STUDY OF BASIC 22nm TRANSISTOR TECHNOLOGY ON SEQUENTIAL CIRCUIT USING PrimeTime

by

NIK AZMAN BIN NIK HASSAN

Thesis submitted in fulfillment of the requirements
for the degree of
Master of Science

July 2013

ACKNOWLEDGEMENT

I would like to thank everyone who contributed to this research, especially my supervisor, Dr. Asrulnizam Bin Abd Manaf for guiding me with his expertise and special thanks to Mr. Leong Chan Ming from Intel Corporation, for his support and technical guidance. My thanks also goes to my colleagues; Saravanan Balakrishnan, Ong Chun Kiat, Seow Syee Chin and Chong Khui Seng for helping through discussions and suggestions. Further, I would like to thank my parents and family especially Deezex Noor Ainizaa Bt Abdullah for their continuous love and support. Lastly, my appreciation goes to Universiti Sains Malaysia (USM) and Intel Corporation for allowing me to use their facilities throughout the research.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENT **ii**

TABLE OF CONTENTS **iii**

LIST OF TABLES **vi**

LIST OF FIGURES **vii**

LIST OF ABBREVIATION **xi**

LIST OF PUBLICATION **xiv**

ABSTRAK **xv**

ABSTRACT **xvii**

CHAPTER 1 - INTRODUCTION

1.0 Introduction **1**

1.1 Problem Statements 4

1.2 The Research Objectives 6

1.3 Project Contributions 6

1.4 Structure of Thesis 7

CHAPTER 2 - LITERATURE REVIEW

2.0 Introduction **9**

2.1 The 22nm tri-gate transistor technology 10

2.1.1 Intel Technology Library 10

2.1.2 Intel 22nm Tri-Gate Transistor Technology. 19

2.1.3 Standard Cell Library 33
2.1.4 Summary 34
2.2 Flip-flop and latch 34
2.2.1 Flip-flop and latch overview 35
2.2.2 Power dissipation in flip-flop and latch 38
2.2.3 Advantages of latch-based designs 42
2.2.4 Summary 58

CHAPTER 3 – METHODOLOGY AND IMPLEMENTATION

3.0 Introduction 60

3.1 Case study. 60
3.2 Design Flow 62
3.2.1 Design idea. 64
3.2.2 Behavioral Description (Register Transfer Level) 65
3.2.3 Logic Synthesis 65
3.2.4 Functional Verification 69
3.2.5 Power, area and timing analysis 71
3.3 Summary 83

CHAPTER 4 – RESULTS AND DISCUSSIONS

4.0 Introduction 84

4.1 Behavioral Description 84
4.2 Logic Synthesis 87
4.3 Functional Verification 90
4.4 Power, area and timing analysis 90
4.4.1 Power analysis 91
4.4.2 Area analysis 93
4.4.3 Timing analysis 94

4.5 Placement, Clock Tree Synthesis (CTS) and Routing 96
4.5.1 Summary of power, area and timing analysis 102
4.6 Summary 104

CHAPTER 5 – CONCLUSIONS AND FUTURE WORKS

5.0 Conclusions 106

5.1 Future Works 107

REFERENCES 109

APPENDIX A 115

APPENDIX B 117

APPENDIX C 119

APPENDIX D 123

APPENDIX E 124

LIST OF TABLES

Page

Table 1.1 Fundamental scaling limits for conventional MOS devices 2

Table 2.1 Flip-flop & Latch in 32nm Transistor Technology 59

Table 4.1 Power analysis result 93

Table 4.2 Area analysis result 94

Table 4.3 Timing analysis result 96

Table 4.4 Summary of power analysis differences 102

Table 4.5 Summary of area and timing analysis 103

LIST OF FIGURES

Page

Figure 2.1 Trend of inversion Tox and gate leakage vs. Intel technology node 11

Figure 2.2 Transistor scaling impact on performance 12

Figure 2.3 Summary of Intel transistor technology evolution 13

Figure 2.4 Intel transistor leadership in semiconductor industry 13

Figure 2.5 90nm transistor technology with strained SiGe film 14

Figure 2.6 2nd Generation of Strained Silicon 15

Figure 2.7 High-k/metal gate PMOS transistor with the embedded SiGe strain 16

Figure 2.8 2nd Generation of High-k Metal Gate; PMOS and NMOS structure 18

Figure 2.9 Transistor technology; (A) Planar MOSFET, (B) finFET 19

Figure 2.10 Tri-gate transistor; (A) A fin connecting the source and drain, (B) The cross-section view 20

Figure 2.11 Tri-gate transistor; (A) Multiple fin, (B) With metal contact 21

Figure 2.12 TEM image; (A) 32nm planar transistor, (B) 22nm tri-gate transistor 22

Figure 2.13 MOS transistor structure and operation 24

Figure 2.14 Bulk transistor with not fully depleted layer 25

Figure 2.15	Fully Depleted Silicon on Insulator (FDSOI)	26
Figure 2.16	Fully Depleted Tri-gate Transistor	26
Figure 2.17	Planar transistor operation	28
Figure 2.18	Planar versus tri-gate transistor operation	28
Figure 2.19	Tri-gate transistor operation with reduced threshold voltage	29
Figure 2.20	Gate delay for 32nm planar transistor	30
Figure 2.21	Gate delay for 22nm planar transistor	30
Figure 2.22	Gate delay for 22nm tri-gate transistor	31
Figure 2.23	Sequential element symbols: A) Latch symbol, B) Flip-flop symbol	35
Figure 2.24	CMOS positive-level-sensitive D latch: a) Latch structure, b) Latch gate level structure, c) Latch timing diagram	36
Figure 2.25	CMOS positive-edge-triggered D flip-flop: a) Flip-flop structure, b) Flip-flop gate level structure, c) Flip-flop timing diagram	38
Figure 2.26	Latch hold time and setup time	42
Figure 2.27	Clock skew and clock jitter	43
Figure 2.28	MIPS processor with 5 stages pipeline. (Source: MIPS processor implementation. Available: http://www.cs.cmu.edu/afs/cs/academic/class/15740-f97/public/info/pipeline-slide.pdf)	45
Figure 2.29	Time borrowing example: (a) Latch system with no feedback, (b) Latch system with feedback.	46
Figure 2.30	Maximum amount of time borrowing	48

Figure 2.31	Theorem 1: Equivalent circuitry using flip-flops and latches	50
Figure 2.32	Theorem 2: Equivalent circuitry using flip-flops and latches	57
Figure 3.1	Block diagram for flip-flop based circuits	61
Figure 3.2	Block diagram for latch based circuits	61
Figure 3.3	Design flow for flip-flop and latch based circuits	63
Figure 3.4	Block diagram for latch based circuits	65
Figure 3.5	Logic synthesis flow	66
Figure 3.6	LEC conformal verification environments	70
Figure 3.7	Power and area analysis environments in Synopsys's Primetime	73
Figure 3.8	Setup time and hold time	75
Figure 3.9	Timing analysis environments in Synopsys's Primetime	76
Figure 3.10	Place and route flow	78
Figure 3.11	Power and ground ring	79
Figure 3.12	Power and ground straps	80
Figure 4.1	RTL for flip-flop based design	85
Figure 4.2	RTL for latch based design	86
Figure 4.3	Schematic view for flip-flop based design	86
Figure 4.4	Schematic view for latch based design	87
Figure 4.5	Gate-level netlist for latch based design	88

Figure 4.6	Gate-level flip-flop for latch based design	88
Figure 4.7	Timing report of flip-flop based design in Design Compiler	89
Figure 4.8	Timing report of latch based design in Design Compiler	89
Figure 4.9	Functional verification by LEC Conformal	90
Figure 4.10	Power analyses for flip-flop based design by Primetime	92
Figure 4.11	Power analyses for flip-flop based design by Primetime	92
Figure 4.12	Timing analyses for flip-flop based design by Primetime	94
Figure 4.13	Timing analyses for latch based design by Primetime	95
Figure 4.14	Placement for flip-flop based design by IC Compiler	97
Figure 4.15	Placement for latch based design by IC Compiler	97
Figure 4.16	Power network for flip-flop based design by IC Compiler	98
Figure 4.17	Power network for latch based design by IC Compiler	99
Figure 4.18	Routing for flip-flop based design by IC Compiler	100
Figure 4.19	Routing for latch based design by IC Compiler	100
Figure 4.20	Timing report for flip-flop based design in IC Compiler	101
Figure 4.21	Timing report for latch based design in IC Compiler	101
Figure 4.22	The percentage of power differences	103
Figure 4.23	The percentage of die area and timing differences	104

LIST OF ABBREVIATION

ABV	Assertion Based Verification
ALU	Arithmetic Logic Units
AOI	AND-OR-Inverter
BCD	Binary Code Decimal
BFM	Bus Functional Model
CAD	Computer-aided Design
CMOS	Complementary Metal Oxide Semiconductor
CPU	Control Processing Unit
DC	Design Compiler
DDR	Double Data Rate
DMA	Direct Memory Access
DRC	Design Rule Constraints
DSP	Digital Signal Processing
DTA	Dynamic Timing Analysis
ECO	Engineering Change Order
EDA	Electronic Design Automation
FEV	Functional Equivalent Verification
FF	Flip-flop
FPU	Floating Point Unit
FSM	Finite State Machine

FU	Functional Unit
GPU	Graphics Processor Unit
HDL	Hardware Description Language
HP	Hyper Pipelined
I/O	Input/Output
IC	Integrated Circuit
IF	Interface
IMEM	Instruction Memory
IN	Input
IP	Internet Protocol
ISA	Instruction Set Architecture
KB	Kilo Byte
LL	Latch
LMEM	Local Memory
MAC	Multiply/ Accumulate
MB	Mega Byte
MHz	Mega Hertz
MMU	Memory Management Units
MPU	Media Processing Unit
MSI	Message Signaled Interrupt
OAI	OR-AND-Inverter
OBP	Output Buffer Pointer

OUT	Output
PCI	Peripheral Component Interconnect
PD	Process Data
PDA	Personal Digital Assistance
PIC	Programmable Interrupt Controller
PM	Power Management
PO	Pin Output
PT	Prime Time
QDR	Quad Data Rate
RISC	Reduced Instruction Set Computing
RTL	Register Transfer Level
SDRAM	Synchronous Dynamic Random Access Memory
SoC	System-on-Chip
SRAM	Static Random Access Memory
STA	Static Timing Analysis
SV	System Verilog
SVA	System Verilog Assertion
TLB	Translation Lookaside Buffer
TR	Transfer Ring
TRB	Transfer Request Block
VCS	Verilog Compiled code Simulator
VLSI	Very Large Scale Intergration

LIST OF PUBLICATION

N. Azman, M., Asrulnizam, A. M., and C. M. Leong, (2011) “Optimization of Circuitry for Power and Area Efficiency by Using Combination Between Latch and Register”. 2011 International Conference on Computer Applications and Industrial Electronics (ICCAIE 2011), 4-7 December 2011, Penang, Malaysia. pp. 240 – 244.

KAJIAN ASAS TEKNOLOGI TRANSISTOR 22nm KE ATAS LITAR JUJUKAN MENGGUNAKAN PrimeTime

ABSTRAK

Teknologi transistor telah melalui proses skala yang pesat selama lebih daripada 30 tahun. Proses skala transistor adalah penting untuk keperluan kuasa yang rendah dan berprestasi tinggi dalam litar digital. Banyak penyelidikan dan kertas kerja telah dijalankan dan diterbitkan berdasarkan kemajuan teknologi transistor. Proses skala terkini dalam teknologi transistor adalah 22nm *tri-gate* transistor teknologi dan ia masih asing kepada penyelidik terutama di dalam penganalisaan mengenai litar yang mengimplementasikan teknologi tersebut. Kajian ini akan mengungkap ciri-ciri flip-flop dan selak berdasarkan teknologi 22nm *tri-gate* transistor kerana kedua-dua peranti tersebut merupakan komponen yang penting di dalam merekabentuk litar digital. Bagi mencapai objektif tersebut, satu aliran pelaksanaan telah direka dan analisa telah dibuat menggunakan *SYNOPSYS® PrimeTime*. Aliran perlaksanaan ini merangkumi idea litar, *Register Transfer Level (RTL),* sintesis, pengesahan persamaan fungsi, '*place and route*' dan pelbagai analisis sebelum melakukan '*sign-off'*. Litar yang berasaskan flip-flop dan litar yang berasaskan selak telah direka dan dijadikan sebagai kes ujian dan kedua-dua litar tersebut beroperasi dengan frekuensi 1 GHz. Keputusan eksperimen yang diperolehi telah dianalisis dan perbandingan telah dibuat antara litar tersebut berdasarkan ciri-ciri kuasa, luas kawasan dan prestasi. Analisis perbandingan menunjukkan bahawa litar berasaskan selak menghasilkan pengurangan sebanyak 45.39%, 15.95% dan 38.51%

dalam kuasa pensuisan bersih, kuasa sel dalaman dan kuasa kebocoran sel berbanding dengan litar yang berasaskan flip-flop. Manakala bagi keseluruhan penggunaan kuasa, litar berasaskan selak menghasilkan penggunaaan kuasa 21.88% kurang berbanding dengan litar yang berasaskan flip-flop. Selain daripada itu, litar berasaskan selak menunjukkan pengurangan sebanyak 44.70% dalam penggunaan kawasan dan peningkatan sebanyak 28.74% dalam kelajuan keseluruhan litar berbanding dengan litar yang berasaskan flip-flop.

STUDY OF BASIC 22nm TRANSISTOR TECHNOLOGY ON SEQUENTIAL CIRCUIT USING PrimeTime

ABSTRACT

Transistor technology has been going through a rapid scaling for more than three decades. Transistor scaling is essential for the needs of low power and high performance digital circuit. Many research and papers has been conducted and published on transistor technology. The latest in line for transistor scaling was 22nm tri-gate transistor technology and it is still unfamiliar to the researchers, especially in the analysis of the circuit that implements the technology. Therefore, this research will reveal the behavior and characteristics of flip-flop and latch in 22nm tri-gate transistor technology as both devices are an important elements in designing a digital circuit. To achieve that, a flow of implementation has been designed and analysis has been done using SYNOPSYS® PrimeTime. This design flow spans from design idea, Register Transfer Level (RTL), synthesis, functional equivalent verification, place and route and various analysis before sign-off. Flip-flop based circuit and latch based circuit has been design as test cases and both circuits are operating frequency at 1 GHz. The experimental results obtained are analyzed and comparisons are made between latch and flip-flop based circuit on its power, area and performance characteristics. Comparative analysis shows that, latch based design consumed 45.39%, 15.95% and 38.51% less in net switching power, cell internal power and cell leakage power respectively compared to flip-flop based design. For overall power consumption, latch based design consumed

21.88% less power compared to flip-flop based design. Meanwhile, latch based design required 44.70% less in die area and 28.74% improvement in timing properties compared to flip-flop based design.

CHAPTER 1

INTRODUCTION

1.0 Introduction

In the era of transistor scaling, Metal Oxide Semiconductor (MOS) device technology has improved at a dramatic rate (Standard 22nm, 2011). A large part of this success is due to the fact of MOS transistor can be scaled to increasingly smaller dimension, which result in higher performance. Complementary Metal Oxide Semiconductor (CMOS) architecture becomes the dominant technology for integrated circuits mainly because of its ability to improve performance consistently while decreasing power consumption. The scaling of CMOS transistors has been the primary factor driving improvements in microprocessor performance. Transistor delay times have decreased more than 30% for every technology generation resulting in a doubling of microprocessor performance every two years.

Conventional scaling of gate oxide thickness, source/drain extension (SDE), junction depths, and gate lengths have enabled MOS gate dimensions to be reduced from 10um in 1970's to a present day size of 22nm. However with conventional CMOS transistor, this scaling has its limits. The key scaling limits were quantified in Table 1.1 for conventional CMOS transistor (Thompson et al., 1998). A traditional silicon dioxide

(SiO2) insulator has reached fundamental leakage limits due to tunneling for an effective electrical thickness below 2.3nm. Meanwhile, decreasing the MOS junction depth to below 30nm will dramatically increase the source/drain resistance, resulting performance degradation. Besides that, channel doping, channel length and gate length has its limits to avoid higher leakage current. Because of these limitations, new transistor technology such as transistor with strained silicon and high dielectric constant material has been investigated to overcome the problem. Transistor scaling is always associated with power dissipation and performance.

Table 1.1: Fundamental scaling limits for conventional MOS devices (Ghani, 2009)

Feature	Limit	Reason
Oxide Thickness	2.3nm	Leakage (I_{GATE})
Junction Depth	30nm	Resistance (R_{SDE})
Channel Doping	$V_T = 0.25$ V	Leakage (I_{OFF})
Source/Drain Extension (SDE) Under Diffusion	15nm	Resistance (R_{INV})
Channel Length	0.06um	Leakage (I_{OFF})
Gate Length	0.10um	Leakage (I_{OFF})

Power dissipation has become one of the important parameter in digital integrated circuitry technology to produce an intelligent and low power consumption of

microprocessor, sensory system and electronic devices. For high performance systems such as workstations and servers, power consumption per chip is often limited to about 150W. The limitation was associated by the amount of heat that can be managed with advances cooled systems and cost-effective heatsinks (Intel's white paper, 2010). Power consumption becomes major factor in designing the product that operation is based on the capability of battery such as laptops, cell phones and PDA. Recently, Intel Corporate News (Intel's Roadmap, 2011) stated that they want to have a lower power envelope (average power) with less leakage power, for all-day battery life. With vision for low power design, Intel wishes to achieve 15W of power consumption for notebook chips compare to 35W previously and less than 1W for Atom-powered systems-on-a-chip (SoC).

Size and structure of transistor are the two most important factors in today's digital technology. The smaller and more power efficient the transistor, the better it is in term of product to be delivered to the end users. Intel introduced strained silicon (introduced by Intel in 2003), high-k/metal gate (introduced by Intel in 2007) and now the 22nm with 3-Dimension (3-D) transistor in a high volume logic process in 2011. With a smaller, 3-D transistor, Intel can design even more powerful processors with incredible power efficiency (Bohr, 2011). The new technology enables innovative microarchitectures, System on Chip (SoC) designs, and new products from servers and personal computers (PCs) to smart phones, and innovative consumer products. Intel demonstrated a computer using a microprocessor code-named Ivy Bridge, which is the first high-volume chip that will use 3-D transistors (Bohr, 2011). However, the journal

and article paper that relate with circuitry interface by implementing 22nm process technology is still lack in publication.

Sequential circuits are usually designed with flip-flops or latches, which are sometimes called memory elements. They are called memory elements because flip-flop and latch can hold data in certain amount of time. Majority of low performance and mid performance designs were flip-flops based design. Flip-flops are easy to use and are well understood by most designers. Other than that, most of the synthesis tools and timing analyzers supported the flip-flops based design compared to latched based design. Unfortunately in systems with few delays per cycle, the sequencing overhead can consume a large fraction of the cycle. Moreover, many standard cell flip-flops are intentionally rather slow to prevent hold time violations at the expense of greater sequencing overhead.

1.1 Problem Statements

Transistor scaling is essential for the needs of low power and high performance digital circuit. Many research and papers has been conducted and published on transistor technology. The 32nm transistor technology has been introduced in 2009 and studies on the characteristics of flip-flop and latch has been published. The latest in line for transistor scaling was 22nm tri-gate transistor technology and it is still unfamiliar to the researchers. The 22nm tri-gate transistor was design to have better power and performance from its predecessor.

Power and performance of a transistor are dependent on the leakage current, threshold voltage (V_T), operating voltage (V_{DD}) and transistor gate delay. The 22nm tri-gate transistor were design to have 10% improvement in leakage current and 25% improvement in threshold voltage that will contribute to improvement in leakage power and active power respectively compared to 32nm transistor technology (Bohr, 2011). Meanwhile for transistor performance, the 22nm tri-gate transistor will improve 0.2V of operating voltage and 37% of transistor gate delay compared to 32nm transistor (Bohr, 2011). With these improvements, the 22nm tri-gate transistor takes the power and performance of flip-flop and latch into a whole new level. Therefore, this research will unfold the behavior and characteristics of flip-flop and latch in 22nm tri-gate transistor technology that one of important element in sequential circuits.

Flip-flop and latch are extremely important circuit elements in any synchronous digital design. Even though flip-flop and latch has a common basic function, they have different power dissipation, area and timing characteristics. These characteristics will be prominent when flip-flop and latch based circuit are experimented using Intel's 22nm tri-gate transistor technology. The 22nm tri-gate transistor is predicted to have significant advantages over planar transistor in performance, leakage power, overall power at low voltage and switching characteristics.

1.2 The Research Objectives

The aim of this research is to investigate the different of flip-flop based circuit and latch based circuit in Intel's 22nm process technology. The main objectives are as follow:

I. To implement a conventional IC design flow that cover from front-end to back-end process with Intel's 22nm process technology.

II. To design a flip-flop and latch based circuit with an identical functionality in 22nm tri-gate transistor technology.

III. To observe and analyze the difference between flip-flop and latch based circuit in term of power consumption, die area and timing requirement of the design.

1.3 Project Contributions

It is well known that performances and power consumption of any digital circuit depend on the circuit design configuration. As consumers look for powerful yet low-power-consuming devices, there is a clear economic interest in the development of low power circuit design. The main reason behind the development of low power circuits is that many portable devices and their applications require low power dissipation and high throughput. Thus, low power design of digital integrated circuits is currently a rapidly developing field in electrical engineering. Equally importance was

the die size of a chip. Therefore, it is important to establish the optimal power-area-time tradeoff. To achieve it, a careful choice of logic elements and circuit design topology is necessary.

Flip-flop and latch are two sequential elements that widely used in digital circuit. It is important to understand the characteristics of flip-flop and latch as it's contribute a major portion in power consumption, area and performance of a digital circuit. This research will present characteristics of flip-flop and latch in 22nm tri-gate transistor technology. This 22nm tri-gate transistor technology will be introduced by Intel in its upcoming microprocessor with a code-named of Ivy Bridge (Intel's white paper, 2010).

1.4 Structure of Thesis

This thesis was organized into five chapters as follows:

Chapter 2 discusses the theoretical and literature review of the Intel's 22nm tri-gate transistor technology and the characteristics of flip-flop and latch. Details of the tri-gate transistor architecture will be explained in this chapter. Meanwhile, the characteristic of flip-flop and latch includes structure and operations, power dissipation and timing properties, and also the timing equivalent between flip-flop and latch also discussed in this chapter.

Chapter 3 presents the design methodology and implementation of this research. The details of design flow and tools applied will be discussed and elaborated.

Chapter 4 covers the result of the design flow implementation. The result will be analyzed and the comparison between latch and flip-flop will be addressed. All the result and analysis will be discussed and verified.

Chapter 5 outlines the conclusions and future works. This chapter concludes the overall research findings; the comparative analysis of flip-flop and latch in 22nm tri-gate transistor technology.

CHAPTER 2

THEORETICAL AND LITERATURE REVIEW

2.0 Introduction

Intel's Tri-Gate transistor uses three gates wrapped around the silicon channel in a 3-D structure, enabling an unprecedented combination of performance and energy efficiency. Intel designed the new transistor to provide unique ultra-low power benefits for use in handheld devices, like smart phones and tablets, while also delivering improved performance normally expected for high-end processors. This 22nm tri-gate transistor technology will be introduced by Intel in its upcoming microprocessor with a code-named of Ivy Bridge. In this research, sequential circuit will be implemented in 22nm tri-gate transistor.

Sequential can be described as circuit in which the output depends on previous as well as current inputs. These circuit are said to have state or memory, where its hold value or state of ‘1’ or ‘0’ at certain amount of time. Two major elements of sequential circuit are flip-flop and latch. This chapter will describe in details the properties of flip-flop and latch.

2.1 The 22nm Tri-gate Transistor Technology

The 22nm tri-gate transistor technology was introduced by Intel to enable the transistor scaling. Tri-gate transistor structure was a new transistor structure and it is different from its predecessor; the planar transistor. Tri-gate transistor is a 'fin' like architecture which originally from finFET architecture. This 22nm tri-gate transistor technology will be introduced by Intel in its upcoming microprocessor with a code-named of Ivy Bridge (Intel's white paper, 2010).

2.1.1 Intel Technology Library

Since the invention of the transistor in 1947, technology has progressed swiftly, paving the way for powerful, yet cost-effective and energy-efficient products. Continuation of these advances, at the pace dictated by Moore's Law, transistor going through numerous innovations; recent notable ones are strained silicon (introduced by Intel in 2003) and high-k/metal gate (introduced by Intel in 2007). Intel is now about to make yet another radical change in its transistor design, the tri-gate transistor. The tri-gate transistor will deliver an unprecedented combination of performance and energy efficiency in a whole range of computers and handheld devices.

One of the key methods to enable transistor gate length scaling over the past several generations has been to scale the gate oxide thickness. This improves the control of the gate electrode over the channel, enabling both shorter channel lengths and higher

performance. As the gate oxide was scaled the gate leakage increased; this increase in gate leakage was insignificant until the 90nm technology node as shown in Figure 3.1. At the 90nm and 65nm nodes, the scaling of the gate oxide slowed as a result of the power limitations from the increase in gate leakage. Figure 2.1 shows the trend of inversion Tox and gate leakage versus the Intel technology node.

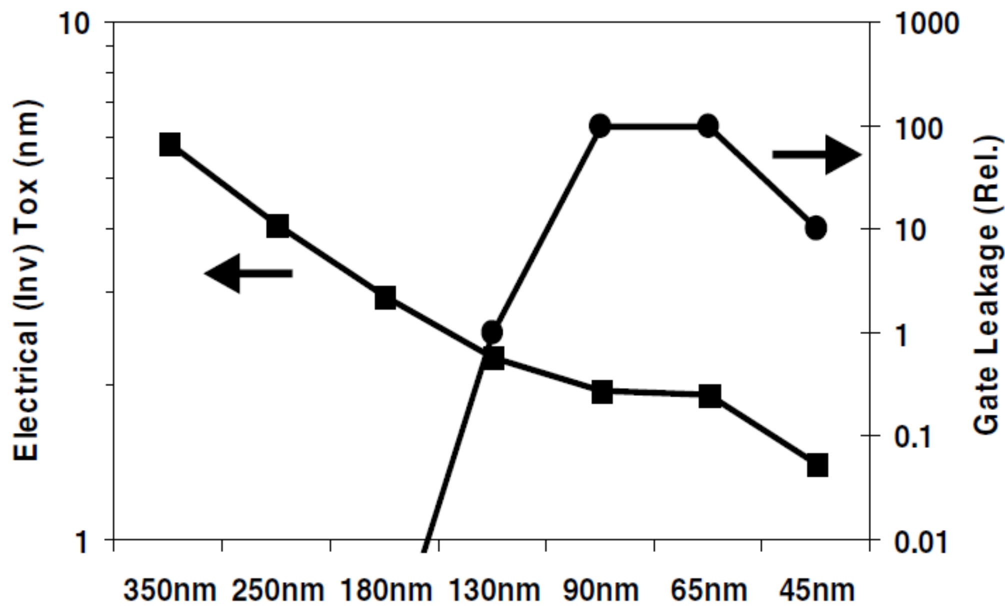

Figure 2.1 Trend of inversion Tox and gate leakage vs. Intel technology node (Intel Presentation, 2011)

As the gate pitch size getting smaller, the performance and drive current for PMOS and NMOS transistor are increased. Figure 2.2 shows the transistor scaling impact on performance. The invention of transistor while scaling the gate length for

leakage efficiency, were also impacting the performance which give better performance and increased the drive current for the transistor.

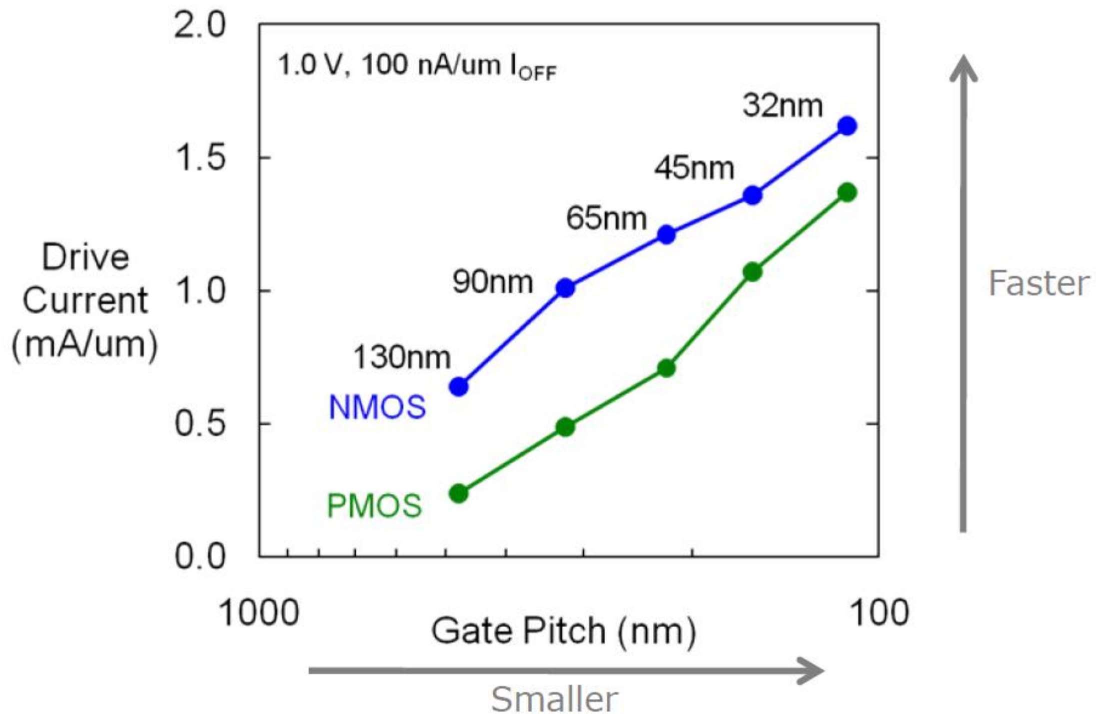

Figure 2.2 Transistor scaling impact on performance (Intel Presentation, 2011)

The summary of Intel transistor technology evolution is shown in Figure 2.3. Continuous innovations in transistor materials and structure are essential to continue scaling down the transistor size. Figure 2.4 show the Intel transistor leadership in semiconductor industry. As can be seen, Intel introduces strained silicon technology 3 years earlier than other major players in transistor technology. Meanwhile for high-k metal gate and tri-gate technology, Intel lead by 3.5 years and 4 years respectively compared to other major players such as TSMC, GlobalFoundries, Samsung and IBM.

Intel leads the industry in introducing new technology generations and revolutionary transistor technology.

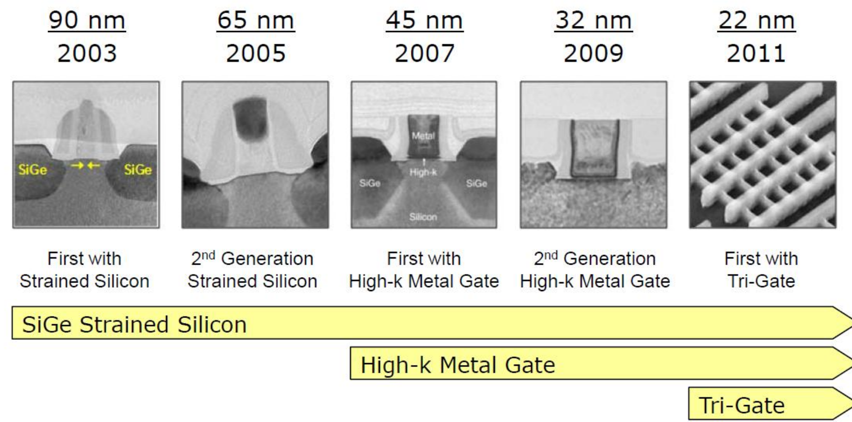

Figure 2.3 Summary of Intel transistor technology evolution (Intel Presentation, 2011)

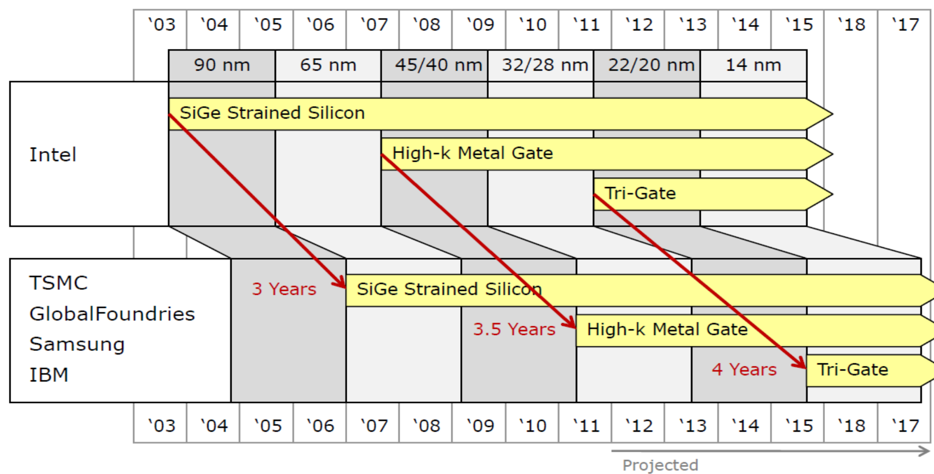

Figure 2.4 Intel transistor leadership in semiconductor industry (Intel Presentation, 2011)

2.1.1.1 90nm Transistor (1st Generation of Strained Silicon)

As MOSFET gate lengths scale below 100nm, it is difficult to maintain high drive current due to mobility degradation. The used of strain effect to improve mobility has been introduced. The unique feature of this transistor was by embedding a compressively strained SiGe (Silicon-Germanium) film in the source drain regions by using a selective epitaxial growth process (Mistry et al., 2004). This strain step is introduced late in the process flow, minimizing integration and defect challenges. The strained silicon NMOS and PMOS transistors have been implemented in a high volume manufacturing 90nm CMOS technology (Mistry et al., 2004). NMOS transistors employ a tensile capping layer to induce strain and improve NMOS drive current by 10%. Meanwhile the Figure 2.5 show a strained SiGe film is embedded into the source drain region to induce compressive strain in the channel region in PMOS transistor. This strained PMOS transistor improved PMOS drive current by 15% (Mistry et al., 2004).

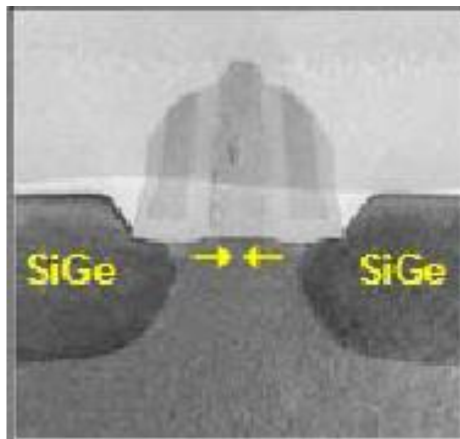

Figure 2.5 90nm transistor technology with strained SiGe film (Intel Presentation, 2011)

2.1.1.2 65nm Transistor (2nd Generation of Strained Silicon)

Intel's 65nm transistor technology still maintained the strained step in improving overall transistor performance (Cappellani et al., 2004). Strained silicon was performed by nickel deposited in the shallow trench isolation around each transistor to stretch the silicon lattice inside the isolation barrier. Strain improves transistor performance by increasing carrier mobility through the lattice. Figure 2.6 show the PMOS strained transistor at 65nm technology. The Electron mobility gain of 35% is observed resulting in 18% drive current improvement. Meanwhile, the Hole mobility is improved by 90% resulting in 50% drive current improvement (Cappellani et al., 2004).

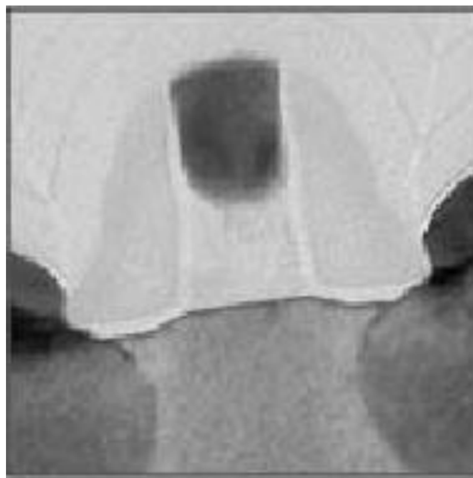

Figure 2.6 2nd Generation of Strained Silicon (Intel Presentation, 2011)

2.1.1.3 45nm Transistor (1st Generation of High-k Metal Gate)

For the 45nm technology node, high-k+metal gate transistors have been introduced for the first time in a high-volume manufacturing process (Auth et al., 2008). The introduction of a high-k gate dielectric enabled a 0.7 times reduction in Tox while reducing gate leakage 1000 times for the PMOS and 25 times for the NMOS transistors. Dual-band edge work function metal gates were introduced, eliminating polysilicon gate depletion and providing compatibility with the high-k gate dielectric (Auth et al., 2008). Figure 2.7 shows a high-k/metal gate PMOS transistor with the embedded SiGe S/D (Source/Drain) strain layer on the PMOS and Ni silicide (Auth et al., 2008). The strained silicon techniques that Intel first introduced at the 90nm and 65nm nodes were further enhanced in this generation. The Ge concentration of the embedded SiGe S/D was increased to 30% from the previous generations of 23% in Intel's 65nm technology (Cappellani et al., 2004) and 17% in the 90nm technology (Mistry et al., 2004).

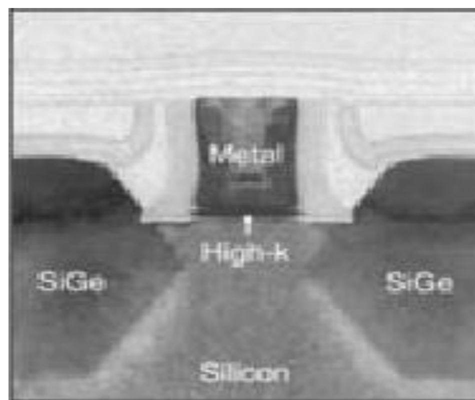

Figure 2.7 High-k/metal gate PMOS transistor with the embedded SiGe strain (Intel Presentation, 2011)

2.1.1.4 32nm Transistor (2nd Generation of High-k Metal Gate)

Intel introduced 32nm process technology with second generation high-k metal gate transistor. This process technology builds upon the tremendously successful 45nm process technology that enabled the launch of Intel® microarchitecture codename Nehalem and the Intel® Core™ i7 processor. This 32nm technology uses 4th generation strained silicon and 2nd generation high-k and replacement metal gate flow (Packan et al., 2009). Using a replacement metal gate flow enables stress enhancement techniques to be in place before removing the poly gate from the transistor. It has been shown that this can further enhance strain and is a key benefit of this flow (Packan et al., 2009). A cross section of the NMOS and PMOS devices are shown in Figure 2.8. The introduction of raised NMOS source and drain regions enables reduced device resistance helping to mitigate the pitch scaling issues discussed above. The proximity of the PMOS SiGe region to the channel continues to be reduced for enhanced channel strain. The combination of 4th generation strained silicon and 2nd generation high-k + metal gate results in PMOS saturated (Vds=1.0V) and linear (Vds=0.05V) drive currents of 1.37mA/um and 0.240mA/um at 1.0V and 100nA/um Ioff (Figure 2.8). These represent a 28% improvement in Idsat and a 35% improvement in Idlin over the 45nm technology (Packan et al., 2009) and are the highest reported drive currents for any 32nm technology.

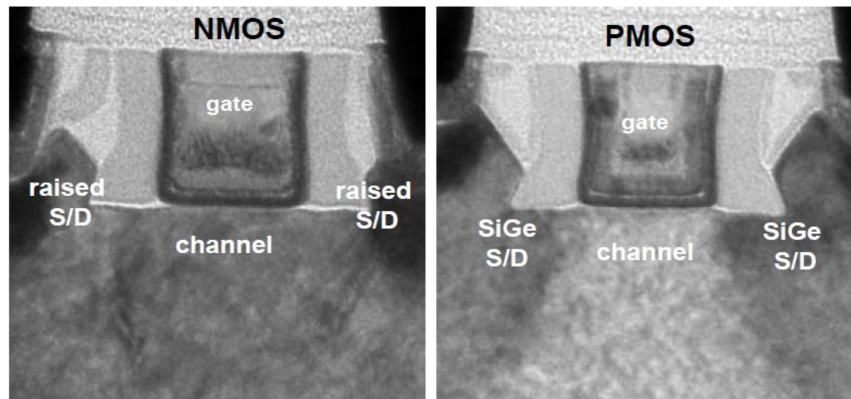

Figure 2.8 2nd Generation of High-k Metal Gate; PMOS and NMOS structure (Intel Presentation, 2011)

2.1.1.5 22nm Tri-Gate transistor technology

The 22nm tri-gate transistor was the latest in line transistor technology by Intel Corporation. For the first time in history, silicon transistors are entering the third dimension. Intel is introducing the tri-gate transistor, in which the transistor channel is raised into the 3rd dimension. Current flow is controlled on 3 sides of the channel (top, left and right) rather than just from the top, as in conventional planar transistors. The net result is much better control of the transistor, maximizing current flow (for best performance) when it is on, and minimizing it (reducing leakage) when it is off (Bohr, 2011). Overall, the 22nm tri-gate transistor will give better performance and power consumption compare to its predecessor. The 22nm tri-gate transistor will be demonstrated in 22nm Intel microarchitecture with codename Ivy Bridge and it will be the first high volume chip to use 3D tri-gate transistors. The 22nm tri-gate transistor technology will be discussed further in the next sub-chapter.

2.1.2 Intel 22nm Tri-Gate Transistor Technology

In the era of classical scaling, transistor performance improved primarily as a result of dimensional scaling. In the past decade, performance has progressed through introduction of transistor architecture innovations, including strained silicon and high-k/metal-gate technologies. While the finFET featured gate control on two sides of a fin, the Tri-Gate transistor extended gate control to three sides of the fin (Tawfik et al., 2008). This 22nm process technology is the first to exploit fin based Tri-Gate devices and combine their benefits with strained silicon and high-k/metal-gate (Bohr, 2011).

Tri-gate transistor is a 'fin' like architecture which originally from finFET ('fin' Field Effect Transistor) architecture. The characteristic of the finFET is the conducting channel is wrapped by thin silicon 'fin', which forms the body of the device. The thickness of the fin (measured in the direction from source to drain) determines the effective channel length of the device. Figure 2.9 shows the normal planar MOSFET and finFET transistor architecture.

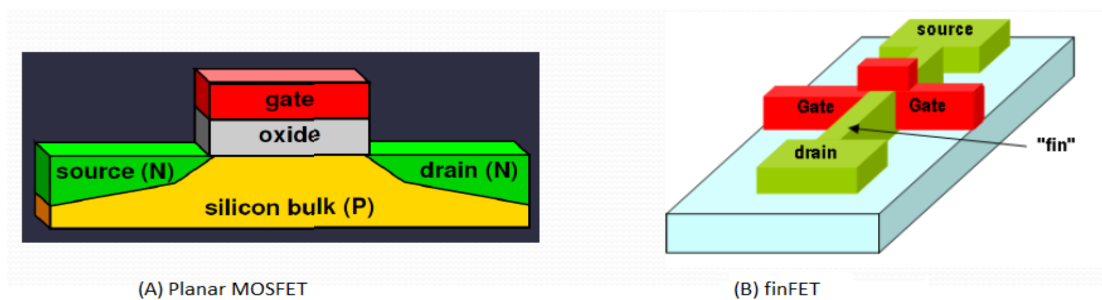

Figure 2.9 Transistor technology; (A) Planar MOSFET, (B) finFET (Bohr, 2011)

2.1.2.1 Tri-gate Transistor Structure

The structure of the tri-gate transistor was evolved from traditional planar transistor. The additional feature, the ‘fin’, was introduced between the source and drain to become the tri-gate transistor. Figure 2.10(A) shows a tri-gate transistor with a fin connecting the source and drain, while Figure 2.10(B) shows the cross-section view of the tri-gate transistor. High-k dielectric still exists in the tri-gate transistor structure and it is depicted as a yellow stripe in the Figure 2.10(A). The conducting channel was form on the three sides of a vertical silicon fin as shown in Figure 2.10(A). This three sides conducting channel will improved the sub-threshold slope resulting an improved in sub-threshold voltage. The increased of the conducting channel area will also increase the inversion layer area which will give a higher drive current (Bohr, 2011). The additional fin on the tri-gate transistor structure will added the process cost of 2-3% only (Bohr, 2011).

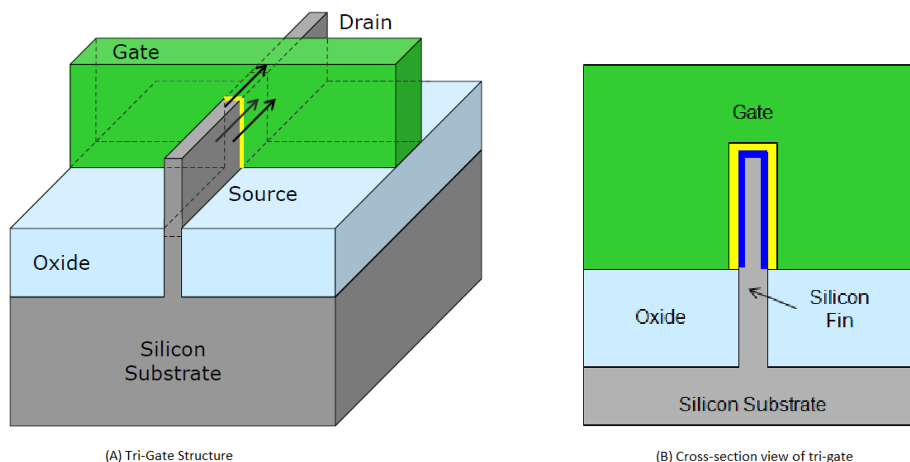

Figure 2.10 Tri-gate transistor; (A) A fin connecting the source and drain, (B) The cross-section view (Bohr, 2011)

Tri-gate transistor can be connected together with multiple fins as shown in Figure 2.11(A). With multiple fins structure, allow the tri-gate transistor to have higher drive current and higher performance. Meanwhile Figure 2.11(B) shows the metal contact added into tri-gate transistor structure with multiple fins.

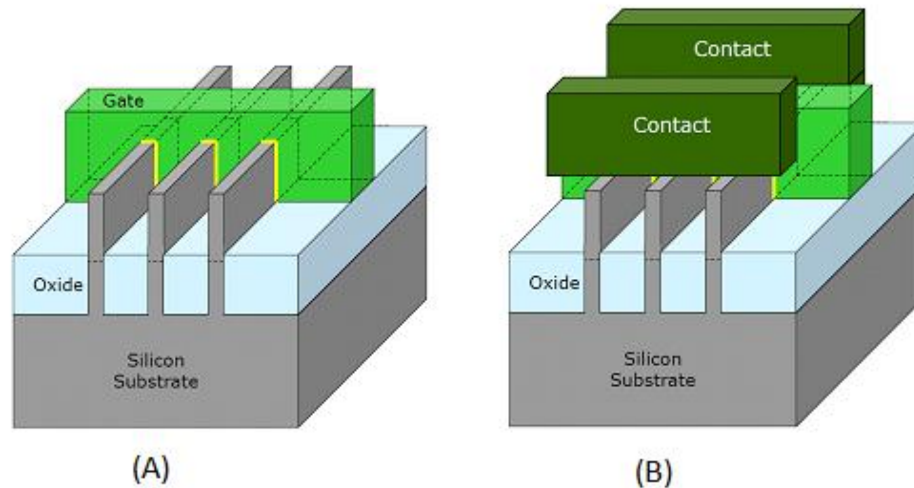

Figure 2.11 Tri-gate transistor; (A) Multiple fins, (B) With metal contact (Bohr, 2011)

Figure 2.13 shows Transmission Electron Microscopy (TEM) image of 32nm planar transistor and 22nm tri-gate transistor. TEM is a microscopy technique whereby a beam of electrons is transmitted through an ultra-thin specimen, interacting with the specimen as it passes through (Nobelprize.Org, 2011). An image is formed from the interaction of the electrons transmitted through the specimen. TEMs are capable of imaging at a significantly higher resolution than light microscopes. From Figure 2.12(A), gates structure is visible on top of the substrate in 32 nm planar transistor.

Meanwhile in Figure 2.12(B), the gate and fin structure are visible on top of the substrate for 22nm tri-gate transistor (Bohr, 2011).

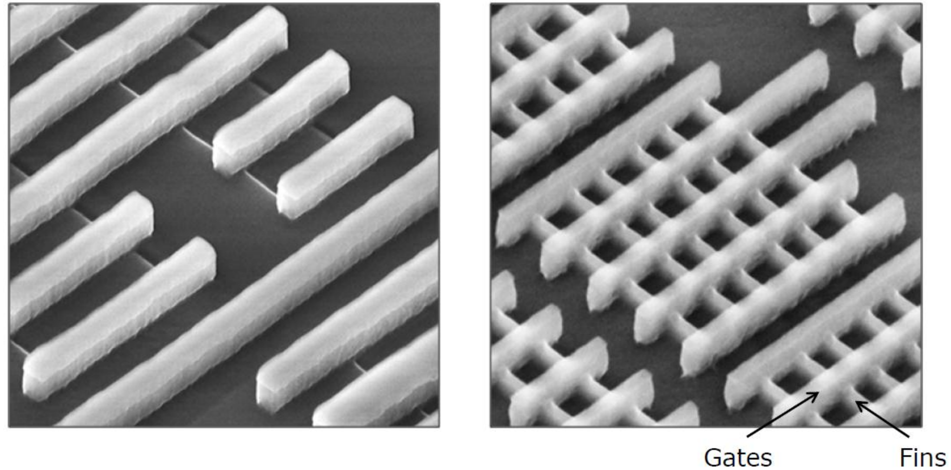

Figure 2.12 TEM image; (A) 32nm planar transistor, (B) 22nm tri-gate transistor (Bohr, 2011)

2.1.2.2 Tri-gate Performance and Power Benefits

Tri-gate transistor was design to give a better performance and power compare to existing transistor. With this smaller, performance and power efficient 3D-transistor, Intel can design even more powerful processors with incredible power efficiency. The performance and power characteristics of 22nm tri-gate transistor will be described in the following sub sections.

2.1.2.2.1 MOS Transistor Theory

The MOS transistor is a majority-carrier device in which the current in conducting channel between the source and drain is controlled by voltage applied to the gate. In an nMOS transistor, the majority carriers are electrons while in pMOS transistor, the majority carriers are holes. Figure 2.13 shows a simple MOS structure (Lee, 2010). The top layer of the structure is a conductor called gate. The middle layer is a very thin insulating film SiO_2 called the gate oxide. The bottom layer is the doped silicon body. The figure shows a p-type body in which the carriers are holes. The body is grounded and a voltage is applied to the gate. The gate oxide is a good insulator so almost zeros current flows from the gate to the body (Lee, 2010).

In Figure 2.13(a), a negative voltage is applied to the gate, so there is negative charge on the gate. The mobile positively charged holes are attracted to the region beneath the gate. This is called accumulation mode. In Figure 2.13(b), a low positive voltage is applied to the gate, resulting in some positive charge on the gate. The holes in the body are repelled from the region directly beneath the gate resulting a depletion region forming below the gate. In Figure 2.13(c), a higher positive potential exceeding a critical threshold voltage V_t is applied, attracting more positive charge to the gate. The holes are repelled further and a small number of free electrons in the body are attracted to the region beneath the gate. This conductive layer of electrons in p-type body is called the inversion layer. The threshold voltage depends on the number of dopants in the body and the thickness t_{ox} of the oxide (Lee, 2010).

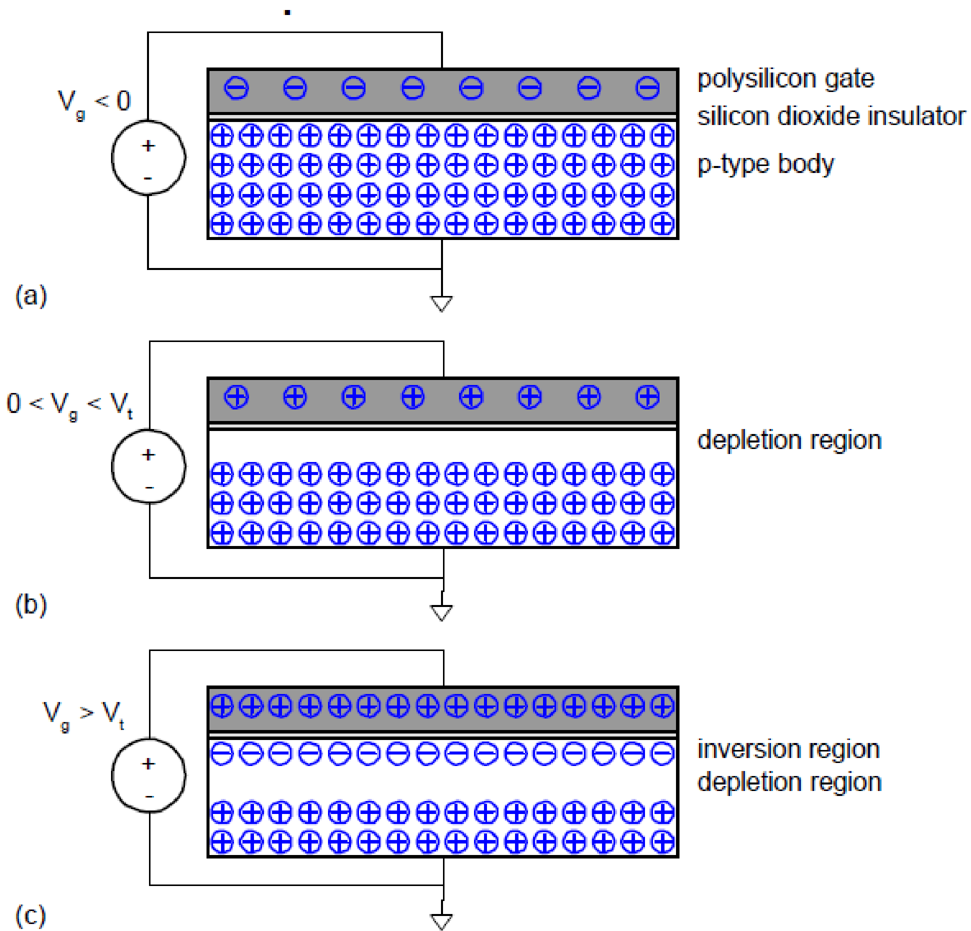

Figure 2.13 MOS transistor structure and operation (Lee, 2010)

2.1.2.2.2 Fully Depleted Tri-Gate Transistor

Tri-gate transistor implements a fully depleted layer in its architecture. This fully depleted layer was evolved from bulk transistor as explained in previous sub chapter. Figure 2.14 shows the bulk transistor with not fully depleted operation. The substrate voltage exerts some of electrical influence on the inversion layer (where the source to drain current flows).