown, followed by the deposition of the Si_3N_4 and the photoresist. The resist is then patterned and the nitride is etched to protect the active areas. A thick field oxide is then grown on areas which are not covered by the silicon nitride mask. After the thermal oxidation, the silicon nitride is then etched off for the next processing steps.



(a) Thin pad oxide is grown followed by the deposition of Si_3N_4 and photoresist



(b) The resist is patterned and the nitride is etched to protect the active areas



(c) Thick field oxide is grown on selective areas which are not covered by the nitride

Figure 2.14: Sequence of LOCOS formation (Reproduced from MIT, 2005)

However, it should be noted that the standard isolation technique such as LOCOS suffers from bird's beak encroachment, which is a result of oxide growth under the edge of the nitride. As shown in Figure 2.15, the nitride is "pushed up" at the edges, resulting in a slope region on both sides and a waste of space areas. Even though

LOCOS may still be a viable technology for certain applications, it should be pointed out that for deep submicron devices, there is a clear trend to switch to more advanced isolation processes such as shallow trench isolation (STI).

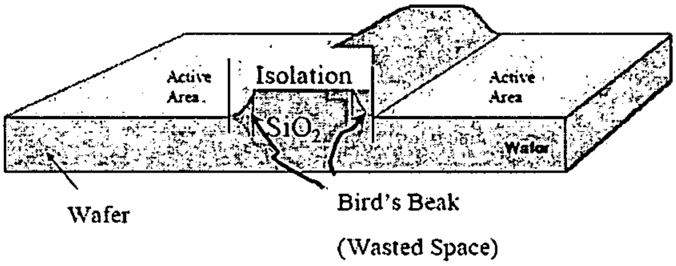
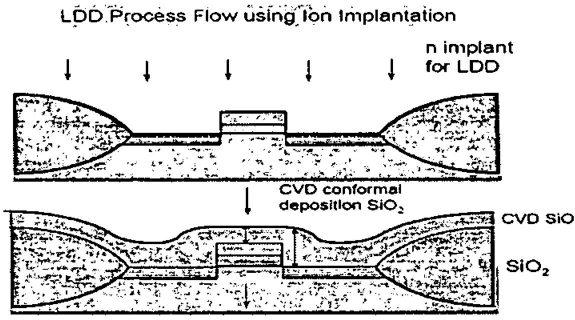


Figure 2.15: Bird's beak and LOCOS cross section
(Reproduced from Schramm *et al.*, 1999)

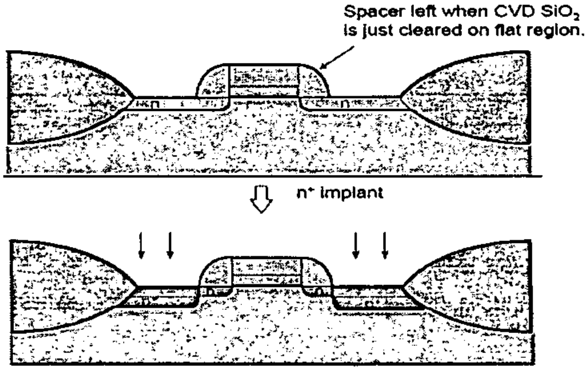
2.3.2 Lightly-doped-drain (LDD) Structure

As MOSFET is scaled down to submicron dimensions, high electric field has been shown experimentally to be present near the drain junction. This effect produces avalanche-induced drain-source breakdown and hot carrier effects (Chua & Liu, 1993). In order to reduce the lateral electric field and suppress the hot electron effect, LDD structure is required. LDD refers to a low energy, low-current implantation process where lightly doped n^- regions are inserted between the drain/source diffusion regions and the channel.

Figure 2.16 shows the sequences of LDD structure formation. In Figure 2.16 (a), a lightly-doped n^- minus regions is first formed between the source/drain diffusion regions and the channel. This is followed by conformal deposition of SiO_2 where the vertical thickness of the oxide layer will be much thicker than that on top of the polysilicon. Next, an anisotropic etch which is an etching mechanism that etches only in the downward vertical direction is carried out to form the sidewall spacer as shown in Figure 2.16 (b). The spacers that are formed are then used as an implant mask to recess the next source/drain implants away from the gate edges. This enables having shallow source/drain extensions at the edge of the channel (with the poly gate being the implant mask) and deeper source/drain away from the channel (with the spacer and gate being the implant masks). This is later followed by the next processing steps of the deep source/drain implant.



(a) Formation of lightly-doped regions followed by conformal deposition of SiO_2

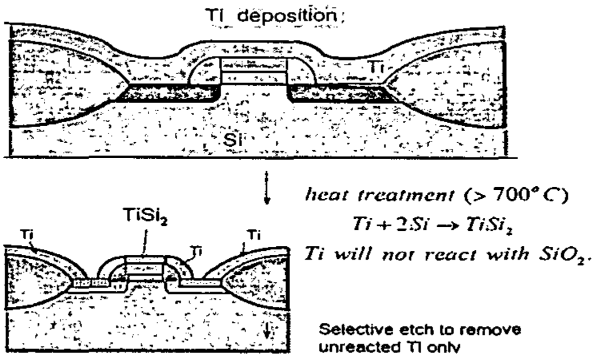


(b) Formation of sidewall spacer followed by deep source/drain implant

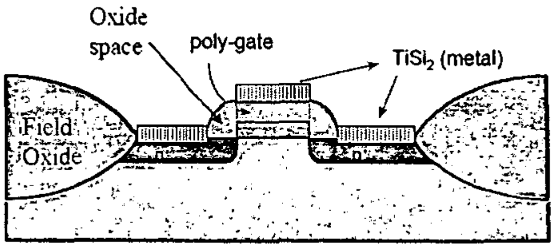
Figure 2.16: Sequence of LDD formation (Reproduced from UC Berkeley, 2006)

2.3.3 SALICIDE Process

SALICIDE or Self-Aligned Silicide process refers to a process in which silicide contacts are formed in those areas in which deposited metal (which after annealing becomes a metal component of the silicide) is in direct contact with silicon, hence are self-aligned. Over insulators, these metals do not react, and thus they can be easily etched away. The result is silicides just over the sources, drains and polysilicon gates. It is commonly implemented in MOS/CMOS processes in which ohmic contacts to the source, drain, and poly-Si gate are formed by “salicide” process. The purpose of forming the SALICIDE layer is to reduce the contact resistance to the source and drain and to reduce the resistance of the polysilicon gate insulators. Reducing resistance is essential to produce fast transistors. Silicides such as Titanium Silicide ($TiSi_2$) and Tungsten Silicide (WSi_2) which have much lower resistivity than polysilicon are commonly used (Xiao, 2001). The following describes the formation of a SALICIDE structure. In Figure 2.17 (a), the metal used to form the silicide is deposited after the source/drain implant. The wafer is then heated, which causes the silicide reaction to occur at areas where the metal is in contact with the silicon. Everywhere else, the metals remain unreacted. In Figure 2.17 (b), it can be seen that the unreacted metal is selectively removed, leaving each exposed source/drain and polysilicon gate region completely covered by a silicide film.



(a) Metal of titanium is deposited followed by heat treatment



*Self-aligned structures are always preferred for process integration

(b) The resulting silicide structure

Figure 2.17: Sequence of SALICIDE formation (Reproduced from UC Berkeley, 2006)

2.4 Summary

In this chapter, an introduction to the concept of transistor scaling is made. Basically, there are three types of scaling guidelines available which are the constant-field scaling (full scaling), constant-voltage scaling and general scaling. Each of the scaling guidelines has its own advantages and limitations and thus, the actual transistor scaling is determined by the designer himself by varying selected parameters to obtain an optimum performance, whereby the scaling guidelines is used as reference only. Descriptions are also made on the adverse effects that arise as a result of scaling-down of devices, known as short-channel effects and techniques used to compensate these effects. As the later chapter will discuss on the fabrication process involved in the formation of MOS transistor, brief descriptions are also made on the important processes involved such as oxidation, diffusion, ion implantation and RTA. Utilization of these basic fabrication processes into formation of certain key CMOS technology such as LOCOS, LDD and SALICIDE structures are also described at the end of this chapter.

CHAPTER 3

TECHNOLOGY COMPUTER AIDED DESIGN (TCAD) TOOLS

This chapter describes the Technology Computer Aided Design (TCAD) Tools used in this research work. It starts with the advantages of using the TCAD Tools, followed by brief descriptions on the features of Process Simulator and Device Simulator used. Important aspects to be considered for each stage of simulation are also highlighted.

3.1 Advantages of TCAD Tools

The concept of virtually fabricated and characterized sub-micron devices using TCAD Tools provides several advantages,

(a) Cost-effectiveness

Advanced wafer manufacturing usually requires a combination of advanced machinery equipments, new innovation of techniques as well as technological expertise which are extremely expensive. In order to develop new processes and devices, costly test lots need to be fabricated. Computer simulation on the other hand is cheap and efficient. It is repeatable and fast in generating the output and thus, is highly sought as an alternative to experiments as a source of data.

(b) Enabling scrutiny into the insight physic of the device studied

By using the TCAD tools, researchers will be able to look deeper into the physic of the device studied for information that is not accessible by experimental diagnostics. For example, the effects of variation of implantation energies and doses on the resulting device structure and the electrical performance can be obtained through simulation. To obtain such information by real experiment is not only expensive, but also time consuming and practically impossible. The use of TCAD tools will undoubtedly expedite technological progress as compared to the conventional approach of trial-and-error through experimentation (empiricism).

The two major components of TCAD that are used in this work are (1) Process simulator and (2) Device simulator.

3.2 Process Simulator – ATHENA

Physically-based process simulator is a software that models the manufacturing process and predicts the structures generated from the specified process sequences. This is done by solving systems of equations that describe the physics and chemistry of semiconductor processes. The process simulator used in this work is Silvaco *ATHENA 2D Simulation Framework*. One of its modules is the *SSUPREM4* which is used particularly in the design, analysis and optimization of silicon semiconductor related structures. In this work, both the 0.35 μm and 0.25 μm CMOS transistor simulation is based on the silicon semiconductor structure which made the *SSUPREM4* module embedded in the *ATHENA Framework* a very important components in conducting the research.

3.2.1 Requirements for Grid Definitions

In process simulation, it is critical to have a correct specification of a grid, as the number of nodes in the grid has a direct influence on simulation accuracy and time. A finer grid should exist in those areas of the simulation structure where ion implantation will occur or where p-n junction will be formed. In MOSFET modelling, an important area for accuracy is the inversion region under the gate, as it is this charge that is responsible for current conduction in the device. Too coarse of a mesh will results in too high a current being simulated (Silvaco Int., 2001).

Another important area that requires proper gridding is in the simulation of oxidation steps. A well gridded oxide is important to properly account for dopant redistribution during the oxidation step. This is as silicon is being consumed, dopants are transported across the Si/SiO_2 interface. Improper gridding can result in jagged oxide shapes and errors in resolving impurity distributions. In the oxidation simulation steps, proper gridding will cause grid points to be added at predefined spacing as the oxide is growing.

3.2.2 Requirements for Substrate Depth Modeling

A correct choice of substrate depth is vital in process simulation in order to accurately model dopant diffusion. This is because the rate of dopant diffusion is highly dependent on the level of damage in the substrate. Therefore, an accurate modeling of dopant diffusion will require an accurate modeling of substrate damage, particularly the movement of interstitials. In general, the interstitials created directly or indirectly by implantation and oxidation, tend to diffuse much greater distances than the dopant. In order to allow the interstitials concentrations to return to background levels at the bottom of the simulated substrate, the substrate depth chosen for modeling purposes must therefore be deep enough, even if no dopant diffusion occurs at this depth. If the substrate depth is made too shallow, there will be no interstitials gradient. The concentration of interstitials will then start to pile up because they are not being removed through diffusion into the bulk of the substrate. If the level of modeled interstitials becomes too high, the diffusion of dopant will also be too high and the simulation will be inaccurate, even near the surface of the substrate. For normal small geometry MOSFET/Bipolar processing, a substrate depth of 20 μm should be more than adequate (Silvaco Int., 2001).

3.2.3 Selection of Physical Models

3.2.3 (a) Oxidation Models

The oxidation models are made up of *vertical*, *compress* and *viscous* model. The *compress* model which is the *default* oxidation model, is recommended for planar and non-planar structure where stress effects play a minor role in determining the oxide shape. It is activated by specifying *compress* in the *method* statement prior to *diffuse* statement. To calculate stresses in a growing oxide, *viscous* model is used and the shape created for the silicon/oxide interface is almost the same as does the *compress* model. This model is activated by specifying the *viscous* parameter in the *method* statement prior to *diffuse* statement. The model that is used in this work is the *compress* model which is used to model 2D oxidation in non-planar structure with linear flow.

3.2.3 (b) Diffusion Models

In *ATHENA*, the diffusion models described how implanted profiles of dopants/defects redistribute themselves during thermal treatment, due to concentration gradients and internal electric fields. The three most basic models are *fermi*, *two-dimensional (two.dim)* and *fully-coupled (full.cpl)* diffusion models. All the three models rely on the concept of Pair Diffusion, which says that a dopant atom cannot diffuse on its own as it needs the assistance of a point defect (a silicon self interstitial or a lattice vacancy) in the near vicinity as a diffusion vehicle.

The *fermi* diffusion model assumes constant level of point defects and therefore does not account for defect enhanced diffusion. The *two.dim* diffusion model on the other hand, solves the two-dimensional distribution of point defects and includes point defects generation during implant and oxidation. It is suitable for oxidation/silicidation enhanced diffusion. Meanwhile, the *full.cpl* diffusion model takes into account coupling between point defects and individual dopants. It can be used for simulation of transient diffusion phenomena, low temperature diffusion and co-diffusion of dopants (emitter push effect). Both *two.dim* and *full.cpl* models include the local point defect concentration in the diffusion rate of the dopants. The model used in this work is the *full.cpl* model which is used in particular after the post high dose implant and co-diffusion effects. However, the usage of this model requires higher execution times.

3.2.3 (c) Ion Implantation Models

The ion implantation hierarchy is made up of two basic model types: *analytic* and *Monte Carlo*. By default, the *analytic* models are used. Analytical models are based on the reconstruction of implant profiles from the calculated or measured distribution moments. The Monte Carlo/statistical technique on the other hand, uses the physically based Monte Carlo calculation of ion trajectories in order to calculate the final distribution of stopped particles.

Analytic Implant Models are made up of *gaussian, pearson* and *dual pearson* models. These models give very good results when applied to ion implantation into simple planar structures (bare silicon or silicon covered with thin layer of other material). The *gaussian* implant model uses the *gaussian* distribution (simplest 1D profiles) which is specified by the *gauss* parameter in the *implant* statement. Generally, the *gaussian* distribution is inadequate because real profiles are asymmetrical in most cases. The *pearson* implant model is the simplest and most widely approved method for calculation of asymmetrical ion-implantation profiles. *ATHENA* uses this function to obtain longitudinal implantation profiles.

In order to extend applicability of the analytical approach toward profiles heavily affected by channelling, the *dual pearson* method is suggested. By default, *ATHENA* uses *SIMS-Verified Dual Pearson (SVDP)* implant models. These are based on tables from the University of Texas at Austin. These tables contain *dual pearson* moments for B, BF_2, P and As extracted from high quality implantation experiments conducted by the University of Texas at Austin. These implantation tables contain dose, energy, tilt, rotation angle, and screen oxide thickness dependence as specified (Silvaco Int., 2001).

Statistical/Monte Carlo Implant Models is used to simulate ion implantation in non-standard conditions (for example, structures containing many non-planar layers (material regions)) and for the cases which have not been studied experimentally yet. This approach allows calculation of implantation profiles in an arbitrary structure with accuracy comparable to the accuracy of analytical models for a single layer structure. Table 3.3 summarizes the implantation models available. The model used in this work is the *svdp* model which is the default model.

The effect of implant damage enhanced diffusion is another important aspect to be discussed as damage generated by implantation leads to an enhancement of the diffusion of the dopants during subsequent heat cycles. Simulation of the enhanced diffusion effects can be divided between two processes. First, *ATHENA* must simulate the implant damage generated by a given implant and secondly, it must model the effect that these defects have on subsequent impurity diffusion.

The most practical model for coupling implant damage to subsequent diffusion calculation is the *+1* model. In its simplest form, the *+1* model adds exactly one interstitial for each implanted ion. This is a reasonable approximation if one assumes that the vacancies and interstitials created by the implant recombine quickly relative to the timescale needed to produce significant diffusion. This leaves one extra interstitial for each ion, assuming the implanted ion has replaced it on the lattice.

The *+1* model is applicable to both Monte Carlo and the default analytic implants and can be invoked by including the *unit.dam* parameter on the *implant* statement. A commonly applied variation to this model is to scale the number of generated interstitials. In *ATHENA*, this can be done using the parameter *dam.fact* on the *implant* statement.

3.3 Device Simulator-ATLAS

Device simulator is being used to predict the terminal characteristics of semiconductor devices for DC, AC small signal and transient stimuli, by solving the fundamental physics equation describing the dynamics of carriers in semiconductor devices. The device simulator used in this work is *Silvaco ATLAS 2D Simulation Framework*. One of the most important modules in this framework which is used extensively in the CMOS simulation is the *S-PISCES*. It is a powerful and accurate two-dimensional device modeling program that simulates the electrical characteristics of silicon based semiconductor devices including MOS, bipolar, SOI, EEPROM and power device technologies. *ATLAS* is often used in conjunction with the *ATHENA* process simulator. In this work, *ATHENA* will be used initially to predict and generate the physical structures associated with the specified process sequence. The resulting physical structures are then used as input by *ATLAS*, which then predicts the electrical characteristics associated with the specified bias conditions. By combining *ATHENA* and *ATLAS*, the impact of process parameters on device characteristics can be easily determined.

3.3.1 Definition of Material Parameters and Models

All materials can be categorized into three classes: semiconductors, insulators and conductors. A different set of parameters are required for each class. However, default parameters for materials properties used in device simulation are already available for many materials. To specify physical model, *models* and *impact* statement are used. This physical models can be grouped into five classes namely *mobility, recombination, carrier statistics, impact ionization* and *tunneling*. To specify interface charge density and surface recombination velocity, *interface* statement is used. Meanwhile, the *contact* statement is used to specify the metal work function of electrodes. In this work, only the physical models for *recombination* and *mobility* are specified.

3.3.2 Selection of Numerical Methods

There are basically three types of solution techniques: *(a) decoupled (gummel), (b) fully-coupled (newton)* and *(c) block.*

In *Gummel* method, each iteration solves for each unknown in turn keeping the other variables constant, repeating the process until a stable solution is achieved. *Gummel* iteration typically converges relatively slowly, but the method will often tolerate relatively poor initial guesses. The *Newton* method on the other hand, solves the total system of unknowns together. This method is useful when the system of equations is strongly coupled and has quadratic convergence. The *Newton* method may however, spend extra time solving for quantities which are essentially constant or weakly coupled. *Newton* also requires a more accurate initial guess to the problem to obtain convergence. The *block* methods will solve some equations fully coupled while others are de-coupled. It can be useful to start a solution with a few *Gummel* iterations to generate a better guess and then switch to *Newton* to complete the solution. In this work the *Gummel Newton* iteration method is selected.

3.3.3 Specification of Solutions

ATLAS is able to calculate DC, AC small signal and transient solutions. In order to obtain solution, the voltages need to be defined on each electrode in the device. *ATLAS* will then calculate the current through each electrode. Internal quantities such as carrier concentrations and electric fields throughout the device can also be calculated by *ATLAS.*

3.3.4 Analysis of Results

There are three types of outputs produced by *ATLAS* tool, namely the *runtime outputs, log files* and *solution files.* The *runtime output* stores the run-time messages produced by *ATLAS* which typically include important values extracted from the simulation. All error messages also go to the run-time output. The *log files* on the other hand, store the *dc, ac small signal* and *transient* terminal characteristics for a sequence of *solve* statements. For example, in *dc* simulations, the terminal characteristics that are saved are current and voltages for each electrode. These *log files* are typically viewed in *Tonyplot*. Parameter extraction on data in *log files* can be done in *Deckbuild*. Meanwhile, the physical quantities of the structure at each grid node for a single bias point are stored in s*olution files*. Any evaluated quantity within the device structure in question, from doping profiles to electron concentrations and electric fields can thus be viewed by the user.

3.4 Summary

The use of numerical simulator proves to be indispensable in this research work as it provides flexibility in between performing virtual fabrication of the device at hand, and in performing instant characterization of the electrical parameters using easy interface. The flexibility features between process and device simulation allows considerable amount of time to be saved. Important parameters such as the requirement for grid definition, the correct modeling of substrate depth and the correct selection of physical models plays an important role in determining and ensuring the success and accuracy of simulation work performed. As such, these parameters should be looked into and noted carefully prior to performing the device design which will be done and discussed more thoroughly in the next chapter.

CHAPTER 4

SIMULATION DESIGN

This chapter provides an overview of the process and device simulation design of the 0.35 µm and the successive 0.25 µm CMOS transistor. Process recipe of 0.35 µm CMOS by UC Berkeley Microfabrication Lab (Horvath *et al.*, 2005) serves as the design basis. Included are the summary of the simulation design methodology, the flow of front-end processes used in the design and the summary of parameters used in both process and device simulation.

4.1 Simulation Design Methodology

In the process of downscaling the 0.35 µm CMOS transistor to 0.25 µm, a *"two-step design"* approach is proposed. The *first design step* involves conversion of physical data of 0.35 µm CMOS to simulation environment and extraction of the associated electrical parameters. This is followed by the *second design step* which constitute miniaturization of 0.35 µm to 0.25 µm CMOS through downscaling. A scaling factor, κ of 1.4 which is derived from direct division of 0.35 µm with 0.25 µm is adopted for selected parameters. Electrical characterization is then performed to evaluate the performance of the downscaled 0.25 µm CMOS. For both 0.35 µm and 0.25 µm CMOS, a separate simulation is performed for NMOS and PMOS respectively in order to save up computation times. Both NMOS and PMOS device structure have been simulated from starting wafer until contact hole opening. A summary of simulation design methodology proposed and adopted in this work is illustrated in Figure 4.1.

The parameters selected for scaling purposes are of physical gate length (L_g), gate oxide thickness (t_{ox}), threshold-adjust implant doping and the supply voltage (V_{dd}). L_g which is the smallest feature of the MOSFET patterned by lithography and etching is scaled down to 0.25 µm from the initial 0.35 µm. As L_g is scaled down, t_{ox} must also be decreased to enhance the gate control in creating or stopping the inversion channel by the gate electrode. According to NTRS (1997), t_{ox} of 40-50 angstrom is required for the 0.25 µm technology. In this work, t_{ox} is scaled down from 80 angstrom in the

0.35 μm simulation, to 57 angstrom in the 0.25 μm simulation. For threshold-adjust implant doping, the doping levels is increased by a factor of 1.4 for each of the respective NMOS and PMOS. An increase in the channel doping is necessary in addition to the reduction in the t_{ox} to reduce short-channel effects (Zeitzoff, 1999). To avoid the break down of oxide and ensuring constant electrical field across source and drain (constant-field scaling), V_{dd} is also decreased as t_{ox} is decreased. In this simulation, V_{dd} is reduced from 3.3 V in the 0.35 μm simulation, to 2.5 V in the 0.25 μm simulation as per guideline given in NTRS (1997).

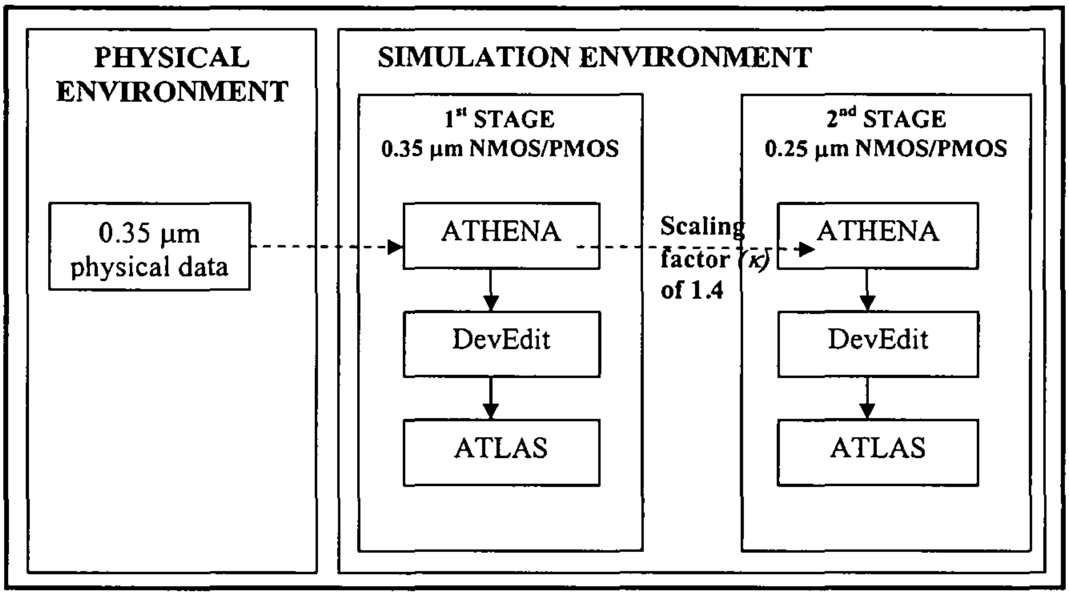
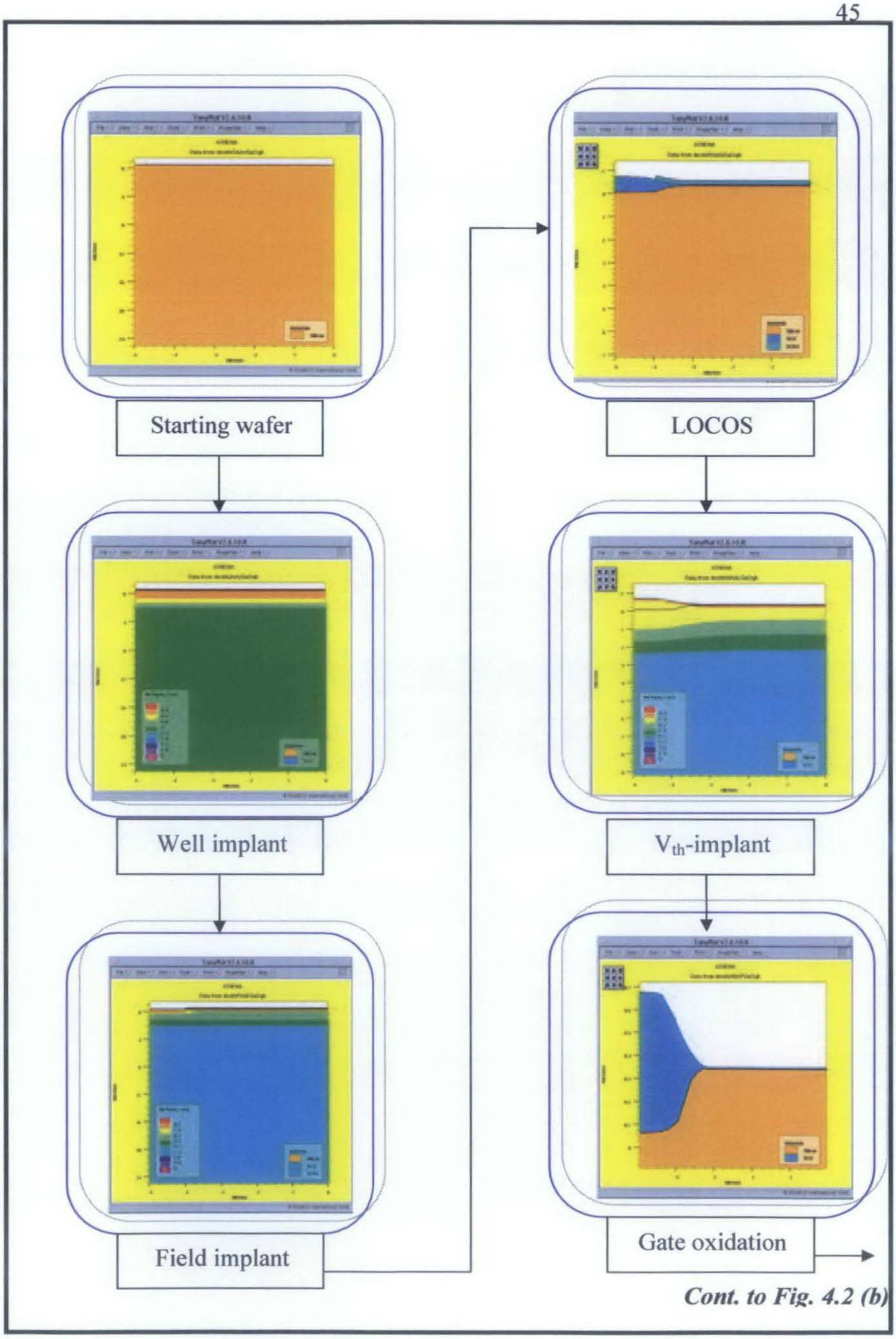


Figure 4.1: Summary of design methodology adopted in the downscaling of 0.35 μm CMOS transistor to 0.25 μm

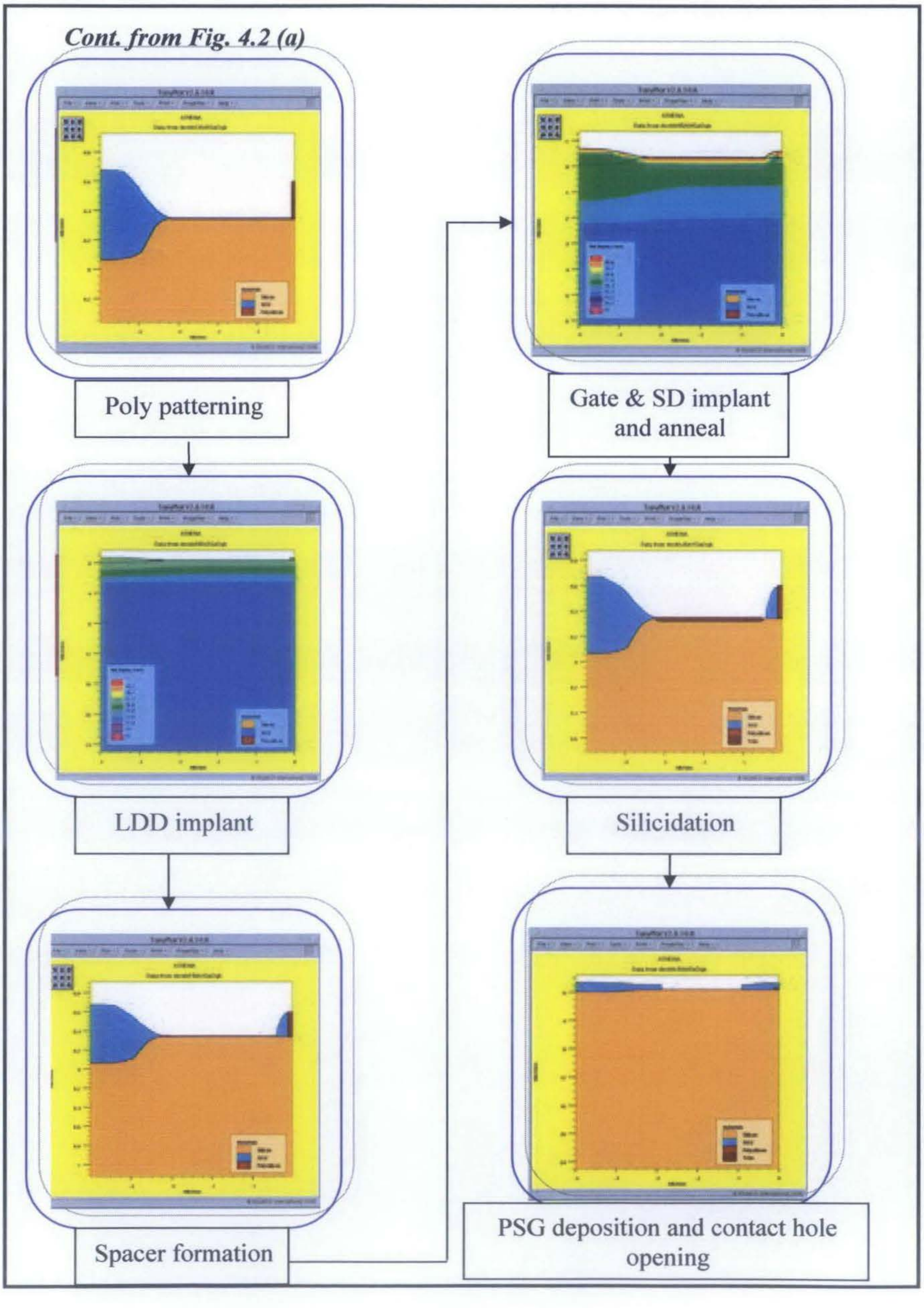
4.2 Summary of Front-end Processes

To lay a ground for further discussion, the processing sequence adopted for the 0.35 μm and 0.25 μm CMOS transistor design is illustrated in Figure 4.2 (a) and 4.2 (b).



Cont. to Fig. 4.2 (b)

(a)



(b)

Figure 4.2: Summary of process flow for the CMOS simulation design

It should be noted that an above illustration is used to show the simulation of half of the transistor structure. To obtain the complete structure, a *"mirror"* command is invoked, and an example of the complete transistor structures generated is shown in Figure 4.3.

Figure 4.3: Complete transistor structure generated by ATHENA for the 0.25 μm NMOS

The following describes each of the processing sequence as illustrated in Figure 4.2 in detailed. A starting wafer of p-type is used at the start of the fabrication process. Two separate wells are then formed for n-channel and p-channel transistor, in a process called a "twin well" process. The "twin-well" process allows the doping profiles of each of the device types to be set independently, unlike the single well where well-doping must always be higher than the doping of the substrate in which one type of device is made (Wolf, 1995). This is then followed by field-implant (channel-stop implant) which is applicable for n-channel transistor only. Ion implantation is used to form a p-type doped isolation region before the thick field oxide growth. The field-implant is used to form guarding rings around the active regions, which helps to electrically isolate neighboring transistors.

Next, field-oxide is formed to isolate neighboring transistors on a chip surface. The LOCOS technique which is based on the principle of growing the field oxide in selected regions instead of selectively etching away the active areas after oxide growth, is being used to form the field-oxide. The threshold adjustment implants are then made with BF_2 and phosphorous to adjust the V_{th} for the respective NMOS and PMOS. This is later followed by gate oxidation and deposition of undoped polysilicon.

An implantation using arsenic for n-channel and BF_2 for p-channel MOS transistors is then carried out to form a shallow lightly-doped source and drain extensions. This is a standard process called the LDD which was devised to improve reliability against hot-carrier instability of a transistor by avoiding electric field concentration. The concept of LDD is the insertion of a graded profiled, n^- drain region between the n^+ part of the drain and the channel region. The n^- region sustains the drain voltage over a longer region and thus reduces the peak electric field which is predominantly lateral at the drain (Tasch *et al.*, 1990).

The sidewall spacers are then formed after the LDD implant by depositing a conformal silicon oxide on the silicon surface. This is followed by an anisotropic etch, a mechanism that etches only in the downward vertical direction and not along the silicon surface. Along the sides of the polysilicon, the vertical thickness of the oxide layer will be much thicker than on top of the polysilicon. All of the deposited insulator is then removed except for a wedge along the sides of the polysilicon lines. These sidewall spacers are used to provide a diffusion buffer for the dopant in the source and drain junction.

This is followed by the formation of deep source and drain in the regions adjacent to the side wall using phosphorous for NMOS and boron for PMOS. At the same time the source/drain contacts are doped, the polysilicon gates are also doped by ion implant. A dual doped poly scheme is employed whereby the poly gate is doped with

n^+ dopants for NMOS and p^+ dopants for PMOS. The fabrication of dual doped poly gates creates a surface-channel PMOS which will have higher immunity against short-channel effects (Lee *et al.*, 2000).

After the source/drain implant, RTA process is employed to repair the damage to the lattice by restoring the single-crystal structure and activating the dopant. RTA is being used instead of conventional furnace diffusion to minimize dopant diffusion and to produce ultra shallow junctions. This is then followed by a silicidation process. Silicidation is a process of forming a surface layer of a refractory metal silicide on silicon. In this work, titanium is being deposited on the silicon (gate polysilicon and source drain regions) and a layer of silicide is formed when the two substances react at elevated temperatures. The resistivity of titanium silicide which is of 12-25 μΩ-cm is the lowest among various metal silicides (Kim *et al.*, 1998). After silicidation, the phosphor-silicate glass (PSG) is deposited and subjected to a reflow process for smoothness. This is followed by contact holes opening made by etching. The summary of process parameters used in the simulation of 0.35 μm and 0.25 μm MOS transistors is tabulated in Table 4.1 and 4.2 respectively.

Table 4.1: Summary of process parameters used in the simulation of 0.35 μm MOS transistor

Process Flow	NMOS	PMOS
Starting wafer	• wafer orientation <100> • boron concentration=$3x10^{14}$ cm^{-3}	
Well-implant	• Boron • $5x10^{12}$ cm^{-2} • 60 keV	• Phosphorous • $1x10^{13}$ cm^{-2} • 150 keV
Field-implant	• Boron • $2x10^{13}$ cm^{-2} • 80 keV	• Nil
LOCOS	• diffuse time = 120 min temp = 1000 °C wet O_2 • diffuse time = 20 min temp = 1000 °C; nitrogen	
V_{th}-implant	• BF_2 • $4x10^{12}$ cm^{-2} • 50 keV	• Phosphorous • $2x10^{12}$ cm^{-2} • 30 keV
Gate oxidation	• diffuse time = 17 min temp = 900 °C dry O_2 • diffuse time = 30 min temp = 1000 °C nitrogen	
Polysilicon deposition	• deposit poly thickness = 0.25 μm	
LDD implant	• Arsenic • $5x10^{13}$ cm^{-2} • 30 keV • tilt = 7 & -7 degree • rotation = 0 degree	• BF_2 • $5x10^{13}$ cm^{-2} • 10 keV • tilt = 7 & -7 degree • rotation = 0 degree
Spacer formation	• deposit oxide thickness=0.3 μm	
Gate & S-D implant	• Phosphorous • $3x10^{15}$ cm^{-2} • 40 keV	• Boron • $3x10^{15}$ cm^{-2} • 20 keV
S-D anneal	• diffuse time = 10 sec temp = 900 °C nitrogen • diffuse time = 5 sec temp = 1050 °C nitrogen	
Silicidation	• deposit titanium thickness = 0.03 μm	
PSG deposition	• deposit oxide thickness = 0.7 μm	

Table 4.2: Summary of process parameters used in the simulation of 0.25 μm MOS transistor

Process Flow	NMOS	PMOS
Starting wafer	• wafer orientation <100> • boron concentration=$3x10^{14}$ cm^{-3}	
Well-implant	• Boron • $5x10^{12}$ cm^{-2} • 60 keV	• Phosphorous • $1x10^{13}$ cm^{-2} • 150 keV
Field-implant	• Boron • $2x10^{13}$ cm^{-2} • 80 keV	• Nil
LOCOS	• diffuse time = 120 min temp = 1000 °C wet O_2 • diffuse time = 20 min temp = 1000 °C nitrogen	
V_{th}-implant	• BF_2 • $5.6x10^{12}$ cm^{-2} • 50 keV	• Phosphorous • $2.8x10^{12}$ cm^{-2} • 30 keV
Gate oxidation	• diffuse time = 25 min temp = 850 °C dry O_2 • diffuse time = 30 min temp = 1000 °C nitrogen	
Polysilicon deposition	• deposit poly thickness = 0.25 μm	
LDD implant	• Arsenic • $5x10^{13}$ cm^{-2} • 30 keV • tilt = 7 & -7 degree • rotation = 0 degree	• BF_2 • $5x10^{13}$ cm^{-2} • 10 keV • tilt = 7 & -7 degree • rotation = 0 degree
Spacer formation	• deposit oxide thickness=0.3 μm	
Gate & S-D implant	• Phosphorous • $3x10^{15}$ cm^{-2} • 40 keV	• Boron • $3x10^{15}$ cm^{-2} • 20 keV
S-D anneal	• diffuse time = 10 sec temp = 900 °C nitrogen • diffuse time = 5 sec temp = 1050 °C nitrogen	
Silicidation	• deposit titanium thickness = 0.03 μm	
PSG deposition	• deposit oxide thickness = 0.7 μm	

4.3 Process Simulations

Process simulation was carried out using *ATHENA* software package which is an advanced 2D process simulation tool set. The most important tool of *ATHENA* is *SSUPREM4* which is the state-of-the-art 1D and 2D semiconductor process simulator used in the design, analysis and optimization of silicon fabrication technologies and device structures. Description of the design parameters used in this work such as the grid definitions, substrate depth modeling and the physical models chosen is highlighted. Unless otherwise specified, the descriptions of the design parameters are applicable to both the 0.35 μm and 0.25 μm process simulations. Figure 4.4 illustrate the sequence for process simulations. It can be seen that the transistor structure generated in *ATHENA* is later fed into *DevEdit for* mesh modifications.

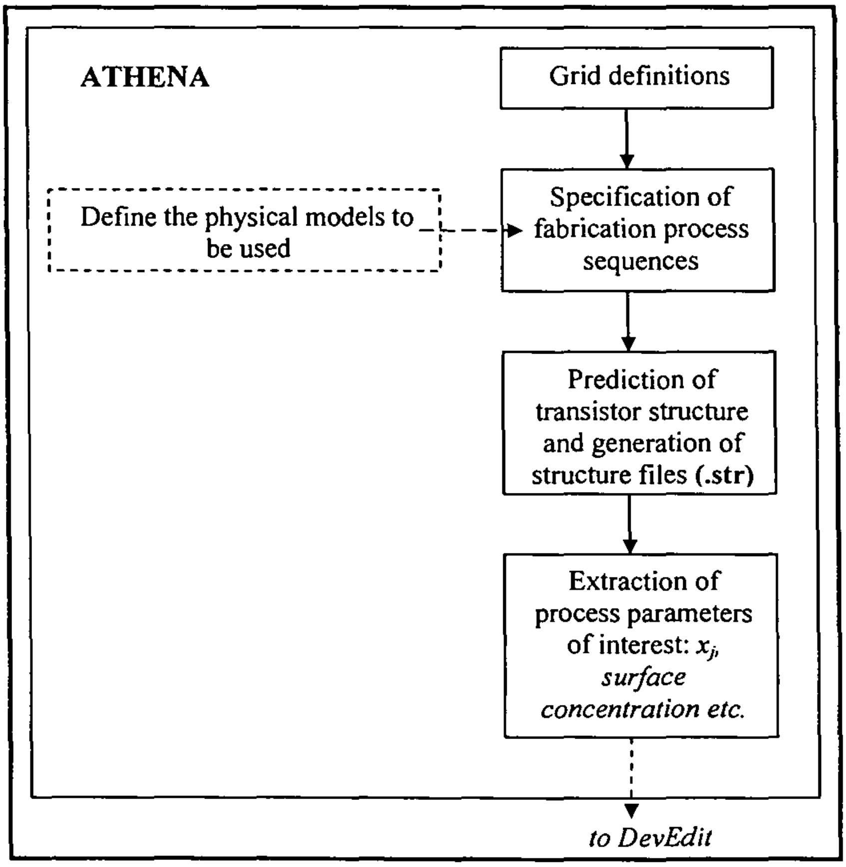


Figure 4.4: Sequences of process simulations in *ATHENA*

4.3.1 Grid Definitions

The process design starts with the definition of grid in the silicon structure of 5 micron in length and 25 micron in depth. Pointed right along the structure (x-axis), the grid size used is of between 0.05 to 0.1 micron. Pointed down to the depth of the structure (y-axis), the grid size used is of between 0.05 to 0.4 micron. The smallest grid size used is of 0.05 micron which extends from the wafer surface into the substrate depth (y-axis) of 0.4 micron. In between the depth of 0.4 to 6 micron, the grid size used is of 0.1 micron. Elsewhere, extending deeper into the substrate, the grid size used is of 0.3 and 0.4 micron. It should be noted that finer grid is needed only in the upper portion of the structure where doping occurs. Near the bottom of the substrate, the grid points may be placed quite far apart as all that is required is that there are sufficient grid points to model the gradient of interstitials in this region (Silvaco Int., 2001). Figure 4.5 shows the grid used in the simulation work of CMOS transistor.

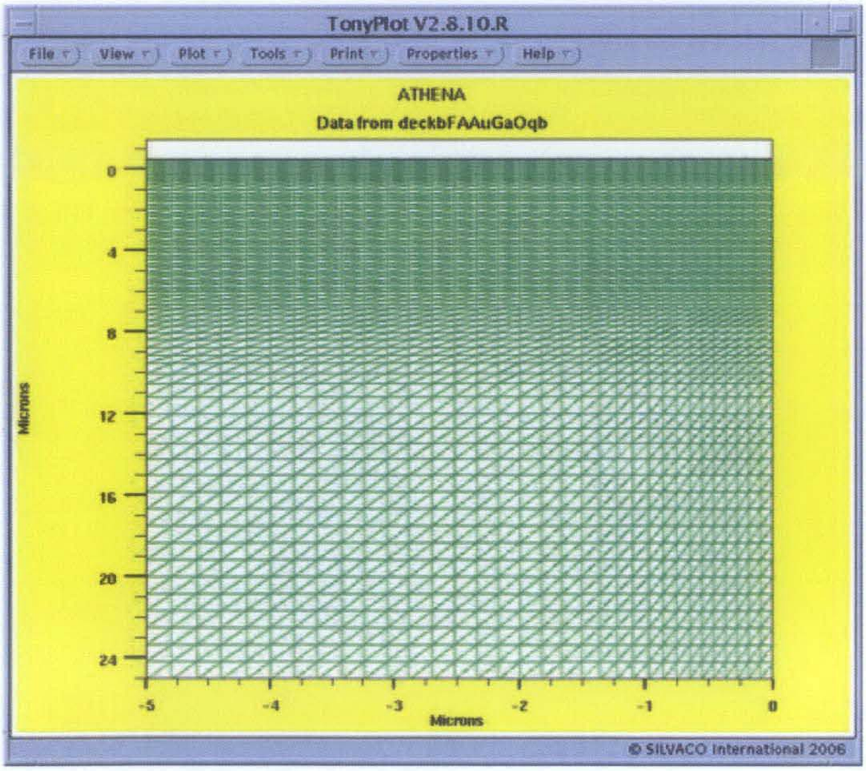


Figure 4.5: Plot of the grid used in the simulation work of CMOS transistor

4.3.2 Modeling of Substrate Depth

As mentioned in Chapter 3, an extended depth of the simulated structure (25 µm) is needed at the start of the simulation to enable adequate sink of point defects during subsequent process simulations. Silvaco Int. (2001) stated that point defects have large diffusivities than dopants and may therefore diffuse down to the bottom of the structure during a simulation and thus, a substrate depth of 20-25 µm is usually recommended. If the simulation structure is too shallow, an unphysically high defect concentration might be obtained in the regions where dopant profiles are present. This will lead to high concentration of dopant diffusion and inaccuracies of simulation results. Figure 4.6 shows an example of the simulated structure of 0.25 µm PMOS transistor with simulated substrate depth of 25 µm, used in this simulation work.

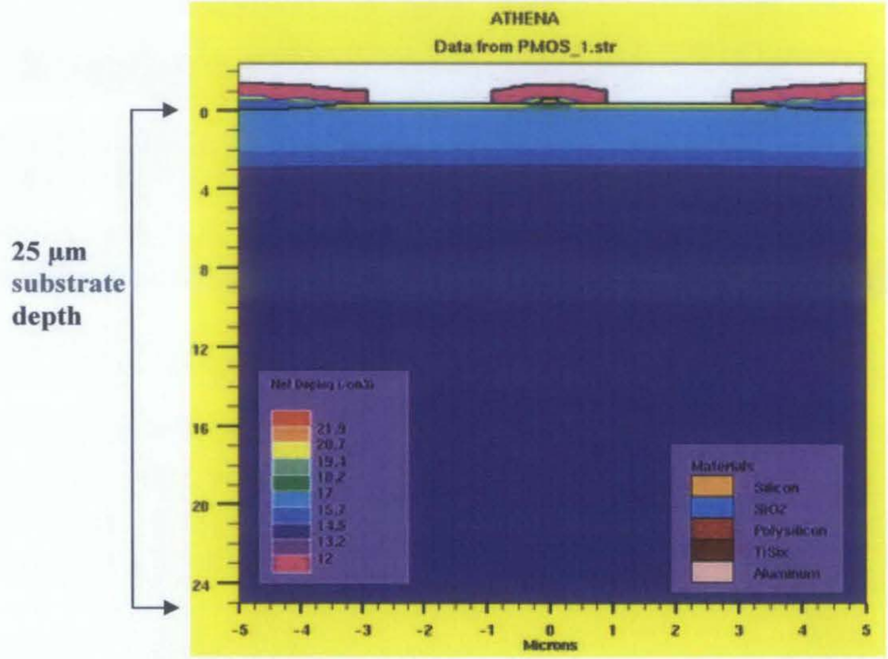


Figure 4.6: Simulated structure of 0.25 µm PMOS transistor with substrate depth of 25 µm

4.3.3 Selection of Physical Models

4.3.3 (a) Oxidation Models

For oxidation, the default *compress* model which is recommended by Silvaco Int. (2001) for planar and non-planar structure is being used throughout the simulation.

4.3.3 (b) Diffusion Models

For diffusion, the *fully-coupled* model is initiated by using the *method full.cpl* statement. With this model, the diffusion of the defects is influenced by the diffusion of the dopants by the addition of the joint pair fluxes to the flux terms in the governing equation of the defects. There is now a two-way interaction between the diffusion of dopants and the diffusion of point defects.

4.3.3 (c) Implant Models

For the simulation of the implant process steps, Silvaco's SVDP (Sims Verified Dual Pearson) models are used as all of the implants fell within the parameter ranges of SVDP look-up tables. These implant tables contains dose, energy, tilt, rotation angle and screen oxide thickness dependence for a particular implant species (Silvaco Int., 2001). The transient-enhanced diffusion (TED) is also accounted for by enabling the *unit.damage* parameter. TED refers to diffusion enhancement due to defects introduced during heavy implants and causes enhanced diffusion being several orders of magnitude higher than the intrinsic values (Ozturk *et al.*, 1998). The residual damage from the S/D implants is a major source of such defects. RTP at relatively high temperatures of 900°C or more is very effective at reducing the TED (Rodder *et al.*, 1995). Both NMOS and PMOS simulation uses the *+1* model, as described in Section 3.2.4(c).

4.4 Device Simulations

After remeshing of grid in *DevEdit*, the subsequent generated structure is exported to *ATLAS* which is a software package for device modeling tool. In the following section, descriptions of parameters used in electrical characterization are highlighted. Included are specifications of work function for gates, selection of physical models, specification of interface properties, specification of number of carriers, selection of numerical methods and summary of descriptions on biasing conditions and subsequent *extract* statement. Unless otherwise specified, the descriptions of the design parameters are applicable to both the 0.35 μm and 0.25 μm device simulations. Figure 4.7 illustrate the sequence of device simulations.

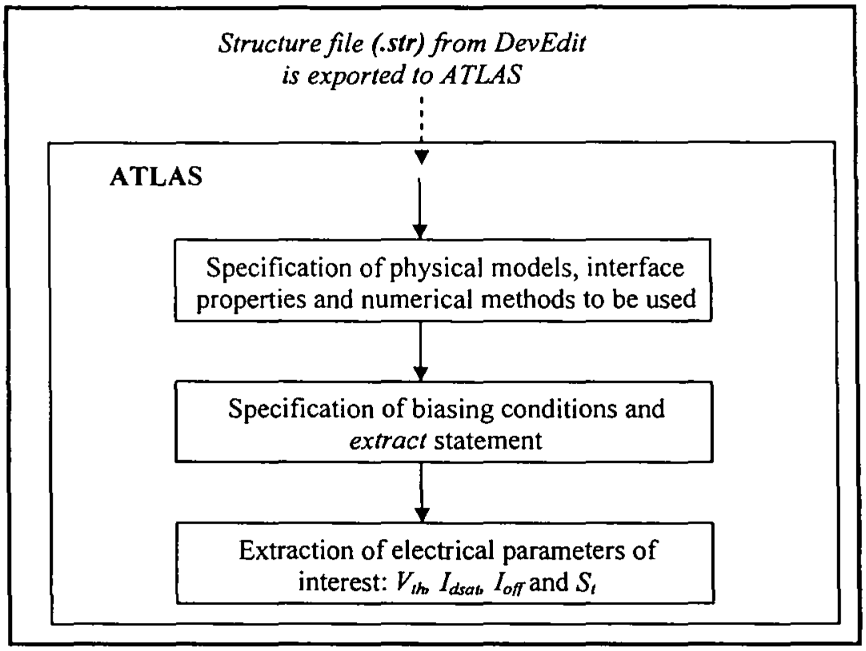


Figure 4.7: Sequences of device simulations in *ATLAS*

4.4.1 Specification of Work Function for Gates

In this work, the dual-doped poly approach is being adopted, whereby the gate is doped by the same dopants that made up the source and drain area (both being a surface-channel device). For NMOS, n^+ dopant is being used while for PMOS, the p^+ dopant is being used to dope the respective gates. In addition, titanium is also being used for silicidation process on the source, drain and gate area. Titanium is chosen because its resistivity of 12-25 μΩ-cm, is the lowest among various metal silicides (Kim *et al.*, 1998). The Titanium Silicide ($TiSi_2$) formed after the silicidation process is then defined as the gate. No work function is assigned to the $TiSi_2$ electrode to ensure that it behaves as an ohmic rather than a Schottky contact. The effective work function of the poly gate will then be correctly calculated from the doping profile in the polysilicon.

4.4.2 Specification of Physical Models

Physical models are specified using the *models* statements. In this work, the physical models adopted for MOS simulation is based on the recommended models by Silvaco Int. (2002) which is the *Lombardi mobility model (cvt)* and *Shockley-Read-*

Hall (srh) recombination model. The *cvt* model enables the modeling of mobility as having a dependence on the transverse field. This is particularly important in MOSFET simulation as it takes into account mobility effects associated with inversion layers. The *srh* recombination model which is also turned on for device simulation is being used in almost all simulations and is based on fixed lifetimes.

4.4.3 Specification of Interface Properties

The *interface* statement is used to define the interface charge density and surface recombination velocity at interfaces between semiconductors and insulators. In this work, the default value which is of 3×10^{10} C-cm^{-2} is being used

4.4.4 Specification of Number of Carriers

The default two-carrier solution is being used in the device simulation. This will cause ATLAS to solve both electron and hole continuity equations.

4.4.5 Specification of Numerical Methods

In this work, a combined algorithm of *Gummel Newton* numerical method is being selected to be used throughout the device simulation. *Gummel* iteration can refine initial guess to a point from which *Newton* iteration can converge. The solver will first start with *Gummel* iterations and then switch to *Newton* to complete the solution. This is the most robust method but consumes higher simulation times.

4.4.6 Specification of Biasing Conditions and Parameter Extractions

In all simulations, the device starts with zero bias on all electrodes. Then, the voltage on each of the electrodes in the device is specified. With that, *ATLAS* can now calculate the current through each electrode. Solutions will be obtained by stepping the biases on electrodes from the initial equilibrium condition and then saving the results using the *log* or *save* statement. The *log* statement has to be inserted in the program before sweeping the bias on the gate electrode (for generating I_d - V_g curve), or on the drain electrode (for generating I_d - V_d curves). After the *logfile* are saved, they can then be displayed in *Tonyplot* by specifying the associated filename. To extract the device parameters, the *extract* command is provided within the *Deckbuild* environment. Details on each of the biasing conditions specified in *ATLAS* to extract the parameters of V_{th}, I_{dsat}, I_{off} and S_t will be further described in Chapter 5. The results of the extraction will be displayed in the run-time output and will be stored in the file *results.final*.

4.5 Summary

In this simulation design work, a "two-step" design approach is proposed and adopted for the downscaling process of 0.35 µm CMOS transistor to 0.25 µm. The first design step involves conversion of the 0.35 µm physical data to simulated environment while the second design step involves miniaturization to a smaller 0.25 µm CMOS transistor geometries by using a scaling factor, κ of 1.4. In this work, the transistor is simulated from starting wafer until the formation of contact holes. Descriptions of the parameters adopted in the simulation such as the grid definitions, the modeling of substrate depth and the selection of the process and device physical models are also discussed in details.

CHAPTER 5

RESULTS AND DISCUSSIONS

This chapter reports on the results obtained from the simulations of the 0.35 µm and 0.25 µm CMOS transistor. Device parameters extracted from I-V curves comprising of threshold voltage *(V_{th})*, saturation current *(I_{dsat})*, off-state leakage current *(I_{off})* and subthreshold swing (S_t) are reported. Descriptions of the method used for extracting each particular parameter and the related biasing conditions are also included.

5.1 Results for 0.35 µm MOS Simulations

The results obtained from the simulations of 0.35 µm MOS transistor are evaluated against the actual measured values and device design specification stated in several literatures. Summary of the comparisons made is tabulated in Table 5.1.

Table 5.1: Comparison of electrical parameters obtained from the 0.35 µm MOS simulation with actual measured values and device design specification reported

	V_{th} (V)		I_{dsat} (µA/µm)		I_{off} (pA/µm)		S_t (mV/dec)	
	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS
Simulation values	0.63	-0.49	552	251	0.1	0.5	83	85
NTRS, 1997	Not stated		600	280	Not stated		Not stated	
Mat Husin *et al.*, 2000	0.55	-0.55	550	260	<10	<10	<90	<90
Horvath *et al.*, 2005	0.67	-0.57	Not stated		12	7.6	85	90
Silvaco Resource Ctr., 1994	Not stated		Not stated		<10	<10	Not stated	

5.1.1 Results on Threshold Voltage *(V_{th})* Extraction

In this work, the *linear extrapolation method* is being used to extract the V_{th} from the simulation. When using the *linear extrapolation method*, the I_d is measured as a function of V_g at a low V_d of typically 50-100 mV to ensure operation in the linear

MOSFET (Schroder, 1990). The I_d versus V_g curve is then extrapolated to $I_d=0$ and the V_{th} is determined from the extrapolated or intercept gate voltage V_{gSi} by the following equation (Schroder, 1990):

$$V_{th} = V_{gSi} - \left(\frac{V_d}{2}\right) \tag{5.1}$$

For extraction of V_{th} in this simulation, the I_d-V_g curve is first plotted and the sequence of solve statements are then set to ramp the gate bias with the drain voltage, V_d = 0.05 V for NMOS simulation and V_d = -0.05 V for PMOS simulation. Solutions are obtained in intervals from 0 V to 1.5 V for the gate voltage, V_g of NMOS and from 0 V to -1.5 V for V_g of PMOS.

The V_{th} obtained from the 0.35 μm simulation is 0.63 V for NMOS and -0.49 V for PMOS as illustrated in Figure 5.1 and 5.2 respectively. For NMOS, the simulation value obtained is found to be much closer to results obtained by Horvath *et al.* (2005) with a difference of only 6%. However, if compared to the specification set in Mat Husin *et al.* (2000), this simulation value is found to be 14% higher. For PMOS, the simulation value obtained is found to be 10% and 14% lower when compared to values reported in Mat Husin *et al.* (2000) and Horvath *et al.* (2005) respectively. These variations could be due to the fact that different references are adopting different methods and parameters in producing the 0.35 μm CMOS transistor. As these methods and parameters are not disclosed explicitly, direct comparisons and explanations on the differences in the values obtained could not be accounted for.

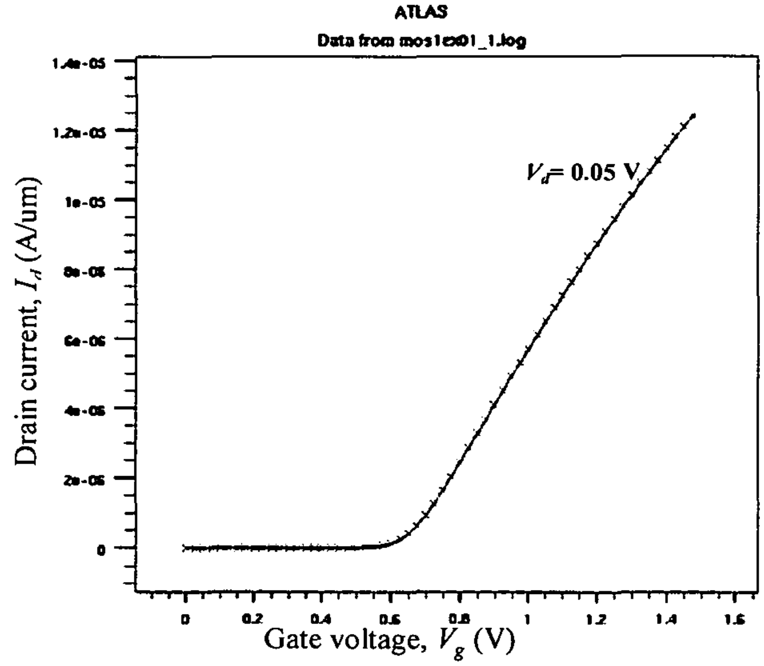


Figure 5.1: Drain current, I_d versus gate voltage, V_g for 0.35 μm NMOS transistor in the linear region with $V_d = 0.05$ V

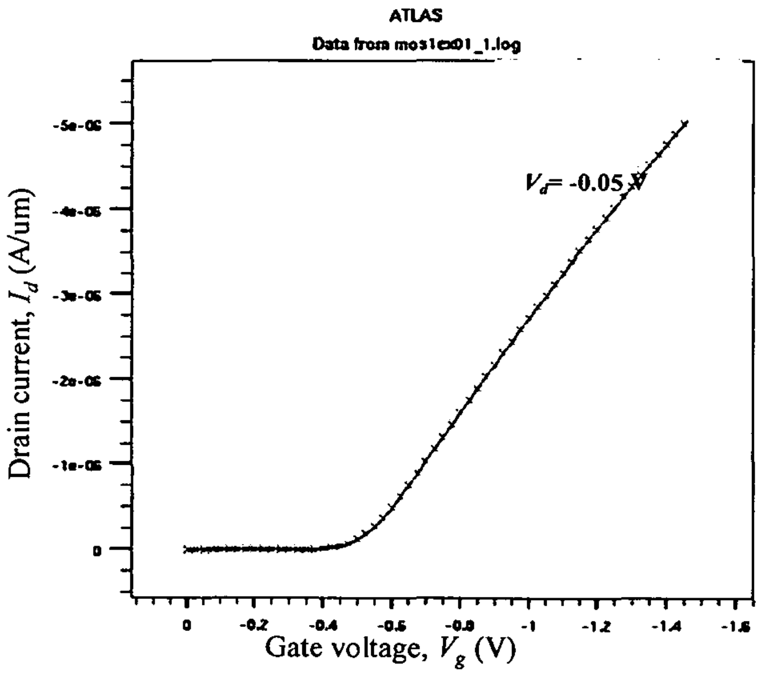


Figure 5.2: Drain current, I_d versus gate voltage, V_g for 0.35 μm PMOS transistor in the linear region with $V_d = -0.05$ V

It should be noted that even though a low V_{th} is desirable, too low V_{th} will lead to excessive leakage current. I_{off}. This is because the transistor turn-off rate is largely driven by thermally activated diffusion and is independent of power supply voltage and channel length. For every 0.1 V reduction of V_{th}, the I_{off} would further increase by about 10 times and thus, a minimum V_{th} of 0.3-0.4 V is required for room temperature CMOS devices, below which the standby power due to I_{off} becomes prohibitively high (Taur *et al.*, 1997). Figure 5.3 shows an exponential relationship between I_{off} and V_{th}.

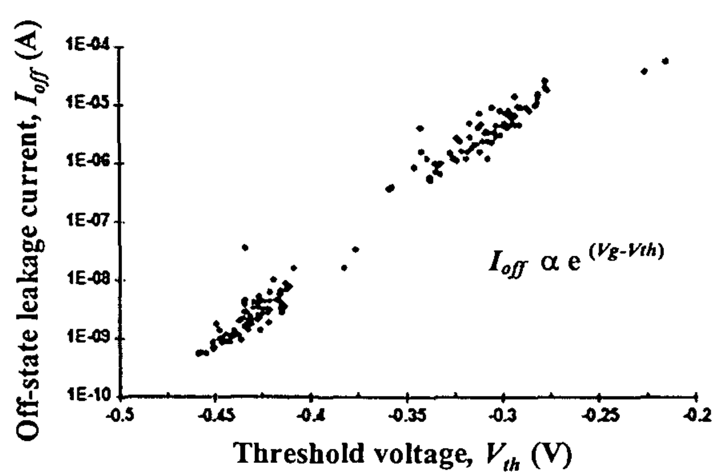


Figure 5.3: Exponential relationship between off-state-leakage current, I_{off} and threshold voltage, V_{th} (Reproduced from Purdue Univ., 2003)

5.1.2 Results on Saturation Current Extraction

Saturation current is also known as I_{dsat} or I_{on}. It refers to how much current is carried in the "ON" state of a MOSFET and provides a measure of current drive capability. For extraction of I_{dsat} in this work, three I_d-V_d curves are required at different V_g. For each of the V_g where V_g = 1.1 V, 2.2 V and 3.3 V for NMOS and V_g = -1.1 V, -2.2 V and -3.3 V for NMOS, a solution with V_d = 0 V is simulated and the results is saved to a solution file. A *log file* is opened and the ramp of V_d is set. The three I_d-V_d curves are then overlaid in *Tonyplot*. At the end of the simulation, the *extract* statement is used to measure the maximum drain current, I_d.

As can be seen from Figure 5.4 and 5.5, I_{dsat} obtained is 552 μA/μm for NMOS and 251 μA/μm for PMOS respectively. A higher value of I_{dsat} is desirable as it implies a faster speed device. For NMOS, the I_{dsat} value obtained from the simulation is found to be closer to specification set in Mat Husin *et al.* (2000) by a difference of only 0.36%. However, when compared to the specification set in NTRS (1997), this simulation value exhibits slightly larger difference of 8%. As for PMOS, lower values of I_{dsat} by 3.5% and 10% are obtained from the simulation as compared to specification set in Mat Husin *et al.* (2000) and NTRS (1997) respectively.

As evident in Figure 5.4 and 5.5, true saturation of I_d can be observed without any drastic increase as compared to Figure 2.5 (a) in Chapter 2. This lead to the confirmation that the 0.35 μm MOS transistor design does not experience punch through phenomenon.

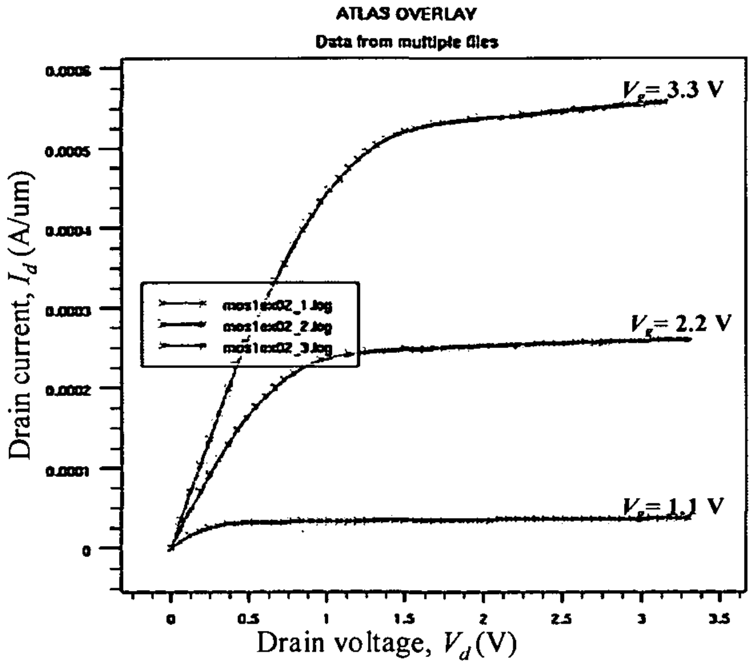


Figure 5.4: Plot of Drain current, I_d versus drain voltage, V_d for 0.35 μm NMOS

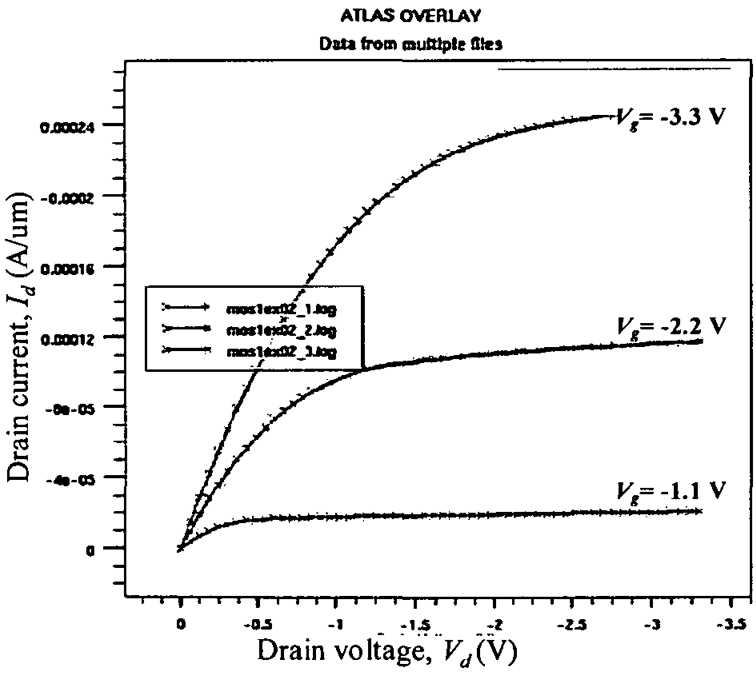


Figure 5.5: Plot of Drain current, I_d versus drain voltage, V_d for 0.35 μm PMOS

5.1.3 Results on Off-state Leakage Current Extraction

Ideally, I_d should fall to zero very quickly after V_g falls below V_{th}. However, below the V_{th}, the transistor is partially conducting and the region is referred to as weak inversion. The corresponding I_d that flows in this weak inversion region is referred to as off-state leakage current, I_{off}. I_{off} refers to how much leakage current flows in the "OFF" state of a MOSFET and it determines the turn-off characteristics of the device. According to Iwai & Ohmi (2000), for small gate length, the high potential region near the drain touches the source in a deeper place where the gate bias cannot control the potential resulting in a leakage electron current from source to drain via the space charge region, even in the "off" state. An increase in I_{off} implies that the gate doesn't have full control of turning the device off. With the goal to realize fast circuits, I_{off} should be made as low as possible.

The off-state leakage current, I_{off} parameter can be extracted from the subthreshold transfer characteristics. A simple $logI_d$-V_g curve is generated with V_d = 3.3 V for NMOS and with V_d = -3.3 V for PMOS. The V_g is ramped from 0 V to 3.3 V in 0.1 V incremental steps for NMOS and from 0 V to -3.3 V in -0.1 V incremental steps for PMOS. The I_d which corresponds to V_g = 0 V is taken as the off-state leakage current, I_{off}. Refer Figure 5.6 and 5.7 for the I_{off} extracted for the 0.35 μm NMOS and PMOS which is of 0.1 pA/μm and 0.5 pA/μm respectively.

The I_{off} obtained in the simulation of both the 0.35 μm NMOS and PMOS is found to be 93%-99% lower as compared to specification set in Mat Husin *et al.* (2000), Horvath *et al.* (2005) and Silvaco Resource Ctr. (1994). A much smaller values of I_{off} obtained in the simulation is desirable as it demonstrates that the leakage current that flows in the off-state operation of the transistor is very small.

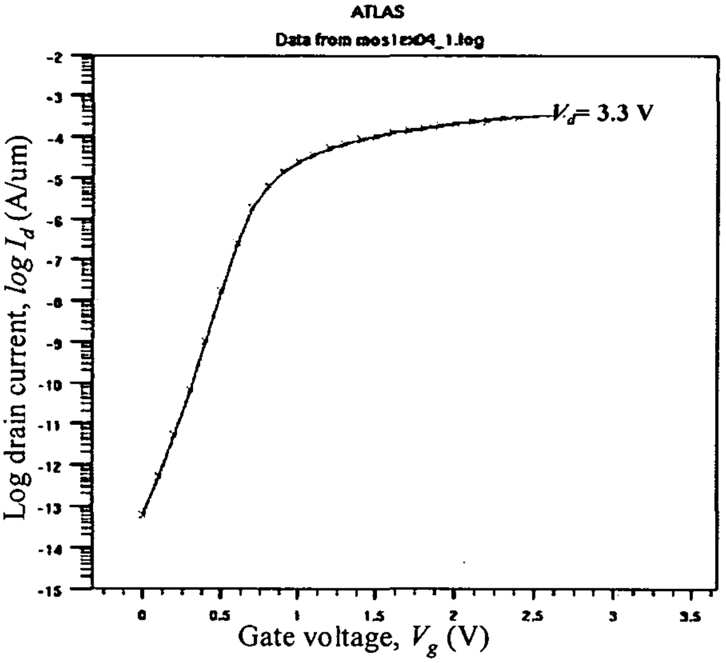


Figure 5.6: Plot of log drain current, $logI_d$ versus gate voltage, V_g with $V_d = 3.3$ V for 0.35 μm NMOS

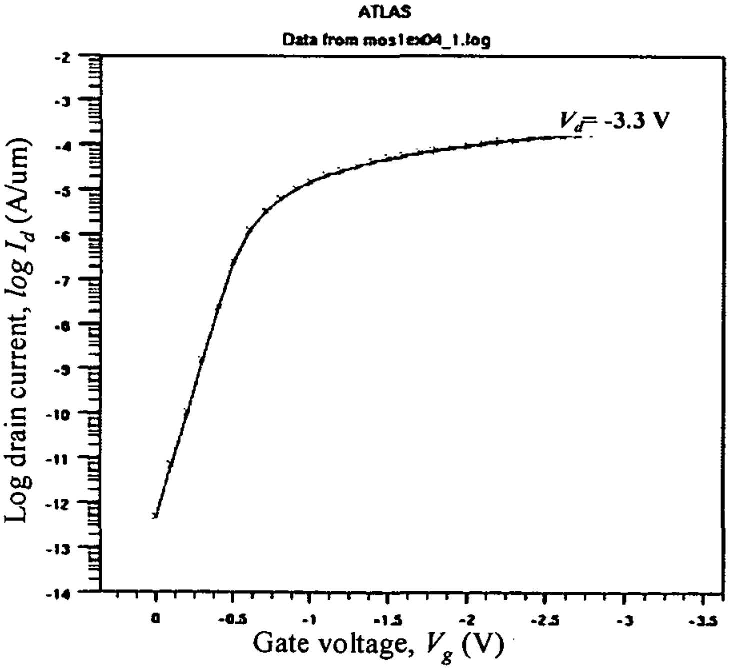


Figure 5.7: Plot of log drain current, $logI_d$ versus gate voltage, V_g with $V_d = -3.3$ V for 0.35 μm PMOS

5.1.4 Results on Subthreshold Swing Extraction

Subthreshold swing, S_t is a measure of how much of a change in V_g (below V_{th}) is required to change the off-current in the device by a decade (magnitude of 10 on a logarithmic scale). The usual subthreshold plot is $logI_d$ versus V_g for $V_d > kT/q$. The inverse of the slope is then defined as the S_t (Wolf, 2000).

$$S_t = \frac{dV_g}{d(\log I_d)} \tag{5.2}$$

If a MOSFET exhibits a steep decline in I_d as V_g is decreased below V_{th}, its S_t value will be small. A small value of S_t (meaning a steep decline in I_d as V_g is decreased below V_{th}) is desirable as it implicate that only a small reduction of V_g below V_{th} can effectively turn off the device, whereas a large S_t value (meaning a gradual slope in $logI_d$ versus V_g) implicates that a significantly large I_d may still flow in the OFF state when $V_g = 0$ V (Wolf, 1995).

In this simulation work, S_t is extracted from a $logI_d$-V_g curve with $V_d = 0.05$ V for NMOS and $V_d = -0.05$ V for PMOS simulation. The V_g is ramped from 0 V to 3.3 V in 0.1 V incremental steps for NMOS and from 0 V to -3.3 V in -0.1 V incremental steps for PMOS. The reciprocal of the maximum slope of the curve V_g versus $logI_d$ is then defined as the S_t. Refer Figure 5.8 to 5.9 for S_t extracted for NMOS and PMOS respectively.

The S_t value obtained in the simulation work of 0.35 μm NMOS and PMOS is between 83-85 mV/dec which means that the subthreshold current drop by a factor of 10 for a reduction in V_g of 83-85 mV. This is well below the specification set in Mat Husin *et al.* (2000) and values reported in Horvath *et al.* (2005). The differences of only 2%-8% as compared to reported literatures implies a transistor design with good turn-off characteristics.

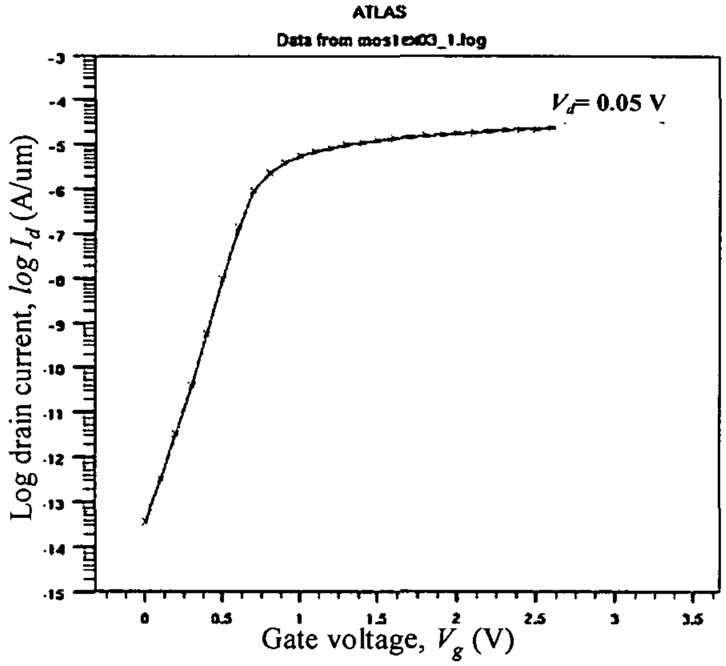


Figure 5.8: Plot of log drain current, $logI_d$ versus gate voltage, V_g with V_d = 0.05 V for 0.35 μm NMOS

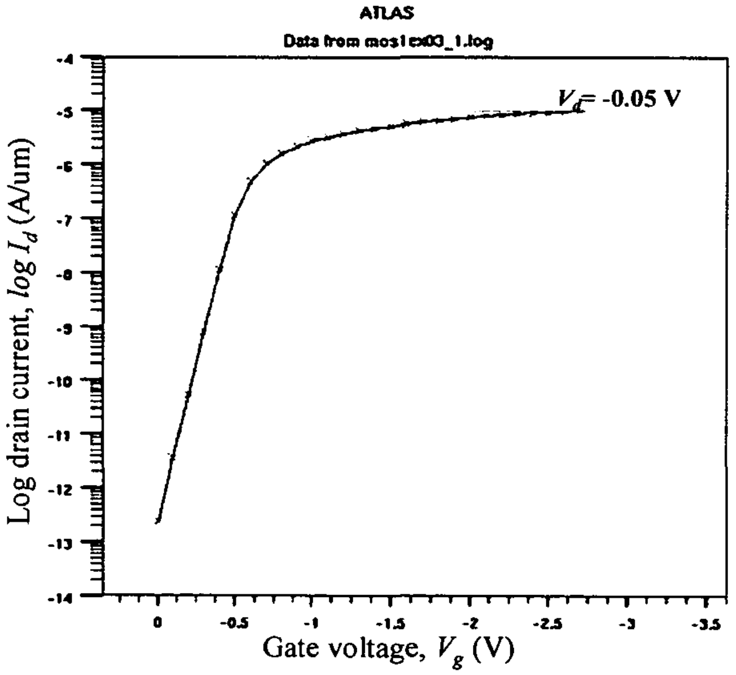


Figure 5.9: Plot of log drain current, $logI_d$ versus gate voltage, V_g with V_d = -0.05 V for 0.35 μm PMOS

5.2 Results for 0.25 μm MOS Simulations

Upon successful conversion of the 0.35 μm MOS physical data to simulated environment, the process of downscaling to geometries of 0.25 μm begun. The results obtained from the simulations of 0.25 μm MOS transistor are again evaluated against the actual measured values and device design specification stated in several literatures for that particular technology. Summary of the comparisons made is tabulated in Table 5.2.

Table 5.2: Comparison of electrical parameters obtained from the 0.25 μm MOS simulation with actual measured values and device design specification reported

	V_{th} (V)		I_{dsat} (μA/μm)		I_{off} (pA/μm)		S_t (mV/dec)	
	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS
Simulation values	0.51	-0.40	296	181	0.4	2.0	82	80
NTRS, 1997	Not stated		600	280	Not stated		Not stated	
Chang *et al.*, 1992	0.40	-0.40	Not stated		Not stated		78	80
Asai & Wada, 1997	Not stated		500-600	250-300	Not stated		80	80
Woerlee *et al.*, 1992	0.60	-0.60	Not stated		<10	<10	Not stated	

5.2.1 Results on Threshold Voltage (V_{th}) Extraction

The method used for extracting the V_{th} is the same as described in section 5.1.1. For the 0.25 μm CMOS, the V_{th} obtained is 0.51 V for NMOS and -0.40 V for PMOS as illustrated in Figure 5.10 and 5.11 respectively. For NMOS, the simulation value obtained is found to be much closer to specification set in Woerlee *et al.* (1992) with a difference of 15%. However, a higher V_{th} difference of 28% is observed when compared to values obtained by Chang *et al.* (1992). For PMOS, the simulation value obtained is found to be in good agreement with values obtained by Chang *et al.* (1992). However, a high V_{th} difference of 33% is observed when compared to specification set in Woerlee *et al.* (1992). As stated previously, the differences in the values obtained in simulation as compared to various reported literatures are very much depended on the various techniques and parameters adopted in each of the transistor fabrication sequences. As this information is not disclosed, the differences in the values obtained could not be accounted for.

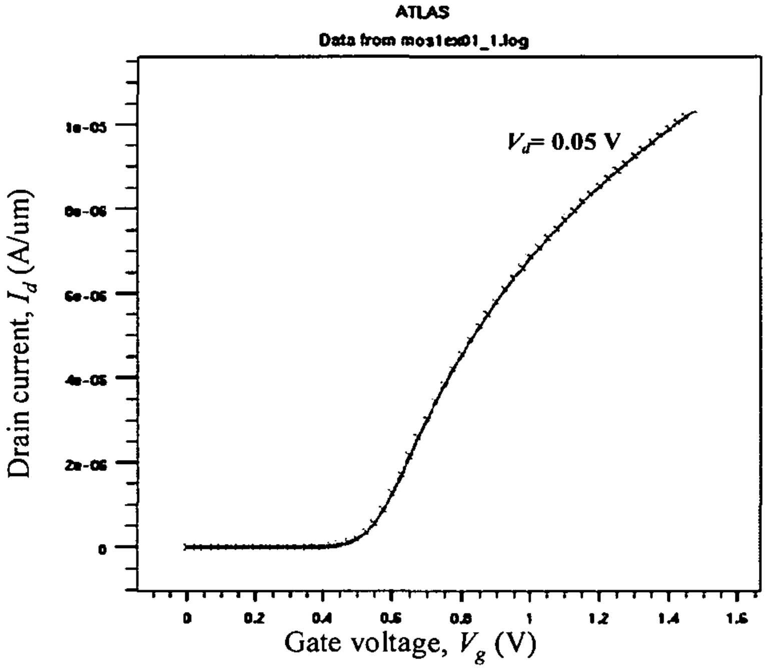


Figure 5.10: Drain current, I_d versus gate voltage, V_g for 0.25 μm NMOS transistor in the linear region with $V_d = 0.05$ V

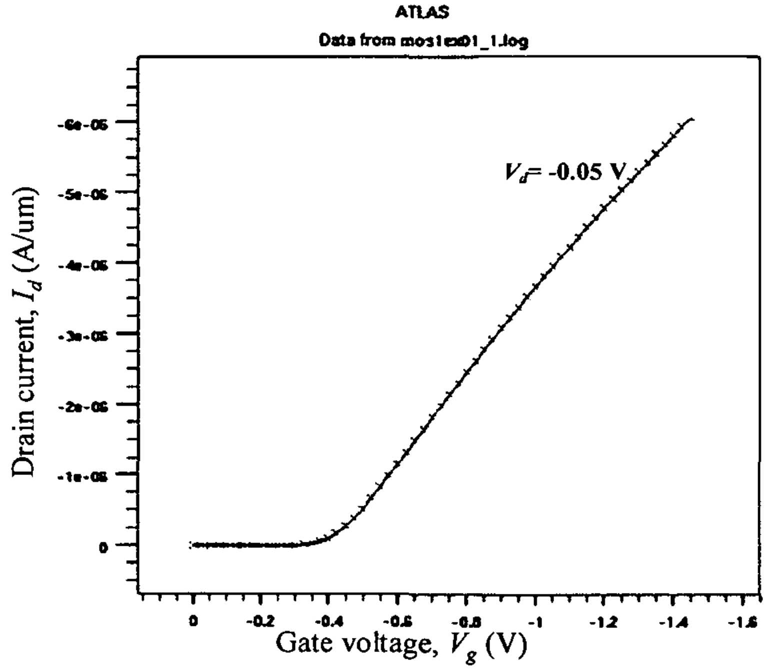


Figure 5.11: Drain current, I_d versus gate voltage, V_g for 0.25 μm PMOS transistor in the linear region with $V_d = -0.05$ V

5.2.2 Results on Saturation Current Extraction

The same sequence as of 0.35 µm is applied to extract the I_{dsat}. However, five I_d-V_d curves are required at different V_g. For NMOS simulation, the V_g used is $V_g = 0.5$ V, 1.0 V, 1.5 V, 2.0 V and 2.5 V while for PMOS simulation, the V_g used is $V_g = -0.5$ V, -1.0 V, -1.5 V, -2.0 V and -2.5 V. From the simulated Figure 5.12, it can be seen that I_{dsat} obtained for 0.25 µm NMOS is 296 µA/µm at $V_d = V_g = 2.5$ V. Meanwhile, Figure 5.13 denotes that for 0.25 µm PMOS, I_{dsat} obtained is 181 µA/µm at $V_d = V_g = -2.5$ V.

The results of I_{dsat} obtained from the 0.25 µm simulation are 296 µA/µm for NMOS and 181 µA/µm for PMOS respectively. For NMOS, the I_{dsat} value obtained from the simulation is found to be 40%-50% lower as compared to specification set in NTRS (1997) and Asai & Wada (1997). A lower I_{dsat} value is also obtained from the simulation of PMOS when compared to specification set in NTRS (1997) and Asai & Wada (1997) by 28%-40%. A low I_{dsat} obtained from the downscaling process could be due to the fact that direct scaling is being performed in the design of 0.25 µm transistors from the 0.35 µm geometry transistor without alterations on the existing structure. It has been reported that an introduction of halo implant to a transistor structure can lead to higher I_{dsat} as with a given V_{th}, the halo device has a more abrupt drain-channel junction and higher channel mobility than a non-halo device (Intel Tech. Journal, 1998).

Again, as evident in Figure 5.12 and 5.13 respectively, the downscale device of 0.25 µm MOS transistor exhibits no punch through phenomenon as true saturation of I_d can be observed without any drastic increase, which is a desired feature.

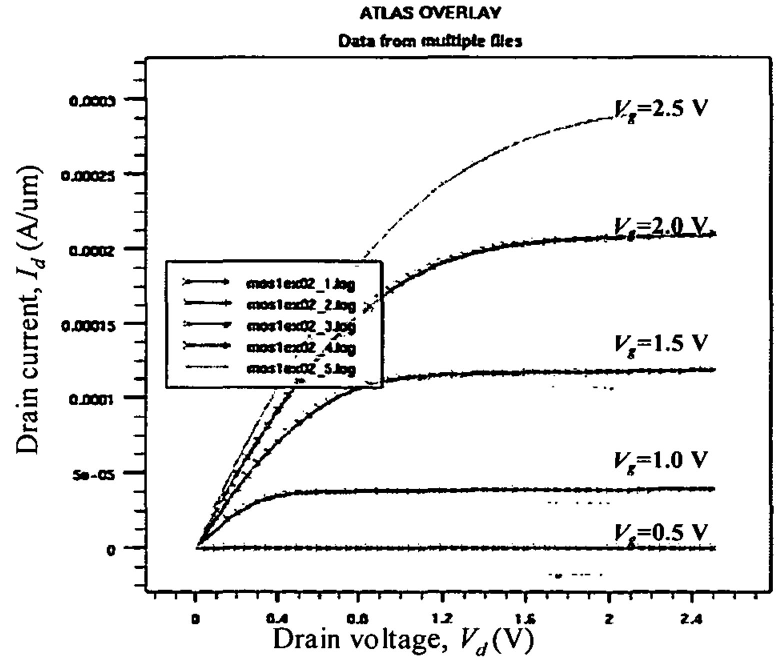


Figure 5.12: Plot of Drain current, I_d versus drain voltage, V_d for 0.25 μm NMOS

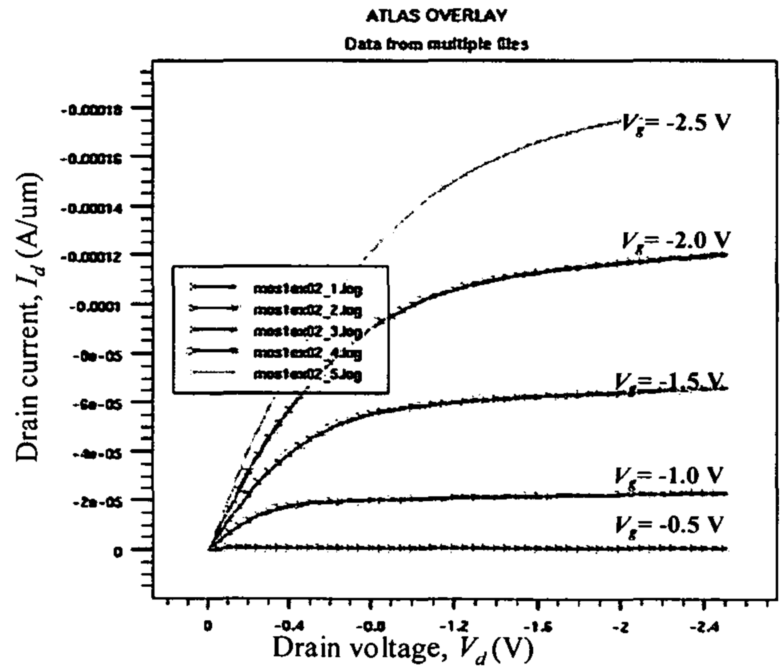


Figure 5.13: Plot of Drain current, I_d versus drain voltage, V_d for 0.25 μm PMOS

5.2.3 Results on Off-state Leakage Current Extraction

The off-state leakage current, I_{off} parameter can be extracted from the subthreshold transfer characteristics. A simple $logI_d$-V_g curve is generated with V_d = 2.5 V for NMOS and with V_d = -2.5 V for PMOS simulation. The V_g is ramped from 0 V to 2.5 V in 0.1 V incremental steps for NMOS and from 0 V to -2.5 V in -0.1 V incremental steps for PMOS. The I_d which corresponds to V_g = 0 V is then taken as the off-state leakage current, I_{off}. Refer Figure 5.14 and 5.15 for the I_{off} extracted for the 0.25 μm NMOS and PMOS respectively.

In this work, the I_{off} obtained is between 0.4–2.0 pA/μm for both the 0.25 μm NMOS and PMOS simulation. This result which is of 80-96% lower is in good agreement with device design specifications stated in Woerlee *et al.* (1992) which specify that I_{off} should be < 10 pA/μm for that particular CMOS technology.

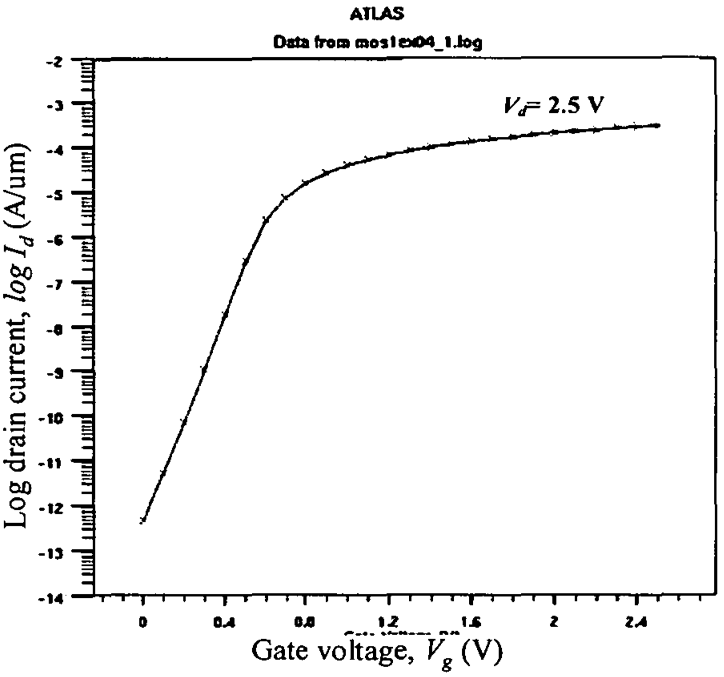


Figure 5.14: Plot of log drain current, $logI_d$ versus gate voltage, V_g with V_d = 2.5 V for 0.25 μm NMOS

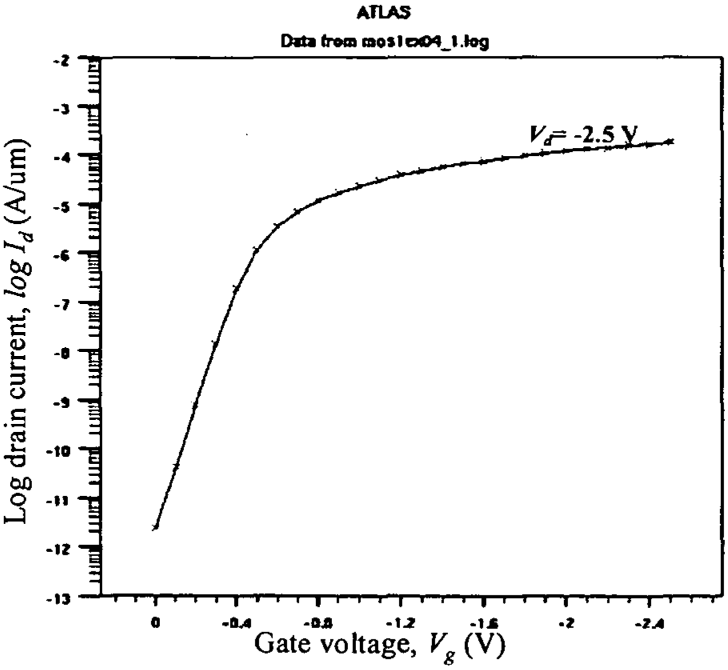


Figure 5.15: Plot of log drain current, $logI_d$ versus gate voltage, V_g with V_d = -2.5 V for 0.25 μm PMOS

5.2.4 Results on Subthreshold Swing (S_t)

In this simulation work, S_t is extracted from a $logI_d$-V_g curve with V_d = 0.05 V for NMOS and with V_d = -0.05 V for PMOS. The V_g is ramped from 0 V to 2.5 V in 0.1 V incremental steps for NMOS and from 0 V to -2.5 V in -0.1 V incremental steps for PMOS. The reciprocal of the maximum slope of the curve V_g versus $logI_d$ is then defined as the S_t.

From the simulated Figure 5.16, it can be seen that S_t obtained for NMOS is 82 mV/dec. Meanwhile, Figure 5.17 denotes that for PMOS, S_t obtained is 80 mV/dec. The S_t value obtained in the simulation work of 0.25 μm is between 80-82 mV/dec, which means that the subthreshold current drop by a factor of 10 for a reduction in V_g of 80-82 mV. The simulation values with less than 5% difference are found to be in good agreement with S_t value set in Asai & Wada (1997) and values reported in Chang *et al.* (1992).

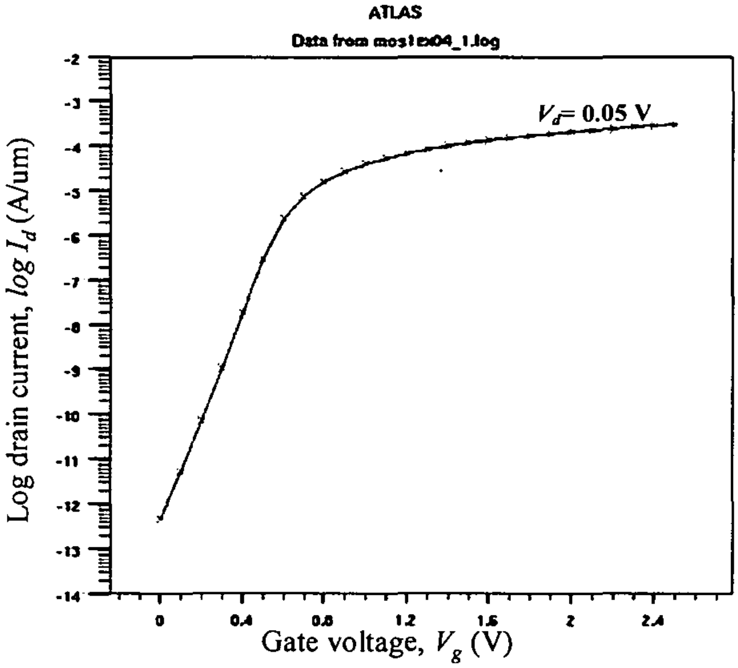


Figure 5.16: Plot of log drain current, $logI_d$ versus gate voltage, V_g with V_d = 0.05 V for 0.25 μm NMOS

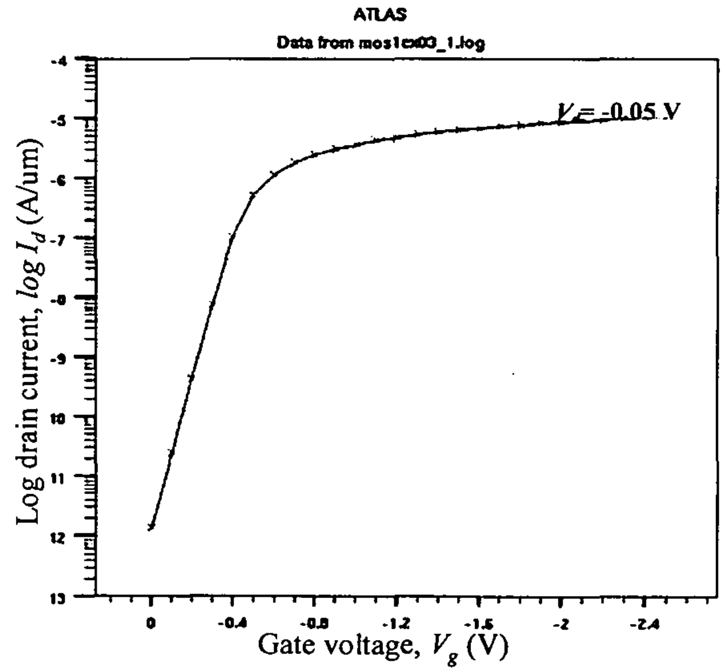


Figure 5.17: Plot of log drain current, $logI_d$ versus gate voltage, V_g with V_d = -0.05 V for 0.25 μm NMOS

5.3 Results Comparison between 0.35 μm and 0.25 μm Simulated data

The following section describes the results obtained when the transistor is being scaled from the 0.35 μm to 0.25 μm CMOS technology using scaling factor,κ of 1.4. Summary of results obtained by simulation of the 0.35 μm and 0.25 μm CMOS technology is tabulated in Table 5.3.

Table 5.3: Comparison of electrical parameters obtained in simulated environment for between 0.35 μm and 0.25 μm CMOS technology

	V_{th} (V)		I_{off} (pA/μm)		S_t (mV/dec)		I_{dsat} (μA/μm)	
	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS	NMOS	PMOS
Values obtained from 0.35 μm MOS simulation	0.63	-0.49	0.1	0.5	83	85	552	251
Values obtained from 0.25 μm MOS simulation	0.51	-0.40	0.4	2.0	82	80	296	181

It is expected that as technology is scaled down, device performance and density of an Integrated Circuit (IC) will be improved. From the above results, it can be seen that as technology is scaled down from 0.35 μm to 0.25 μm, the values of V_{th} will also decreased as expected when miniaturization is being performed. For NMOS, the V_{th} values show a reduction from 0.63 V to 0.51 V while for PMOS, the V_{th} values show a reduction from -0.49 V to -0.4 V for the shift from the 0.35 μm technology to 0.25 μm technology. With each technology generation, lower V_{th} value is desirable as it implies that only a small V_g is needed to trigger the transistor operation. However, as described in the previous section, too low of V_{th} will lead to an excessive leakage current, I_{off}. Therefore, during miniaturization process, I_{off} is another important parameter that should be looked into during the transistor design process. Proper trade-off should exist in between obtaining the desired value of V_{th} and at the same time maintaining acceptable tolerable values of I_{off}. In this downscaling process, small values of I_{off} of between 0.1-2.0 pA/μm have been successfully maintained with the reduction in V_{th}. This value which is well below the specification set by Woerlee *et al.* (1992) indicates that only a small value of leakage current flows when the transistor is not in operation. Another indicator of the turn-off characteristics of a MOSFET is the subthreshold swing, S_t. Throughout the 0.35 μm and 0.25 μm simulation, a small value of S_t of between 80-85 mV/dec has also been successfully maintained. These values show that the subthreshold current drop by a factor of 10 for a reduction in V_g of 80-85 mV, which is in good agreement with values reported and specification set in Mat Husin *et al.* (2000), Horvath *et al.* (2005), Chang *et al.* (1992) and Asai & Wada (1997). As the transistor is scaled down to much smaller dimension, higher value of saturated drain current, I_{dsat} is expected to be obtained as the carrier is now traveling much shorter distances between the source and drain. However, a lower value of I_{dsat} is obtained when the transistor is scaled down from 0.35 μm to 0.25 μm which is in contradiction as per expected. As explained previously, this could be due to the fact that direct scaling is being performed on the 0.35 μm transistor without alterations on the existing structure. Further optimization work is needed to obtain higher values of I_{dsat}, one of which is an introduction of halo implant (Intel Tech. Journal, 1998). In both the 0.35 μm and 0.25 μm simulations, it can be seen that the I_{dsat} value of PMOS is lower than its NMOS counterpart. This is due to the fact that mobility of holes, μ_p is less than mobility of electrons, μ_n.

5.4 Summary

In both the 0.35 μm and 0.25 μm CMOS simulation, the values obtained are compared to several reported literatures where differences in the values obtained are discussed. It is found that in general, the simulation values obtained for both technology is within the range of values reported by other literatures. For the simulation of 0.35 μm NMOS and PMOS, a difference of 6%-14% is observed for V_{th} when compared to values obtained in other literatures. Meanwhile, a difference of between 0.4% to 10% is obtained for comparison made on values of I_{dsat} obtained in simulation and other reported works. For simulation results on I_{off} and S_t, both parameters exhibits a much better value as compared to the maximum allowable specification set in other reported work. I_{off} values is about 93%-99% lower than the maximum allowable specification while the S_t values only shows difference of between 2%-8% as compared to specification. For the results on the 0.25 μm NMOS and PMOS simulations, the V_{th} values obtained from the simulation differ from the reported literatures by about 15%-33%. However, for results of I_{dsat}, the values obtained is much lower than other reported work by 28%-50%. For results on I_{off} and S_t, the values obtained is well below the maximum allowable specification set in other reported work. I_{off} values is about 80%-96% lower than the maximum allowable specification while the S_t values only shows difference of less than 5% as compared to specification. It should be noted that detailed discussions on the differences obtained in the simulation values as compared to the actual measured data and device design specification could not made as the methods and parameters used for designing the transistor varies and is not explicitly disclosed. However, the results obtained (except for I_{dsat} of 0.25 μm CMOS) falls in the range of accepted values of 0.35 μm and 0.25 μm CMOS technology.

CHAPTER 6

CONCLUSION

This chapter summarizes and concludes the research work, together with recommendations for future studies.

6.1 Conclusion

In this work, a downscale device of 0.25 μm CMOS transistor has been designed and characterize from the initial 0.35 μm CMOS technology by means of Silvaco ATHENA and ATLAS simulation tools. Besides being a cheaper alternative as compared to an expensive and time consuming fabrication process, another important benefit of using the TCAD Tools is that it enable a quick prediction of the electrical characteristics of a device design. Process recipe of 0.35 μm CMOS technology developed by UC Berkeley Microfabrication Lab (Horvath *et al.*, 2005) serves as the design basis. A "two-step design" approach is proposed in this work to study the feasibility of miniaturization process by scaling method. A scaling factor, κ of 1.4 (derived from direct division of 0.35 with 0.25) is adopted for selected parameters. The first design step involves conversion of physical data of 0.35 μm CMOS technology to simulation environment. The results obtained are then used for subsequent miniaturization process which constitutes the second design step. Here, downscaling to a smaller geometry of 0.25 μm CMOS transistor is carried out and the subsequent electrical characterization is conducted to evaluate its performance.

The validation of the simulation results obtained is being made through comparison to actual measured data and also device design specification stated in other research work. It can be seen that for the 0.35 μm simulations, the electrical characteristics obtained is comparable to those obtained in other reported literatures. Results on V_{th} obtained shows a difference of between 6%-14% as compared to values obtained in other reported work while values of I_{dsat} obtained show a difference of between 0.4% to 10%. For simulation results on I_{off} and S_t, both parameters exhibits a much better value as compared to the maximum allowable specification set in other reported work.

I_{off} values is about 93%-99% lower than the maximum allowable specification while the S_t values only shows 2%-8% differences as compared to specification.

In the case of 0.25 μm simulations, the electrical parameters obtained is also comparable to the reported works and device design specifications, except for the results of I_{dsat} which is lower than expected. For the V_{th} results, the values obtained from the simulation differ from the reported literatures by about 15%-33%. However, for results of I_{dsat}, the values obtained are much lower than other reported work by 28%-50%. This could be due to the fact that direct scaling is being performed in the design of 0.25 μm transistor from the 0.35 μm geometry transistor without alterations on the existing structure. It has been reported that an introductions of halo implant to a transistor structure can lead to higher I_{dsat} as with a given V_{th}, the halo device has a more abrupt drain-channel junction and higher channel mobility than a non-halo device (Intel Tech. Journal, 1998). For results on I_{off} and S_t, the values obtained are well below the maximum allowable specification set in other reported work. I_{off} values is about 80%-96% lower than the maximum allowable specification while the S_t values only shows less than 5% differences as compared to specification. Even though comparison can be made between the simulated values and the actual measured data and device design specification, detailed descriptions on the reason for the differences could not be accounted for. This is as different literatures adopt different methods and parameters in designing their particular transistor where these detailed information is not made available to the public.

From the values obtained in simulation, it is concluded that the use of numerical simulation software provides a pretty accurate results in predicting the main features of the transistors. To ensure successful simulation of the 0.35 μm technology in the first design step, the basis of some of the important parameters that need to be studied in depth would be the correct substrate depth modeling, correct choice of physical models, correct specification of gates work function and others. These constitute the critical part of the transistor design. Electrical parameters obtained from the simulation which is found to be comparable to other reported work and device design specification shows that a successful simulation of 0.35 μm technology has been developed. The know-how on the important aspects of simulation that need to be properly understood and accounted for in order to achieve a successful simulation is acquired during this process.

The acquisition of this technical know-how has provided the capability to perform device miniaturization to 0.25 μm CMOS technology, which forms the main contribution of this work. This miniaturization process involves scaling of four parameters namely of physical gate length *(L_g)*, gate oxide thickness *(t_{ox})*, threshold-adjust implant doping and the supply voltage *(V_{dd})*. Electrical parameters obtained are found to be comparable to other reported literature and device design specification except for the value of I_{dsat} which is much lower. This implies that even though a direct scaling approach is feasible in performing device miniaturization, it is not sufficient. Rather, new structures or techniques need to be incorporated during the design to obtain an optimized performance as discussed previously.

6.2 Recommendations for Future Works

As the miniaturization trends is moving towards the deep submicron regime with fabrication of transistor of less than 100 nm gate length, the results obtained in this work can be used as a reference basis for designing transistors with a much smaller geometries, provided that the related process recipe is acquired in the process. Conventional CMOS transistor with 30 nm physical gate length developed by Intel has been reported to be using aggressively scaled junctions, polysilicon gate electrode, gate oxide and Ni silicide (Chau *et al.*, 2000). Besides the conventional CMOS technology, there are also alternative device architectures waiting to be explored such as Silicon-on-Insulator (SOI) device, $Si_{1-x}Ge_x$ device and Dynamic V_{th} device. In the SOI device, a thin insulating layer, such as silicon oxide or glass, is placed between a thin layer of silicon and the silicon substrate. This architecture helps to reduce the amount of electrical charge that the transistor has to move during a switching operation, thus making it faster and allowing it to switch using less energy. SOI chips can be as much as 15 percent faster and use 20 percent less power than today's bulk-CMOS-based chip (Freescale Semicon., 2006). However, SOI chips tend to cost more than the standard silicon chips and therefore, it has been used primarily for high-end applications. The next alternative device architecture is the $Si_{1-x}Ge_x$ device where the transistor performance is enhanced by fabricating it in a $Si_{1-x}Ge_x$

channel. This $Si_{1-x}Ge_x$ channel region has been shown to increase hole mobility (Ismail *et al.*, 1994). For low supply voltage operation (<0.6 V), a dynamic threshold voltage MOS device (DTMOS) has been proposed (Assaderaghi *et al.*, 1994 and Shibata *et al.*, 1998). A DTMOS is formed by connecting the gate to the well. This connection causes the threshold voltage of the device to be lowered during switching, and thus the transistor drive current is increased. From the above, it can be concluded that there is still a wide area waiting to be explored in the domain of MOS devices, all with the aim to produce a high performance device.

REFERENCES

Arora N., "MOSFET Models for VLSI Circuit Simulation", Springer, 1993

Asai S., and Wada Y., "Technology Challenges for Integration Near and below 0.1μm", *Proc. of the IEEE,* Vol. 85, No. 4, April 1997

Assaderaghi F., Sinitsky D., Parke S., Bokor J., Ko P.K., and Hu C., " A Dynamic Threshold Voltage MOSFET (DTMOS) for Ultra-Low Voltage Operation", *IEDM Technical Digest*, 1994 p. 809

Bohr M., "High Performance Logic Technology and Reliability Challenges", International Reliability Physics Symposium, Keynote Address, 2003

Brown University, 2005, Lecture Notes on "Nanosystem Design Lecture: CMOS Fabrication" by Brown University, retrieved from http://www.lems.brown.edu/~iris/en291s10-05/Lectures/lectures.html on 10 December 2005

Chang W.H, Davari B., Wordeman M.R., Taur Y., Hsu C.H., and Rodriguez M.D., "A High-Performance 0.25 μm CMOS technology: I-Design and Characterization", *IEEE Trans. On Electron Devices*, Vol. 39, No. 4, April 1992

Chau R., Kavalieros J., Roberds B., Schenker R., Lionberger D., Barlage D., Doyle B., Arghavani R., Murthy A., and Dewey G., "30 nm Physical Gate Length CMOS Transistors with 1.0 ps n-MOS and 1.7 ps p-MOS Gate Delays", *IEDM Technical Digest, 2000*

Chua L-M., and Liu P-C., "Subthreshold Current for Submicron LDD MOS Transistor", *Proceedings of the 36th Midwest Symposium on Circuits and Systems*, 1993

Cowern N.E.B., Jaraiz M., Cristiano F., Claverie A., and Mannino G., "Fundamental Diffusion Issues for deep-Submicron Device Processing", *Electron Devices Meeting, 1999,* IEDM Technical Digest International

Critchlow D.L., "MOSFET Scaling – The Driver of VLSI Technology", *Proc. of the IEEE*, Vol. 87, 1999

Davari B., "CMOS Technology Scaling: 0.1 µm and Beyond", *Proceedings of the International Electron Devices Meeting*, 1996

Dennard R., Gaensslen F., Yu H., Rideout L., Bassous E., and LeBlanc A., "Design of Ion-Implanted MOSFET's with Very Small Physical Dimensions", *IEEE J. Solid-State Circuits*, vol. SC-9, No. 5, October 1974

Freescale Semiconductor, 2006, Semiconductor R&D and Laboratories, Silicon-on-Insulator, available at http://www.freescale.com

Horvath A., Parsa S., and Wong H.Y., "0.35 µm CMOS Process on Six-Inch Wafer-Baseline Report IV", Memorandum No. UCB / ERL M05 / 15, Electronics Research Laboratory , University of California, Berkeley, 2005

Hwang H., Lee D-H., and Hwang J.M., "Degradation of MOSFETs Drive Current Due to Halo Ion Implantation", *International Electron Device Meeting*, Technical Digest, 1996

Intel Technology Journal, Volume 2, Issue 3 (3rd Quarter 1998), "Intel's 0.25 Micron, 2.0 V Logic Process Technology", 1998

Intel Corporation, retrieved from http://www.intel.com/technology/silicon/mooreslaw/ on 24 November 2005

Ismail K., Chu J.O., and Meyerson B.S., "High Hole Mobility in SiGe Alloys for Device Applications", *Applied Physics Letter*, Vol. 64, 1994, p. 3124

Iwai H., and Ohmi S., "Problems and Solutions for Downsizing CMOS Below 0.1 μm", *ICSE 2000 Proceedings*, 2000

Kim H.S., Ko D.H., Bae D.L., Lee N.I., Kim D.W., Kang H.K., and Lee M.Y., "Gate Oxide Reliabilities in MOS (Metal-Oxide-Semiconductor) Structures with Ti-Polycide Gates", *Journal of Electronic Materials*, Vol. 27, No. 4, 1998

KNU, 2002, Lecture Notes on "NMOS/PMOS", from Kyungpook National University, Korea, retrieved from http://nanosys.knu.ac.kr/lecture/ on 14 March, 2005

Lee Y-T, Woo D-S, Lee J.D., and Park B-G, "Threshold Voltage Reduction Model for Buried Channel PMOSFETs Using Quasi-2-D Poisson Equation", *IEEE Transactions on Electron Devices*, Vol. 47, No. 12, December 2000.

Lim K.Y., and Xing Z., "Design, Modeling, and Characterization of Submicron MOSFETs", First Year Conversion Report 1997/98, Nanyang Technological University

Liou J.J., Ortiz-Conde A., and Garcia Sanchez F., "Extraction of the threshold voltage of MOSFETs : an overview (invited) ", *Proc. of Hong Kong Electron Device Meeting*, Hong Kong, 1997, pp. 31-38

Mat Hussin M.R., Mohd Saari S.A., Abdul Rahim A.I., Mat Ayub R., and Ahmad M.R., "Blanket and Pocket Anti Punchthrough Device Design Approaches in 0.35 μm CMOS Technology Development", *ICSE 2000 Proceedings*, Nov 2000

MIT, 2005, Lecture Notes on "Physics of Microfabrication: Advanced Front-End Processing", by Massachusetts Institute of Technology, retrieved from http://web.mit.edu/6.774/www/handout_1.pdf on 11 January 2006

NTRS, 1997, The National Technology Roadmap for Semiconductors, Technology Needs, 1997 Edition, Semiconductor Industry Association

NUS, 2003, Lecture Notes on "Operation of MOSFET" by National University of Singapore, retrieved from http://www.ece.nus.edu.sg/stfpage/elezhucx/myweb/Operation_of_MOSFET.pdf on 07 August 2004

Ozturk M., Pas S., Roozeboom F., and Timans P., "Rapid Thermal and Integrated Processing VII, ed.", *MRS Proc.* Vol. 525, 1998

Purdue University, 2003, Lecture Notes on "Advanced VLSI Design" by Purdue University, retrieved from http://web.ics.purdue.edu/~hanantha/academics.shtml on 12 July 2005

Purdue University, 2005, Lecture Notes on "Leakage and Leakage Reduction Techniques for Nano-scale CMOS: Device and Circuit Perspective" by Purdue University, Electrical and Computer Engineering, retrieved from http://www.research.ibm.com/aceed/2005/proceedings/roy.ppt on 11 December 2005

RIT, 2005, Lecture Notes on "RIT's Sub-CMOS Process (Leff < 1.0 μm)", by Rochester Institute of Technology, Microelectronic Engineering, retrieved from http://www.rit.edu/~lffeee/Cmos2003.pdf on 11 August 2005

Rodder M., Aur S., and Chen I-C., "A Scaled 1.8 V, 0.18 μm Gate Length CMOS Technology: Device Design and Reliability Considerations", *IEDM Tech. Digest*, pp. 415-418, 1995

Schramm D., Bowles D., Mastovich M., Knutrud P.C., and Tyurina A., "Algorithm Implementation and Techniques for Providing More Reliable Overlay Measurements and Better Tracking of the Shallow Trench Isolation (STI) Process", *Proceedings of SPIE,* Volume 3677, 1999

Schroder D.K., "Semiconductor Material and Device Characterization", Wiley, N.Y. 1990

Shibata A., Matsuoka T., Kakimoto S., Kotaki H., Nakano M., Adachi K., Ohta K. and Hashizume N., "Ultra Low Power Supply Voltage (0.3V) Operation with Extreme High Speed Using Bulk Dynamic Threshold Voltage MOSFET (B-DTMOS) with Advanced Fast Signal Transmission Shallow Well", *VLSI Symposium Digest,* 1998, p. 76

SIA, 1999, Semiconductor Industry Association, International Technology Roadmap for Semiconductors, 1999 edition, Austin, TX: SEMATECH, 1999

Silvaco International 2001, *ATHENA User's Manual, 2D Process Simulation Software,* Silvaco International 2001

Silvaco International 2002, *ATLAS User's Manual, Device Simulation Software, Volume 1 & 2* and *VWF Interactive Tools, User's Manual Volume 1 & 2,* Silvaco International 2002

Silvaco Resource Ctr., 1994, Silvaco Resource Center, Simulation Standard, Volume 5, Number 3, "A Modern Approach to 0.35 μm Technology Optimization", 1994

SJSU, 2005, Lecture Notes on "Virtual Wafer Fab (VWF) Framework from Silvaco" by San Jose State University, retrieved from http://kaflatooni.yasna.com/atlas_tutorial/pdf on 15 November 2005

Slisher D.K., Filippi R.G. Jr., Storaska D.W., and Gay A.H., “Scaling of Si MOSFETs for Digital Applications”, 1999

Stanford University, 2005, Lecture Notes on “Advanced CMOS Structures”, retrieved from http://www.stanford.edu/class/ee410/AdvCMOS.pdf on 13 December 2005

Sze S.M., “Physics of Semiconductor Devices”, Bell Laboratories, Inc.,Murray Hill, New Jersey, Second Edition, John Wiley & Sons, New York, 1981

Tasch A.F., Shin H., Bordelon T.J., and Maziar C.M., “Limitations of LDD Types of Structures in Deep-Submicrometer MOS Technology”, *IEEE Electron Device Letters*, Vol. 11, No. 11, 1990

Taur Y., Buchanan D.A., Chen W., Frank D.J., Ismail K.E., Lo S-H., Sai-Halasz G.A., Viswanathan R.G., Wann H-J. C, Wind S.J., and Wong H-S., “CMOS Scaling into the Nanometer Regime”, *Proceedings of the IEEE*, Vol. 85, No. 4, April 1997.

Taur Y., “Design Near the Limits of Scaling”, *Journal of Research and Development*, Volume 46, No. 23, 2002

Thompson S., Packan P., and Bohr M., “MOS Scaling: Transistor Challenges for the 21st Century”, *Intel Technology Journal*, Q3’98, pp. 1-19

Toan N.N., “Spin-On-Glass Materials and Applications in Advanced IC Technologies”, 1999

UMBC, 2004, Lecture Notes on “Advanced VLSI Design” by University of Maryland, Baltimore County, retrieved from http://www.csee.umbc.edu/~plusquel/vlsiII/slides/scaling.pdf on 05 January 2005

UC Berkeley, 2006, Lecture Notes on "Process Integration" by University of Berkeley, California, retrieved from http://www-inst.eecs.berkeley.edu/~ee143/sp06/lectures/Lec_20.pdf on 12 February 2006

Uyemura J.P., "Circuit Design for CMOS VLSI", Kluwer Academic Publishers, Boston, 1992

Woerlee P.H., Juffermans C.A.H., Lifka H., Manders W., Pomp H., Paulzen G., Walker A.J., and Woltjer R., "A Low Power 0.25 μm CMOS Technology", *IEDM 92-31*

Wolf S., "Silicon Processing for the VLSI Era, Volume 3: The Submicron MOSFET", California, Lattice Press, 1995

Wolf S., "Silicon Processing for the VLSI Era, Volume 2: Process Integration", California, Lattice Press, 2000

Xiao H., "Introduction to Semiconductor Manufacturing Technology", Prentice Hall, 2001

Yu B., Wann C.H.J., Nowak E.D., Noda K., and C. Hu, "Short Channel Effect Improved by Lateral Channel-Engineering in Deep-Submicrometer MOSFET's", *IEEE Transactions on Electron Devices*, Vol. ED-44, 1997

Zeitzoff P.M., "Front-End Trends, Challenges and Potential Solutions for the 180 – 100 nm IC Technology Generations", Semiconductor Fabtech –10th Edition, pp. 275-282, 1999

Publications

1. Norani Muti Mohamed and Noraini Othman, "Simulations of 0.35 μm CMOS Technology for Advanced Semiconductor Courses", presented at the *Regional Conference on Solid State Science & Technology 2005 (RCSST '05)*, Pahang, Malaysia, 18-21 December, 2005. Will be published in the *Journal of Solid State Science & Technology 2006.*

2. Norani Muti Mohamed and Noraini Othman, "Miniaturization and Simulation of CMOS Transistor", presented at the *International Conference on Computer and Communication Engineering 2006 (ICCCE'06)*, Kuala Lumpur, Malaysia, 9-11 May, 2006. Published in the conference proceedings

APPENDIX A

Input deck for 0.25 µm NMOS

```
go athena

#
line x loc=-5.0 spac=0.1
line x loc=-2.0 spac=0.1
line x loc=-0.65 spac=0.05
line x loc=0 spac=0.05

#
line y loc=-0.45 spac=0.05
line y loc=0 spac=0.05
line y loc=0.01 spac=0.05
line y loc=0.2 spac=0.05
line y loc=0.4 spac=0.05
line y loc=0.6 spac=0.1
line y loc=1.2 spac=0.1
line y loc=6 spac=0.1
line y loc=10 spac=0.3
line y loc=12 spac=0.4
line y loc=15 spac=0.4
line y loc=20 spac=0.4
line y loc=25 spac=0.4

#
init orientation=100 c.boron=3e14 space.mul=2 two.d

#
method full.cpl

# PAD OXIDATION 250A
method grid.oxide=0.005
diffus time=21 temp=1000 dryo2
diffus time=15 temp=1000 nitro

#
extract name="Pad oxide_1" thickness material="SiO~2" mat.occno=1 x.val=-2.0

# P-WELL IMPLANT
implant boron dose=5e12 energy=60 tilt=0 rotation=0 pears unit.damage

# WELL DRIVE-IN
```

```
diffus time=60 temp=750 t.final=1100 nitro
diffus time=150 temp=1100 dryo2
diffus time=15 temp=1100 nitro

# ETCH OXIDE THAT RESULTS FROM WELL DRIVE-IN AND THE #PREVIOUS
PAD OXIDE
etch oxide all

# PAD OXIDATION 250A
diffus time=21 temp=1000 dryo2
diffus time=15 temp=1000 nitro

#
extract name="Pad oxide_2" thickness material="SiO~2" mat.occno=1 x.val=-2.0

# NITRIDE DEPOSITION
deposit nitride thick=0.18 divisions=10

# ETCH NITRIDE OUTSIDE OF THE ACTIVE REGION
#(below LOCOS area)
etch nitride left p1.x=-4.0

# P-FIELD IMPLANT (underneath LOCOS area, active area are covered)
implant boron dose=2e13 energy=80 tilt=0 rotation=0 pears unit.damage

# LOCOS OXIDATION 5500A
method grid.oxide=0.055
diffus time=120 temp=1000 weto2
diffus time=20 temp=1000 nitro

#
extract name="LOCOS" thickness material="SiO~2" mat.occno=1 x.val=-4.5

structure outfile=LOCOS_1.str

# ETCH ALL NITRIDE
etch nitride all

# ETCH PAD OXIDE
etch oxide dry thick=0.025

# SCREEN OXIDATION 250A
method grid.oxide=0.005
diffus time=21 temp=1000 dryo2
diffus time=15 temp=1000 nitro
```

```
#
extract name="Screen oxide" thickness material="SiO~2" mat.occno=1 x.val=-2.0

structure outf=screenox.str

# NVT IMPLANT BORON TO SHIFT THE THRESHOLD
#(throughout active region)
implant bf2 dose=5.6e12 energy=50 tilt=0 rotation=0 pears unit.damage

# ETCH SCREEN OXIDE
etch oxide dry thick=0.025

# OXIDIZE THE GATE 57A
method grid.oxide=0.004
diffus time=25 temp=850 dryo2
diffus time=30 temp=1000 nitro

#
extract name="Gate oxide" thickness material="SiO~2" mat.occno=1 x.val=-2.0

structure outf=gateox.str

# DEPOSIT UNDOPED POLY GATE 2500A
deposit poly thick=0.25 divisions=10

# ETCH POLYSILICON
etch poly left p1.x=-0.125

# N-TYPE LDD IMPLANT
implant arsenic dose=5e13 energy=30 tilt=7 rotation=0 pears unit.damage
implant arsenic dose=5e13 energy=30 tilt=-7 rotation=0 pears unit.damage

# LDD SPACER DEPOSITION (SPACER WIDTH TARGET=3000A)
deposit oxide thick=0.3 divisions=10

# LDD SPACER FORMATION
etch oxide dry thick=0.3

# N+ GATE AND S/D IMPLANT
implant phosphor dose=3e15 energy=40 tilt=0 rotation=0 pears unit.damage

# GATE AND S/D ANNEAL (RTA)
diffus time=10/60 temp=900 nitro
diffus time=5/60 temp=1050 nitro
```

```
structure outf=aftersd.str

# ETCH GATE OXIDE THICKNESS
etch oxide dry thick=0.0057

# SILICIDATION
deposit titanium thick=0.03

#
diffus time=0.25 temp=650 nitro

structure outfile=silicide.str

# ETCH TITANIUM
etch titanium all


# PSG DEPOSITION 700nm
deposit oxide thick=0.7 divisions=10

# PSG DENSIFICATION
diffus time=0.167 temp=900 nitro

structure outfile=PSG.str


# ETCH CONTACT HOLES
etch oxide start x=-2.9 y=-1.5
etch cont x=-2.9 y=0
etch cont x=-0.9 y=0
etch done x=-0.9 y=-1.5

structure outfile=1st_contact.str

# EXTRACT DESIGN PARAMETERS

# Extract final S/D Xj
extract name="nxj" xj silicon mat.occno=1 x.val=-2.0 junc.occno=1

# Extract the surface conc under the channel
extract name="chan surf conc" surf.conc impurity="Net Doping"\
material="Silicon" mat.occno=1 x.val=-0.1

structure mirror right
```

```
electrode name=gate x=0 y=-0.59
electrode name=source x=-2.5
electrode name=drain x=2.5
electrode name=substrate backside

structure outfile=NMOS_1.str

#PLOT THE STRUCTURE
tonyplot NMOS_1.str -set NMOS_1.set

go devedit
work.area x1=-5 y1=-1.3701461 x2=5 y2=5
# devedit 2.6.0.R (Thu Dec 12 12:40:19 PST 2002)
# libSvcFile 1.8.3 (Sat Dec  7 17:56:58 PST 2002)
# libsflm 4.14.3 (Sat Dec  7 18:02:49 PST 2002)
# libSDB 1.4.3 (Tue Dec 10 19:51:05 PST 2002)
# libDW_Version 2.0.0.R (Thu Nov 28 05:44:29 PST 2002)
region reg=1 mat=Silicon \
        polygon="-5,-0.063 -4.589,-0.063 -4.387,-0.071 -4.182,-0.084 -3.952,-0.115 -
3.603,-0.291 -3.425,-0.331 -3.244,-0.344 -3.242,-0.34 -3.194,-0.325 "\
        "-3.185,-0.323 -3.058,-0.315 -0.581,-0.315 -0.556,-0.316 -0.464,-0.34 0.464,-0.34
0.556,-0.316 0.625,-0.316 0.65,-0.315 3.058,-0.315 "\
        "3.185,-0.323 3.194,-0.325 3.242,-0.34 3.244,-0.344 3.425,-0.331 3.603,-0.291
3.952,-0.115 4.182,-0.084 4.387,-0.071 4.788,-0.071 "\
        "5,-0.063 5,5 -5,5"
#
constr.mesh region=1 default

region reg=2 mat="Silicon Oxide" \
        polygon="-5,-1.37 -4.657,-1.37 -4.581,-1.367 -3.997,-1.276 -3.901,-1.254 -3.801,-
1.227 -3.731,-1.203 -3.573,-1.144 -3.412,-1.09 -3.36,-1.075 "\
        "-3.271,-1.055 -3.185,-1.045 -2.9,-1.045 -2.9,-0.345 -3.241,-0.345 -3.244,-0.344 -
3.425,-0.331 -3.603,-0.291 -3.952,-0.115 -4.182,-0.084 "\
        "-4.387,-0.071 -4.589,-0.063 -5,-0.063"
#
constr.mesh region=2 default

region reg=3 mat="Silicon Oxide" \
        polygon="-0.464,-0.34 -0.556,-0.346 -0.581,-0.345 -0.9,-0.345 -0.9,-1.045 -
0.768,-1.045 -0.721,-1.081 -0.622,-1.147 -0.588,-1.167 -0.507,-1.207 "\
        "-0.42,-1.24 0,-1.277 0.42,-1.24 0.507,-1.207 0.588,-1.167 0.622,-1.147 0.721,-
1.081 0.768,-1.045 0.9,-1.045 0.9,-0.345 "\
        "0.65,-0.345 0.625,-0.346 0.557,-0.346 0.464,-0.34" \
        polygon="0.125,-0.346 0.125,-0.585 0.126,-0.591 0.126,-0.596 0,-0.602 -0.126,-
0.596 -0.126,-0.591 -0.125,-0.585 -0.125,-0.346"
```

```
#
constr.mesh region=3 default

region reg=4 mat=PolySilicon \
        polygon="-0.125,-0.346 -0.125,-0.585 -0.119,-0.59 -0.12,-0.591 0,-0.571 0.12,-
0.591 0.119,-0.59 0.125,-0.585 0.125,-0.346"
#
constr.mesh region=4 default

region reg=5 name=source mat="Titanium Silicide" elec.id=2 work.func=0 \
        polygon="-0.9,-0.345 -0.581,-0.345 -0.556,-0.346 -0.464,-0.34 -0.556,-0.316 -
0.581,-0.315 -3.058,-0.315 -3.185,-0.323 -3.194,-0.325 -3.242,-0.34 "\
        "-3.244,-0.344 -3.241,-0.345 -2.9,-0.345"
#
constr.mesh region=5 default

region reg=6 name=gate mat="Titanium Silicide" elec.id=1 work.func=0 \
        polygon="-0.12,-0.591 -0.119,-0.59 -0.125,-0.585 -0.126,-0.591 -0.126,-0.596 0,-
0.602 0.126,-0.596 0.126,-0.591 0.125,-0.585 0.119,-0.59 "\
        "0.12,-0.591 0,-0.571"
#
constr.mesh region=6 default

region reg=7 mat="Silicon Oxide" \
        polygon="5,-0.063 4.788,-0.071 4.387,-0.071 4.182,-0.084 3.952,-0.115 3.603,-
0.291 3.425,-0.331 3.244,-0.344 3.242,-0.345 2.9,-0.345 "\
        "2.9,-1.045 3.194,-1.045 3.271,-1.055 3.36,-1.075 3.412,-1.09 3.573,-1.144
3.731,-1.203 3.801,-1.227 3.901,-1.254 3.997,-1.276 "\
        "4.581,-1.367 4.71,-1.367 4.788,-1.37 5,-1.37"
#
constr.mesh region=7 default

region reg=8 name=drain mat="Titanium Silicide" elec.id=3 work.func=0 \
        polygon="2.9,-0.345 3.242,-0.345 3.244,-0.344 3.242,-0.34 3.194,-0.325 3.185,-
0.323 3.058,-0.315 0.65,-0.315 0.625,-0.316 0.556,-0.316 "\
        "0.464,-0.34 0.557,-0.346 0.625,-0.346 0.65,-0.345 0.9,-0.345"
#
constr.mesh region=8 default

substrate name="substrate" electrode=4 workfunction=0


# Set Meshing Parameters
#
base.mesh height=10 width=10
```

```
#
bound.cond !apply max.slope=28 max.ratio=300 rnd.unit=0.001 line.straightening=1
align.points when=automatic
#
imp.refine imp="Net Doping" scale=log
imp.refine min.spacing=0.02
#
constr.mesh max.angle=90 max.ratio=300 max.height=10000 \
        max.width=10000 min.height=0.0001 min.width=0.0001
#
constr.mesh type=Semiconductor default
#
constr.mesh type=Insulator default
#
constr.mesh type=Metal default
#
constr.mesh type=Other default
#
constr.mesh region=1 default
#
constr.mesh region=2 default
#
constr.mesh region=3 default
#
constr.mesh region=4 default
#
constr.mesh region=5 default
#
constr.mesh region=6 default
#
constr.mesh region=7 default
#
constr.mesh region=8 default
Mesh Mode=MeshBuild
refine mode=both x1=-2.89 y1=-0.33 x2=2.88 y2=0.38
refine mode=both x1=-2.96 y1=-0.35 x2=2.93 y2=0.55


structure outf=AB.str

############# Vth Test ###############
go atlas

# set material models
models cvt srh print
```

```
interface qf=1e10

method gummel newton
solve init

# Bias the drain
solve vdrain=0.05

# Ramp the gate
log outf=mos1ex01_1.log master
solve vgate=0 vstep=0.025 vfinal=1.5 name=gate
save outf=mos1ex01_1.str

# plot results
tonyplot  mos1ex01_1.log -set mos1ex01_1_log.set

# extract device parameters
extract name="nvt" (xintercept(maxslope(curve(abs(v."gate"),abs(i."drain")))) \
        - abs(ave(v."drain"))/2.0)

############# Subvt ################
go atlas

# set material models
models cvt srh print
interface qf=1e10

# get initial solution

solve init

method gummel newton
solve prev


# Bias the drain a bit...
solve vdrain=0 vstep=0.01 vfinal=0.05 name=drain

# Ramp the gate to a volt...
log outf=mos1ex03_1.log master
solve vgate=0 vstep=0.1 vfinal=2.5 name=gate
save outf=mos1ex03_1.str

# extract the device parameter SubVt...
```

```
extract init inf="mos1ex03_1.log"
extract name="nsubvt" 1.0/slope(maxslope(curve(abs(v."gate"),log10(abs(i."drain")))))

tonyplot mos1ex03_1.log -set mos1ex03_1_log.set

############# Ioff ################
go atlas

# set material models
models cvt srh print
interface qf=1e10

# get initial solution

solve init

method gummel newton
solve prev

# Bias the drain a bit...
solve vdrain=0 vstep=0.05 vfinal=2.5 name=drain

# Ramp the gate to a volt...
log outf=mos1ex04_1.log master
solve vgate=0 vstep=0.1 vfinal=2.5 name=gate
save outf=mos1ex04_1.str

# extract the device parameter Ioff...
extract init inf="mos1ex04_1.log"
extract name="Ioff" 10^(y.val from curve (v."gate",log10(i."drain"))\
where x.val=0)

tonyplot mos1ex04_1.log -set mos1ex04_1_log.set

tonyplot -overlay -st mos1ex03_1.log  mos1ex04_1.log -set mos1ex03_1.set

############# Idsat with Vdd=3.3V ################
go atlas

# Define the Gate Qss
interface qf=1e10

# Use the cvt mobility model for MOS
models cvt srh print numcarr=2
```

```
#
method gummel newton

# set gate biases with Vds=0.0
solve init
solve vgate=0.5 outf=solve_tmp1
solve vgate=1.0 outf=solve_tmp2
solve vgate=1.5 outf=solve_tmp3
solve vgate=2.0 outf=solve_tmp4
solve vgate=2.5 outf=solve_tmp5


#load in temporary files and ramp Vds
load infile=solve_tmp1
log outf=mos1ex02_1.log
solve name=drain vdrain=0 vfinal=2.5 vstep=0.05


load infile=solve_tmp2
log outf=mos1ex02_2.log
solve name=drain vdrain=0 vfinal=2.5 vstep=0.05

load infile=solve_tmp3
log outf=mos1ex02_3.log
solve name=drain vdrain=0 vfinal=2.5 vstep=0.05


load infile=solve_tmp4
log outf=mos1ex02_4.log
solve name=drain vdrain=0 vfinal=2.5 vstep=0.05


load infile=solve_tmp5
log outf=mos1ex02_5.log
solve name=drain vdrain=0 vfinal=2.5 vstep=0.05

# extract max current and saturation slope
extract name="nidsmax" max(abs(i."drain"))

#
tonyplot -overlay -st mos1ex02_1.log  mos1ex02_2.log mos1ex02_3.log mos1ex02_4.log
mos1ex02_5.log -set mos1ex02_1.set


quit
```

APPENDIX B

```
**Input deck for 0.25 µm PMOS**

go athena

#
line x loc=-5.0 spac=0.1
line x loc=-2.0 spac=0.1
line x loc=-0.65 spac=0.05
line x loc=0 spac=0.05

#
line y loc=-0.45 spac=0.05
line y loc=0 spac=0.05
line y loc=0.01 spac=0.05
line y loc=0.2 spac=0.05
line y loc=0.4 spac=0.05
line y loc=0.6 spac=0.1
line y loc=1.2 spac=0.1
line y loc=6 spac=0.1
line y loc=10 spac=0.3
line y loc=12 spac=0.4
line y loc=15 spac=0.4
line y loc=20 spac=0.4
line y loc=25 spac=0.4

#
init orientation=100 c.boron=3e14 space.mul=2 two.d

#
method full.cpl

# PAD OXIDATION 250A
method grid.oxide=0.005
diffus time=21 temp=1000 dryo2
diffus time=15 temp=1000 nitro

#
extract name="Pad oxide_1" thickness material="SiO~2" mat.occno=1 x.val=-2.0


# N-WELL IMPLANT
implant phos dose=1e13 energy=150 tilt=0 rotation=0 pears unit.damage
```

```
# WELL DRIVE-IN
diffus time=60 temp=750 t.final=1100 nitro
diffus time=150 temp=1100 dryo2
diffus time=15 temp=1100 nitro

# ETCH OXIDE THAT RESULTS FROM WELL DRIVE-IN AND THE #PREVIOUS
PAD OXIDE
etch oxide all

# PAD OXIDATION 250A
diffus time=21 temp=1000 dryo2
diffus time=15 temp=1000 nitro

#
extract name="Pad oxide_2" thickness material="SiO~2" mat.occno=1 x.val=-2.0

# NITRIDE DEPOSITION
deposit nitride thick=0.18 divisions=10

# ETCH NITRIDE OUTSIDE OF THE ACTIVE REGION
#(below LOCOS area)
etch nitride left p1.x=-4.0

# LOCOS OXIDATION 5500A
method grid.oxide=0.055
diffus time=120 temp=1000 weto2
diffus time=20 temp=1000 nitro

#
extract name="LOCOS" thickness material="SiO~2" mat.occno=1 x.val=-4.5

structure outfile=LOCOS_1.str

# ETCH ALL NITRIDE
etch nitride all

# ETCH PAD OXIDE
etch oxide dry thick=0.025


# SCREEN OXIDATION 250A
method grid.oxide=0.005
diffus time=21 temp=1000 dryo2
diffus time=15 temp=1000 nitro
```

```
#
extract name="Screen oxide" thickness material="SiO~2" mat.occno=1 x.val=-2.0

structure outfile=screenoxide.str

# PVT IMPLANT BORON TO SHIFT THE THRESHOLD
#(throughout active region)
implant phos dose=2.8e12 energy=30 tilt=0 rotation=0 pears unit.damage

# ETCH SCREEN OXIDE
etch oxide dry thick=0.025

# OXIDIZE THE GATE 57A
method grid.oxide=0.004
diffus time=25 temp=850 dryo2
diffus time=30 temp=1000 nitro

#
extract name="Gate oxide" thickness material="SiO~2" mat.occno=1 x.val=-2.0

structure outf=gateox.str

# DEPOSIT UNDOPED POLY GATE 2500A
deposit poly thick=0.25 divisions=10

# ETCH POLYSILICON
etch poly left p1.x=-0.175

# P-TYPE LDD IMPLANT
implant bf2 dose=5e13 energy=10 tilt=7 rotation=0 pears unit.damage
implant bf2 dose=5e13 energy=10 tilt=-7 rotation=0 pears unit.damage

# LDD SPACER DEPOSITION (SPACER WIDTH TARGET=3000A)
deposit oxide thick=0.3 divisions=10

# LDD SPACER FORMATION
etch oxide dry thick=0.3

# P+ GATE AND S/D IMPLANT
implant boron dose=3e15 energy=20 tilt=0 rotation=0 pears unit.damage

# GATE AND S/D ANNEAL (RTA)
diffus time=10/60 temp=900 nitro
diffus time=5/60 temp=1050 nitro
```

```
structure outf=aftersd.str

# ETCH GATE OXIDE THICKNESS
etch oxide dry thick=0.008

# SILICIDATION
deposit titanium thick=0.03 divisions=3

#
diffus time=0.25 temp=650 nitro

structure outfile=silicide.str

# ETCH TITANIUM
etch titanium all

structure outf=afteretch.str

# PSG DEPOSITION 700nm
deposit oxide thick=0.7 divisions=10

# PSG DENSIFICATION
diffus time=0.167 temp=900 nitro

structure outfile=PSG.str

# ETCH CONTACT HOLES
etch oxide start x=-2.9 y=-1.5
etch cont x=-2.9 y=0
etch cont x=-0.9 y=0
etch done x=-0.9 y=-1.5

structure outfile=1st_contact.str

# EXTRACT DESIGN PARAMETERS

# Extract final S/D Xj
extract name="pxj" xj silicon mat.occno=1 x.val=-2.0 junc.occno=1

# Extract the surface conc under the channel
extract name="chan surf conc" surf.conc impurity="Net Doping"\
material="Silicon" mat.occno=1 x.val=-0.1

structure mirror right
```

```
electrode name=gate x=0 y=-0.59
electrode name=source x=-2.5
electrode name=drain x=2.5
electrode name=substrate backside

structure outfile=PMOS_1.str

#PLOT THE STRUCTURE
tonyplot PMOS_1.str -set PMOS_1.set

go devedit
work.area x1=-5 y1=-1.3701461 x2=5 y2=5
# devedit 2.6.0.R (Thu Dec 12 12:40:19 PST 2002)
# libSvcFile 1.8.3 (Sat Dec  7 17:56:58 PST 2002)
# libsflm 4.14.3 (Sat Dec  7 18:02:49 PST 2002)
# libSDB 1.4.3 (Tue Dec 10 19:51:05 PST 2002)
# libDW_Version 2.0.0.R (Thu Nov 28 05:44:29 PST 2002)
region reg=1 mat=Silicon \
        polygon="-5,-0.063 -4.589,-0.063 -4.387,-0.071 -4.182,-0.084 -3.952,-0.115 -
3.603,-0.291 -3.425,-0.331 -3.244,-0.344 -3.242,-0.34 -3.194,-0.325 "\
        "-3.185,-0.323 -3.058,-0.315 -0.581,-0.315 -0.556,-0.316 -0.464,-0.34 0.464,-0.34
0.556,-0.316 0.625,-0.316 0.65,-0.315 3.058,-0.315 "\
        "3.185,-0.323 3.194,-0.325 3.242,-0.34 3.244,-0.344 3.425,-0.331 3.603,-0.291
3.952,-0.115 4.182,-0.084 4.387,-0.071 4.788,-0.071 "\
        "5,-0.063 5,5 -5,5"
#
constr.mesh region=1 default

region reg=2 mat="Silicon Oxide" \
        polygon="-5,-1.37 -4.657,-1.37 -4.581,-1.367 -3.997,-1.276 -3.901,-1.254 -3.801,-
1.227 -3.731,-1.203 -3.573,-1.144 -3.412,-1.09 -3.36,-1.075 "\
        "-3.271,-1.055 -3.185,-1.045 -2.9,-1.045 -2.9,-0.345 -3.241,-0.345 -3.244,-0.344 -
3.425,-0.331 -3.603,-0.291 -3.952,-0.115 -4.182,-0.084 "\
        "-4.387,-0.071 -4.589,-0.063 -5,-0.063"
#
constr.mesh region=2 default

region reg=3 mat="Silicon Oxide" \
        polygon="-0.464,-0.34 -0.556,-0.346 -0.581,-0.345 -0.9,-0.345 -0.9,-1.045 -
0.768,-1.045 -0.721,-1.081 -0.622,-1.147 -0.588,-1.167 -0.507,-1.207 "\
        "-0.42,-1.24 0,-1.277 0.42,-1.24 0.507,-1.207 0.588,-1.167 0.622,-1.147 0.721,-
1.081 0.768,-1.045 0.9,-1.045 0.9,-0.345 "\
        "0.65,-0.345 0.625,-0.346 0.557,-0.346 0.464,-0.34" \
        polygon="0.125,-0.346 0.125,-0.585 0.126,-0.591 0.126,-0.596 0,-0.602 -0.126,-
0.596 -0.126,-0.591 -0.125,-0.585 -0.125,-0.346"
```

```
#
constr.mesh region=3 default

region reg=4 mat=PolySilicon \
        polygon="-0.125,-0.346 -0.125,-0.585 -0.119,-0.59 -0.12,-0.591 0,-0.571 0.12,-
0.591 0.119,-0.59 0.125,-0.585 0.125,-0.346"
#
constr.mesh region=4 default

region reg=5 name=source mat="Titanium Silicide" elec.id=2 work.func=0 \
        polygon="-0.9,-0.345 -0.581,-0.345 -0.556,-0.346 -0.464,-0.34 -0.556,-0.316 -
0.581,-0.315 -3.058,-0.315 -3.185,-0.323 -3.194,-0.325 -3.242,-0.34 "\
        "-3.244,-0.344 -3.241,-0.345 -2.9,-0.345"
#
constr.mesh region=5 default

region reg=6 name=gate mat="Titanium Silicide" elec.id=1 work.func=0 \
        polygon="-0.12,-0.591 -0.119,-0.59 -0.125,-0.585 -0.126,-0.591 -0.126,-0.596 0,-
0.602 0.126,-0.596 0.126,-0.591 0.125,-0.585 0.119,-0.59 "\
        "0.12,-0.591 0,-0.571"
#
constr.mesh region=6 default

region reg=7 mat="Silicon Oxide" \
        polygon="5,-0.063 4.788,-0.071 4.387,-0.071 4.182,-0.084 3.952,-0.115 3.603,-
0.291 3.425,-0.331 3.244,-0.344 3.242,-0.345 2.9,-0.345 "\
        "2.9,-1.045 3.194,-1.045 3.271,-1.055 3.36,-1.075 3.412,-1.09 3.573,-1.144
3.731,-1.203 3.801,-1.227 3.901,-1.254 3.997,-1.276 "\
        "4.581,-1.367 4.71,-1.367 4.788,-1.37 5,-1.37"
#
constr.mesh region=7 default

region reg=8 name=drain mat="Titanium Silicide" elec.id=3 work.func=0 \
        polygon="2.9,-0.345 3.242,-0.345 3.244,-0.344 3.242,-0.34 3.194,-0.325 3.185,-
0.323 3.058,-0.315 0.65,-0.315 0.625,-0.316 0.556,-0.316 "\
        "0.464,-0.34 0.557,-0.346 0.625,-0.346 0.65,-0.345 0.9,-0.345"
#
constr.mesh region=8 default

substrate name="substrate" electrode=4 workfunction=0


# Set Meshing Parameters
#
base.mesh height=10 width=10
```

```
#
bound.cond !apply max.slope=28 max.ratio=300 rnd.unit=0.001 line.straightening=1
align.points when=automatic
#
imp.refine imp="Net Doping" scale=log
imp.refine min.spacing=0.02
#
constr.mesh max.angle=90 max.ratio=300 max.height=10000 \
        max.width=10000 min.height=0.0001 min.width=0.0001
#
constr.mesh type=Semiconductor default
#
constr.mesh type=Insulator default
#
constr.mesh type=Metal default
#
constr.mesh type=Other default
#
constr.mesh region=1 default
#
constr.mesh region=2 default
#
constr.mesh region=3 default
#
constr.mesh region=4 default
#
constr.mesh region=5 default
#
constr.mesh region=6 default
#
constr.mesh region=7 default
#
constr.mesh region=8 default
Mesh Mode=MeshBuild
refine mode=both x1=-2.89 y1=-0.33 x2=2.88 y2=0.38
refine mode=both x1=-2.96 y1=-0.35 x2=2.93 y2=0.55

structure outf=AB.str

############# Vth Test ###############
go atlas

# set material models
models cvt srh print
```

```
interface qf=1e10

method gummel newton
solve init

# Bias the drain
solve vdrain=-0.05

# Ramp the gate
log outf=mos1ex01_1.log master
solve vgate=0 vstep=-0.025 vfinal=-1.5 name=gate
save outf=mos1ex01_1.str

# plot results
tonyplot  mos1ex01_1.log -set mos1ex01_1_log.set

# extract device parameters
extract name="pvt" (xintercept(maxslope(curve(abs(v."gate"),abs(i."drain")))) \
        - abs(ave(v."drain"))/2.0)

############# Subvt ################
go atlas

# set material models
models cvt srh print
interface qf=1e10

# get initial solution

solve init

method gummel newton
solve prev

# Bias the drain a bit...
solve vdrain=0 vstep=-0.01 vfinal=-0.05 name=drain

# Ramp the gate to a volt...
log outf=mos1ex03_1.log master
solve vgate=0 vstep=-0.1 vfinal=-2.5 name=gate
save outf=mos1ex03_1.str

# extract the device parameter SubVt...
extract init inf="mos1ex03_1.log"
extract name="psubvt" 1.0/slope(maxslope(curve(abs(v."gate"),log10(abs(i."drain")))))
```

```
tonyplot mos1ex03_1.log -set mos1ex03_1_log.set

############# Ioff #################
go atlas

# set material models
models cvt srh print
interface qf=1e10

# get initial solution

solve init

method gummel newton
solve prev

# Bias the drain a bit...
solve vdrain=0 vstep=-0.05 vfinal=-2.5 name=drain

# Ramp the gate to a volt...
log outf=mos1ex04_1.log master
solve vgate=0 vstep=-0.1 vfinal=-2.5 name=gate
save outf=mos1ex04_1.str

# extract the device parameter Ioff...
extract init inf="mos1ex04_1.log"
extract name="Ioff" 10^(y.val from curve (v."gate",log10(i."drain"))\
where x.val=0)

tonyplot mos1ex04_1.log -set mos1ex04_1_log.set

tonyplot -overlay -st mos1ex03_1.log  mos1ex04_1.log -set mos1ex03_1.set

############# Idsat with Vdd=3.3V ################
go atlas

# Define the Gate Qss
interface qf=1e10

# Use the cvt mobility model for MOS
models cvt srh print numcarr=2

#
method gummel newton
```

```
# set gate biases with Vds=0.0
solve init
solve vgate=-0.5 outf=solve_tmp1
solve vgate=-1.0 outf=solve_tmp2
solve vgate=-1.5 outf=solve_tmp3
solve vgate=-2.0 outf=solve_tmp4
solve vgate=-2.5 outf=solve_tmp5

#load in temporary files and ramp Vds
load infile=solve_tmp1
log outf=mos1ex02_1.log
solve name=drain vdrain=0 vfinal=-2.5 vstep=-0.05

load infile=solve_tmp2
log outf=mos1ex02_2.log
solve name=drain vdrain=0 vfinal=-2.5 vstep=-0.05

load infile=solve_tmp3
log outf=mos1ex02_3.log
solve name=drain vdrain=0 vfinal=-2.5 vstep=-0.05

load infile=solve_tmp4
log outf=mos1ex02_4.log
solve name=drain vdrain=0 vfinal=-2.5 vstep=-0.05

  load infile=solve_tmp5
log outf=mos1ex02_5.log
solve name=drain vdrain=0 vfinal=-2.5 vstep=-0.05

# extract max current and saturation slope
extract name="pidsmax" max(abs(i."drain"))

#
tonyplot -overlay -st mos1ex02_1.log  mos1ex02_2.log mos1ex02_3.log mos1ex02_4.log
mos1ex02_5.log -set mos1ex02_1.set


quit
```