Scaled MOSFET Evolution with Device Physics Based Assessment

Conventional and Mainstream MOSFET Device Architectures

Authors: You will notice that some questions/corrections offered by the copyeditor are marked with “Comp”: This denotes that the copyeditor believes the question is one for the compositor rather than you, the author. However, please review all such notes and state any disagreement you might have regarding them. If you agree with the plan proposed under such an item, do not do anything and the compositor will implement the suggested changes. If you disagree with the plan proposed under such an item, please delete the note for that item.

Synthesis Lectures on Emerging Engineering Technologies

Editor
Kris Iniewski, *CMOS ET*

Scaled MOSFET Evolution with Device Physics Based Assessment: Conventional and Mainstream MOSFET Device Architectures
Nabil Shovon Ashraf

Advanced Concepts and Architectures for Plasma-Enabled Material Processing
Oleg O. Baranov, Igor Levchenko, Shuyan Xu, and Kateryna Bazaka

Low Substrate Temperature Modeling Outlook of Scaled n-MOSFET
Nabil Shovon Ashraf

Radiation Imaging Detectors Using SOI Technology
Yasuo Arai and Ikuo Kurachi

Layout Techniques for MOSFETs
Salvador Pinillos Gimenez

Resistive Random Access Memory (RRAM)
Shimeng Yu

The Digital Revolution
Bob Merritt

Compound Semiconductor Materials and Devices
Zhaojun Liu, Tongde Huang, Qiang Li, Xing Lu, and Xinbo Zou

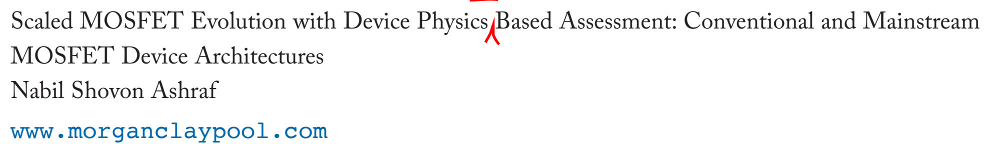
New Prospects of Integrating Low Substrate Temperatures with Scaling-Sustained Device Architectural Innovation
Nabil Shovon Ashraf, Shawon Alam, and Mohaiminul Alam

Advances in Reflectometric Sensing for Industrial Applications
Andrea Cataldo, Egidio De Benedetto, and Giuseppe Cannazza

Sustaining Moore's Law: Uncertainty Leading to a Certainty of IoT Revolution
Apek Mulay

Copyright © 2022 by Morgan & Claypool

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other except for brief quotations in printed reviews, without the prior permission of the publisher.



Scaled MOSFET Evolution with Device Physics Based Assessment: Conventional and Mainstream MOSFET Device Architectures
Nabil Shovon Ashraf
www.morganclaypool.com

ISBN: 9781636392745 Paperback
ISBN: 9781636392752 PDF
ISBN: 9781636392769 Hardcover

DOI 10.2200/S01138ED1V01Y202110EET012

A Publication in the Morgan & Claypool Publishers series
SYNTHESIS LECTURES ON EMERGING ENGINEERING TECHNOLOGIES
Lecture #12
Series Editor: Kris Iniewski, CMOS ET

Series ISSN 2381-1412 Print 2381-1439 Electronic

Scaled MOSFET Evolution with Device Physics Based Assessment

Conventional and Mainstream MOSFET Device Architectures

Nabil Shovon Ashraf
North South University, Dhaka, Bangladesh

SYNTHESIS LECTURES ON EMERGING ENGINEERING TECHNOLOGIES #12

MORGAN & CLAYPOOL PUBLISHERS

ABSTRACT

Toward sustenance of Moore's Law based evolution, semiconductor device engineering professionals and researchers have successfully envisioned derivatives of conventional and other mainstream device architectures that have been efficiently integrated into high-volume batch manufacturing yield. Due to the scaling impacted architectural and device parametric based innovation that was imperative at every process and technology node, crucial and deepening understanding of device performance improvement projected at every node by well-probed device physics based analytical reasoning are of utmost requirement to enable highly reliable, fault-resilient and high yield manufacturing of today's complex device architectures. For conventional bulk MOSFET below 100 nm node, new aspects on scaling issues such as advanced lithography, source and drain contact resistance, high field transport, ultra-shallow junction technology and 2D scaling length optimization are becoming mainstream concerns. For III-V high-mobility FETs, inclusive of conventional MOSFET device physical features, additional effect of interface trap density on drive current is receiving special attention. For silicon-on-insulator (SOI) devices, device physical treatments of self-heating in thin silicon film, and quantum confinement effect of thin film over buried oxide (BOX) are discussed along with other relevant device physical concerns. For strained silicon n and p MOSFETs device physical analysis is provided on strain's effect on mobility and drive current. For multigate transistors including nanowire architectures, physical attributes on gate integrity, volume inversion and nanowire core-shell architecture and pore radius effect on drive current are notably discussed. For tunnel FET, device physical attributes of enhancing drive current mechanisms are discussed with considerations of inter-band tunneling kinetics and ambipolar tunneling reduction and specially significant trap-assisted tunneling (TAT) process.

KEYWORDS

semiconductor device engineering, device physics, device architechtures, MOSFET, FETs

Contents

Foreword xi

Acknowledgments xv

Introduction 1

1 Device Physics-Based Scaling Insights on Conventional Short Channel n-type MOSFETs 29

1.1 Device Physical Attributes of Contact Resistance for Scaled n-Channel Conventional MOSFET 40

2 Device Physics-Based Scaling Insights on III-V Short Channel n-Type MOSFETs and its Derivatives 45

3 Device Physics-Based Scaling Insights on Silicon-on-Insulator (SOI)-Based Short Channel n-Type MOSFETs 55

3.1 Self-Heating in FDSOI MOSFETs Its Origin and Impact on Device Performance 65

4 Device Physics-Based Scaling Insights on Strained Si-Based Short Channel n-Type MOSFETs 69

4.1 Important Observations on the Imprecise Practice of Quoting Effective Masses for Electrons and Holes in Si at T = 300 K 78

5 Device Physics-Based Scaling Insights on Multiple Gate and Surrounding Gate nm Dimension n-Type MOSFETs 83

6 Device Physics-Based Scaling Insights on Tunneling Based n-Type MOSFET Device Architectures 91

7 Conclusions 99

References 103

Author's Biography: 123

Foreword

There are gazillions of articles on semiconductor de[illegible]ng and performance improvement in-line with Moore's Law-based scaling node evolution. But the author discovered the need of device physics based thoroughly probed realization of modeling aspects and reliability aspects of these devices specially when considering ultra short MOSFET devices where transport is based on non-equilibrium kinetics and non-stationary transport dominates the modeling treatment of these devices. The author append[illegible]nce list at the end of the book with references for textbooks and other research based books, plethora of articles on semiconductor device scaling structural and parametric improvization with different device architectures currently in manufactured state other than conventional silicon MOSFETs. The author selected a few papers relevant for discussion in Chapter 1–6 and enumerated the main findings and outcomes of these papers supplemented by the author 's own assessment and realization with device physical attributes always being the central theme. For instance, Chapter 1 is focused on device physics based assessment of conventional MOSFETs and today's most critical device performance hindrances emanate from how nanometer centric resolution of critical dimension (CD) can be enabled by today's most sophisticated and evolved lithographic techniques and instruments and as author, I found that interferometric lithography, self-assembly, thermal scanning probe lithography are potential contenders with extreme ultraviolet lithography and application of computational lithography to each of these tools will accommodate the device manufacturers to optimize their critical dimension resolution. Next in Chapter 1 as a sub-se[illegible]dentifies the contact technology or contact resistance of source, drain, and access regions as another hindrance when nanometer sized device will be probed at contact points for drive current measurement alog with other parameters extraction. The author, after reviewing the excellent articles on contact technology in silicon and other III-V MOSFETs, originated his own understanding that only those metal materials degenerately doped but do [illegible]hen silicon's number of atoms per cell 5 × $10^{22}/cm^3$ can be chosen to form salicided contacts on source and drain contact junctions and th[illegible]e is a need to study the crystal structure of these different metals referred in various references and we have to know the number of atoms per cell in these metals also so that salicided ontacts retain stability from performance point of view. Besides, those metal salicides for which metal atomic radius is such that they can occupy interstitial positions in silicon crystal should be preferred as degenerate doping will tend to drive the salicided structure out of stability and this we should keep in mind. I[illegible] Chapter 2, the author found that III-V materials need mo[illegible]h focus from the device engineers as they generate high carrier mobility and drift velocity but some of these materials have low effective mass

for electrons which reduce the number of density of states (DOS) and maximum possible inversion carrier density at chosen gate and drain bias but their mobility advantage is fundamentally strong to provide good on-current at room temperature and higher on-current at reduced substrate temperatures. In Chapter 3, the author found that ultra thin body (UTB) silicon-on-insulator (SOI) MOSFET with double gate and multigate structures can also provide high on current with better short channel immunity. And device physics based assessment of key reference articles for this device are analyzed for better understanding of the readers. In Chapter 4, strained silicon and SOI MOSFET structures are discussed with information from references how compressive and tensile strain decrease the longitudinal m_l and transverse effective mass m_t in silicon to impact the gain in carrier mobility. After more close probing I found that most of these m_l and m_t parameters are wrongly referenced at T = 300 K while their actually selected values are at 4 K. So, I found two equations based on m_l and m_t variation with temperatures from 4 K to 500 K should be analytically equation based formulated along with density of effective mass mn for electron and mp for hole from 4 K to 500 K, otherwise the mobility enhancement values reported in these references by application of strain may be overvalued or undervalued. In Chapter 5, multi gate structures with Fin-FET a 3 D variant of multigate structures are discussed along with nanowire form with wrapped around cylindrical gate. These devices show very high I_{on} and low I_{off} with low substrate doping and ideal for 10 nm scaled node operations. In Chapter 6, steep slope device with ultra low power operation capable tunnel FET structure is studied from various reference sources. Discussion centered on staggered or broken gap III-V Tunnel FET structures for higher on current and interband tunneling physics applied to improve the tunneling efficiency. The minority carriers which are generally in lower density and therefore result in lower I_{on} but gate-all-around nanowire tunnel FET is still suitable for architectures and cores that do not consume more power and require less switching energy from logic transitions, hence less drive current. As an author of this book, I also found various articles that focus on superior performance of MOSFET devices at lower temperatures down to cryogenic level but while the mobility gain is really multifold, incomplete ionization may be strong in these devices if the substrate doping is not high and must be degenerate level at $10^{20}/cm^3$ so that even after carrier freeze out at low temperatures, substantial fraction of carriers will remain ionized and comparable high I_{on} current can still be realizable in these devices at low temperatures. Also as an author citing some references I postulated the observation that when a device is degenerately doped with donors and acceptors, these donor and acceptor sites become increasingly indistinguishable from one another forming a band and may overlap with conduction band energy minima E_c, so E_c-E_d difference is virtually non-existent and almost all degenerately doped carrier remain ionized to full 100 % in all devices Si, Ge, and III-V when considered but textbook equations when applied show severe freeze out or low ionization percentage for degenerate doping levels, so a transition point needs to be determined below which we can apply the textbook ionization equations but above this transition point of doping value, analytical equations may be

AU: This sentence is entirely too long. I've begun to break it up but the highlighted section shows the text that is too unclear for me to edit/break up further. Please revise.

semi-empirical need to be constructed so that proper ionization percentage can be calculated. In this book, I also found the need of modeling neutral impurity based scattering for lower substrate temperatures than 300 K for donor and acceptor doping as references are very scarce on this where for low temperature mobi[illegible]ulation, neutral impurity based scattering has been incorporated with other scattering events. In the Conclusions, I focused on the backscattering reduction of ballistic n-MOSFET with gate length in 10 nm range. I proposed an engineering of channel potential in the gate to drain side and immediately adjacent to the gate to drain profile within 0.2 of gate length L counted from source point. By providing a slightly higher doping that raises the channel potential in the vicinity of gate-to-drain potential regions where the carriers usually backscatter to the source, the barrier for back scattering can be raised so that there will be fewer carriers that backscatter to the source and ballistic current will be improved along with ballistic source injection velocity at the source inception point. The doping near the source then can be further graded and aligned with original substrate doping somewhere from mid channel position to the drain side allowing for drift field enhancement in addition to ballistic transport and this suggestion of mine also came after reading the article of Mark Lundstrom where he found that significant backscattering to source does not allow ballistic transport to translate into high drive current. With this note to the readers, I conclude my Foreword to this book.

Acknowledgments

The author would like to gracefully thank the Morgan & Claypool publishing staff members for initiating the contract of the book. The author would also like to thank the Librarian of North South University for providing access to the high quality textbooks, research books, journals, and conference articles by his own collaboration and also from North South University subscription.

Introduction

Comprehensive and minutely reasoned knowledge and understanding of device physics based principles, device parametric and structural features with associated attributes are almost singularly pivotal to ensuring reliable and efficient performance of today's evolved scaled device architectures integrated in giga-scale array sustaining Moore's Law-based trend line (referring the readers to these excellent text and research books [1–24]). Even though modeling engineers over the years have developed robust and near-precise numerical device performance and parametric modeling equations based simulatory software, the structural and parametric constraints imposed by giga-scale manufacturing process sequences readily cause anomaly and aberration in device performance testing conducted by the test and product engineers. Therefore in the vicinity of ultra-scaled Moore's technology node, the reliable performance of these field effect device arrays in chips and systems fails the yield window at a rapid rate which is the fundamental reason for the device physicists to probe deeper into all-inclusive device physical attributes of function FET including scaling-based architectural evolution and engineered adaptation. The key conceptual understanding derived by analytical and reasoning based device physical analysis and study, centering on extracting the real causes behind operational mal-performances of highly integrated devices, forms the benchmark standard with which the device modeling numerical software can be adjusted for coherent performance simulation at the front-end-of-line (FEOL) of the device manufacturing process sequences and also for properly regulating the testing and reliability-aware software developed by reliability and test engineers at the back-end-of-line (BEOL) of the device manufacturing process sequences. When almost all device physical perspectives for different evolved scaled device architectures are identified, taken into account and analyzed for conceptual development and understanding, reliable performance and operation of GSI chips and systems can be ensured with consistent batch manufacturing yield (referring the readers to these excellent research articles and some of these articles employ the most plausible option of ensuring device scaling at 10 nm node and below by operating device near 4 K substrate temperatures [25–81]). With this fundamental note in focus, Volume I? of this book systematically explores and elucidates the fundamentally well-developed device physical concepts and attributes of conventional MOSFET at the scaled technology node, high mobility III-V and Ge FETs, silicon-on-insulator FET device architectures, strained-Si FET devices, multigate and 3D-FETs including FinFET, double-gate (DG) FET and gate-all-around nanowire FET devices, and finally tunnel FET devices all of which have been production-enabled as comprising mainstream device architectures and have undergone high-volume manufacturing based product development continued until today. Other key device physical attributes

AU: We are currently in the Introductory chapter so do you mean Chapter 1? Or do you mean 'In this Introduction...'?

and principles, that are extracted scussed in certain referenced articles in connection to accurately targeting precise reliable operation of giga-scale device array, but which are still lagging sound and well-developed factual formulations associated with different FET device architectures are specifically described and analyzed with needed reserved attention. Now the introductory chapter and exposition of this volume I series of the book is sub-divided in paragraphed array highlighting the focus and content attributes of the different chapters of this volume I series arranged chronologically.

giga-scale integration (GSI)

For the conventional MOSFET devices, one of the fundamental limitations from the point of scaled node operation is the ability to define precise critical dimension (CD) of the feature size from the mask to wafer exploiting advanced photolithographic process and techniques wit summarized sophisticated and intricately coordinated instrumentation and metrology. As Lloyd R. Hariott in his classic review paper in *Proceedings of the IEEE*, March 2001, titled "Limits of Lithography" [82] summarized, roughly half of the density improvements in today's GSI have been enabled by the improvements in lithographic resolution capability. The resolution of a lithography system is: (i) directly constrained by the complexity or number of process steps being used in device architectural fabrication; (ii) directly constrained by the dimension of the wavelength of the optical projection system; and (iii) inversely constrained by the numerical aperture (NA) of the lenses used in optical projection system. Therefore the smallest feature size that can be printed on a die in a wafer is almost the same order of the smallest dimension of the wave length of light constrained by the Bragg's diffraction limit of $\lambda/2$. From the context of image size reduction, in the regime where feature sizes that are required to be printed on a die are less than the exposure wave length, the non-linearity in imaging process introduces a complication referred to as mask error enhancement factor (MEEF). MEEF is an indicator which measures the relative deviation in critical dimension due to a deviation of the feature pattern on the mask which can be enhanced due to the stringent multi-step process in pattern transfer. While numerical aperture (NA) can be upshifted to allow improved resolution, the depth-of-focus (DOF) decreases at a rate $(NA)^{-2}$ causing deficiency in controlling adequate image transfer and CD uniformity by enhanced scattering due to the decrease in absorption depth through the photoresist. High-NA lenses are also not economical in pattern replication and require refractive index manipulation such as water-immersed lithography and frequently requires structural design and shaping adding weight and size to the lenses. Certain resolution enhancement technologies (RET), for instance, optimal proximity correction (OPC) and multiple patterning will be discussed in connection to Chapter 1 write-up on the advanced photolithographic processes and techniques. For the next generation lithography (NGL) techniques, extreme ultra-violet lithography (EUVL), electron-beam lithography (EBL), focused ion beam lithography (FIBL), nanoimprint lithography (NIL), directed self-assembly (DSA), scanning probe lithography (SPL), and thermal scanning probe lithography (t-SPL) will be potential lithographic techniques with physics of lithographic attributes for all these versatile technologies available in

the following references [83–94] . Although sparsely discussed as high volume photolithographic manufacturing tool, a burgeoning arena of research focused upon quantum lithography exists where quantum interferometric optical lithography reduces the minimum feature size to λ/2N in an N-photon absorbing substrate. Computational lithography and modeling add additional RET devising techniques by development of intricate mathematical models of the various physical phenomena taking part in the transferring of the layout patterns onto the wafer surface using an exhaustive synergistic approach. Additionally computational metrology based on scatterometry involving conventional reflectometry or spectroscopic ellipsometry is inferred for the readers in relation to Chapter 1. As the references above on computational lithography modeling schemes enable efficient improvements in the overlay accuracy and systematically maintains the stringent CD uniformity requirements across the number of die per wafer. The reades are here mainly made abreast of issues that are determinants of conventional MOSFET bottleneck at the limits of scaling and therefore horizons are identified and the references will guide the readers on developed technology on this.

According to the author, after overcoming the sub-nm CD patterning and definition challenges, the second most taxing and intriguing bottleneck for reliable operation of ultra-scaled FET devices will arise from raised or elevated contact resistances of source-drain junctions, between sidewall spacer and S/D junctions, source to channel access resistances, and gate contact to high-k dielectric. As stated in this classic reference paper, aside from being able to reduce the physical dimensions of the contact, it is paramount to minimize the metal/semiconductor intrinsic resistivity otherwise which a major road block exists to switch operation like maximum drive current, minimum off current and I_{on}/I_{off} ratio. Recently, the cumulative efforts of the scientific and engineering communities have substantially extended the physical limits of the S/D contact resistances thanks to the ingenious process optimization techniques to near 100% activation of highly degenerate dopant concentrations in S/D junctions and to the engineering of their Schottky barrier. This sub-section of Chapter 1 thus highlights key physics and chemistry concepts behind Schottky Barrier Height (SBH), its inhomogeneities and impact on S/D contact resistance (most of these physical aspects are covered in references [95–105] Critical device physical aspects of dipole formation due to band polarization at the metal-semiconductor interface and metal-induced gap states (MIGS) on Fermi level pinning effect are discussed. This sub-section also enunciates theoretical analysis of formation mechanisms for ohmic contacts with the allowance made for reduction of surface states charges at high doping levels. Considerations are given to experimental findings that at a given carrier concentration fast approaching the maximum density of semiconductor atoms per cell, there is a lower limit below which the contact resistivity cannot be reduced any further. This lower limit results from the finite interface transmission probability, the finite carrier velocity and the finite number of interface gap states (IGS) available for carrier trapping based transport across the metal-semiconductor interface. The role of intrinsic transmission probability of electrons being

injected from the highly band-bent metal-semiconductor interface further depends on the Fermi energy of the metal and on the effective masses of the metal and semiconductor and are discussed in this sub-section. Temperature dependence of contact resistivity for ohmic contacts are also discussed with lower than 300 K operation showing an advantage of reduced contact resistance extracted at sub-300 K temperature. Various saliciding techniques and dopant segregation techniques along with dangling bons passivating agents like sulphur and selenium of S/D contacts and gate contacts in providing interfacial dipole induced SBH tuning are discussed. Effects of ultrashallower source drain extension, polysilicon gate depletion effects, lateral and vertical doping gradient effects, relatively enlarged sidewall length, high-k dielectric sidewall effect on series resistance of nanoscale CMOS are also analyzed from a good selected reference articles. The reades are here mainly made abreast of contact resistance issues that are determinants of conventional MOSFET bottleneck at the limits of scaling (some of these contents introduced here in this Introduction will be understandable from the references related to contact resistance mentioned above).

Stationary and non-stationary transport characteristics of ultra-scaled MOSFET are also important including ballistic and semi-ballistic transport in room temperatures and around 4 K, the references will guide the readers more on this [106–116]. As illustrated in their classic paper entitled "Current Transport in Nanoelectronic Semiconductor Devices" by S. Selberherr et al. [116], technology CAD (TCAD) tools are designed to assist in development and engineering at all stages ranging from process simulation to device and circuit optimization. Due to the aggressive downscaling of CMOS device feature sizes and newly emerging nanoelectronic devices, various shortcomings of presently applied TCAD tools appear. These tools are frequently based on semiclassical macroscopic transport models. However inaccuracies originate from the non-local nature of carrier propagation in ultra-scaled devices. Classical non-localities appear when the mean free path between scattering is comparable to the device feature size. Quantum mechanical non-local effects start to determine the transport properties when the device channel length is of the order of the De-Broglie electron wavelength. Size quantization of carrier motion in inversion layers of MOSFETs and in ultra scaled multigate devices as well as tunneling current, including the gate leakage current are the most important examples of quantum effects in MOSFETs. After documenting semiclassical transport model, quantum ballistic transport model making carrier scattering irrelevant is discussed. Finally, dissipative quantum transport theory is discussed which represents the most complete description of transport combining the coherent carrier motion between the scattering events with coherence (or phase) breaking due to carrier scattering. The ballistic saturation velocity in a nanoscale MOSFET is further revealed to be limited to Fermi velocity in a degenerately induced channel appropriate for the quasi two dimensional nature of the inverted channel. Quantum confinement effect degrades the channel mobility out of the confining gate electric field as well as increases the effective thickness of the gate oxide thus making the total gate capacitance responsible for sheet inversion density lower than acceptable as per scaled

structure. Next a classic paper authored by G. Iannaccone et al. is analyzed and referenced with the observation that in this paper the authors provided a physically based model valid in any operating regime and for any ratio of the channel length L to the mean free path λ describing the whole intermediate regime from drift-diffusion to ballistic transport in MOSFETs on the basis of Büttiker model of virtual probes. In part II of this paper, the authors investigated the effects of degeneracy for accurately describing vertical electrostatics and carrier transport where degeneracy's impact on inversion layer capacitance adds an additional cause of degradation for size-quantized ballistic MOSFET. The transport model as described by the author takes account of the effect of degeneracy capacitance on electrostatics, mobility reduction caused by degenerate statistics, triangular confinement in bulk MOSFETs, variable depletion charge in bulk MOSFETs and the transition from drift-diffusion to ballistic transport regime. Taking referen[illegible]n classic Kenji Natori paper "Ballistic Metal-Oxide-Semiconductor Field Effect Transistor," ballistic transport device physics and intrinsic ballisticity of carrier transport are referred and analyzed further to deduce key inferences from the paper. Using published article by Eric Pop et al. titled "Heat Generation and Transport in Nanometer-Scale Transistors" [111], the daunting problem of increasing thermal resistance in thin bulk silicon film due to heat generation and conduction mechanisms is also referred and analyzed to abreast the readers about certain deleterious impact of increasing thermal resistance of bulk ultrathin semiconductor film constrained by gate stack and source-drain alignment giving way to self-heating induced mobility and transconductance degradation which is particularly important for FDSOI, FinFET and GAA NWFET devices. Although almost all scattering processes present in ultrascaled MOSFET at high vertical and lateral field are identified and researched comprehensively, a particularly important scattering process neutral impurity scattering has received less attention in terms of its relative contribution in derivation of experimentally characterized channel mobility as a function of vertical gate electric field for 300 K and 77 K in a classic paper by S. Takagi et al. For highly degenerately doped channel as is usually the case with extremely scaled MOSFETs, both at low and moderate to high temperature a significant fraction of dopants in the channel remain unionized. The proper modeling of these neutral impurity scattering impacted mobility at the Coulomb or ionized impurity scattering regime (low vertical field) and surface roughness scattering regime (high vertical field considering variation of dopant ionization of degenerately doped channel) is extremely important and its physical process will be covered in this book to inspire the modeling engineers to take into account neutral impurity scattering induced [illegible]ge specially for recent research area like cryogenic operation of MOSFETs. Due to shortage of space of the book contents, semi classical transport will not be discussed in any sub section of Chapter 1.

As the final modeling focus in Chapter 1, a sub-section is geared towards keeping the readers abreast on the methods of forming ultra shallow junction of sub 5 nm feature used for sub 10 nm gate lengths to achieve superior gate controlled electrostatics and characteristic penetration depth improving threshold voltage roll-off and DIBL impacted I_{off} leakage current is discussed in

the context of self-assembled dopant monolayers followed by conventional laser spike annealing [117–127]. Using reference of a key paper by Ali Javey et al. on "Controlled Nanoscale Doping of Semiconductors Via Molecular Monolayers" [118], the controlled monolayer doping (MLD) is presented which attains well controlled and uniform formation of nanoscale doping profiles by adjusting the rapid-thermal-anealing (RTA) step. By controlling the spatial variation in number and positioning of the dopants in source/drain ultra shallow junctions and channel by the proposed method in the above paper, contact resistance fluctuations can be minimized where ultra shallow junction generally leads to higher contact resistance due to the shallow thickness of the junction. Different ion implantation strategies and dopant activation schemes like laser thermal annealing, plasma-immersion ion implantation, scanning tunneling microscopy (STM), and single-ion implantation (SII) techniques. Other methods for superior ion dose and straggle control of channel dopants is discussed using high mass molecular (HMM) beam sources for dopant implantations and using a combination of direct write focused ion beams and solid state ion detectors to deterministically fabricate single atom devices and also for controlling dopant fluctuation induced threshold voltage spread within a die and fro[illegible] scaling length optimization by enhancing gate-to-channel electrostatics employing ALD controlled high-k dielectric along with channel doping optimization technique such as super steep retrograde doping and reducing drain junction field encroachment towards the source by drain junction profile engineering are brought to reader's attention for device performance improvements. The readers here are made aware of the novel versatile technology that are keeping the device sustain scaling and maintain required performance. Due to short space of the book, contents on channel engineering and ultra shallow junction formation for improved MOSFET performance will not be discussed as a sub-section for Chapter 1.

Chapter 2 will focus upon device physical aspects of high mobility III-V compound semiconductors. As illustrated by Jesús A. del Alamo in a classic paper "Nanometer-Scale Electronics With Iii-V Compound Semiconductors," III-V compound semiconductors such as GaAs, AlAs, InAs, InP, and their ternary and quaternary alloys find applications in high-speed logic and high-frequency RF and analog electronic systems and other optoelectronic systems such as lasers, light-emitting diodes, and detectors for optical communications, instrumentation, and sensing [128–143]. Another classic paper by del Alamo et al. titled "Nanometer-Scale III-V MOSFETs" [129], featuring important n-type InGaAs FET and p-channel InGaSb FET for optimized III-V CMOS performance superior than all Si CMOS devices, is referred to the readers with documentation of essential physics behind enhanced performance.

The readers are referred to the classic paper providing a survey of ohmic contact materials and properties to GaAs, InP and GaN will be presented along with critical issues pertaining to each semiconductor material. In-line with this discussion, another paper titled "Silicon CMOS Ohmic Contact Technology for Contacting III-V Compound Materials," underlying contact

physics of making ohmic contacts to buried III-V films using Silicide metallurgies is identified and narrated. Novel contact engineering by NiSi/Si/III-V dual heterojunction contact structures using p-Si/GaAs interface and GaAs/$Al_xGa_{1-x}As$ laser with NiSi top contact is reported in this paper. Additional discussion is centered around relationship between specific contact resistivity and heterojunction barrier width. Other experimental demonstrations of ultra low resistance contacts to n-type III-V materials will be further discussed in this subsection.

The enthusiastic readers are directed to references from above on high-field transport in wide-band gap semiconductors. High-field transport properties and estimated saturation drift velocities in several wide band gap semiconductors based upon drifted Maxwellian distribution as discussed by D. K. Ferry in the paper entitled "High Field Transport In Wide Band Gap Semiconductors" will be further discussed. Since nanoscale InGaAs MOSFETs exhibit ballistic transport at the operating longitudinal drift field, ballistic resistance becomes comparable to external source drain resistance R_{sd} and needs to be analytically modeled. Using the paper by del Alamo et al. where the authors documented about the MIT virtual source model (MVS-2) that has been developed to analyze the behavior of nanometer scaled self-aligned InGaAs that exhibits a very small external resistance, a case where the ballistic resistance is of particular importance. Another highly relevant paper on the effect of source starvation in high transconductance III-V quantum well MOSFETs in another paper by del Alamo et al. has been investigated with underlying device physics. Theoretical study of transport property of InAsSb quantum well heterostructure to extract electron mobility in presence of 2D degeneracy induced band nonparabolicity and strain with major scattering mechanisms by acoustic phonon, polar optical phonon and alloy disorder have been included in a paper by Yiping Zeng et al. This paper makes special contribution to incorporation of dislocation scattering, intra-sub band scattering and inter subband scattering usually found to be neglected and deemed inconsequential. The readers are directed to the device physical study on this paper further discussing impacts of quantum well width, spacer thickness, barrier thickness and remote doping concentration for AlSb/InAlSb heterostructures. Focusing high electron mobility in AlGaN/GaN heterostructure, device physics based study is centered on a paper where the authors have illustrated by the molar fractional compositional doping of Al in AlInGaN/GaN heterostructures, condition for the highest thermodynamic stability owing to the significant reduction in total elastic strain in the barrier can be identified that leads to the highest experimentally extracted electron mobility. In graded heterojunction devices such as AlGaN alloys, electron scattering rates are different in different parts of the channel depending upon the local alloy composition and therefore certain classic papers by authors specifically treat alloy scattering and piezoelectric scattering in Wurtzite structures such as GaN and are also referred to interested readers wanting to probe further. The notion that alloy scattering becomes increasingly significant for lightly doped semiconductors at low temperatures is further analyzed. Apart from alloy scattering, in a paper entitled "Neutral Impurity Scattering in AlGaAs" the authors have found that for larger electron ratios and doping levels on

AU: Please add this Ref. to end of Bibliography.

the order of $10^{18}/cm^3$, neutral impurity scattering proved to be the dominant scattering mechanism at all substrate temperatures. Employing strain and heterostructure design (such as valence band offset), splitting between the light and heavy hole bands, K. C. Saraswat et al. discussed in the enhancement of the hole mobility in III-V p-MOSFETs a crucial requirement for best optimized CMOS logic switches in CMOS devices. A fundamental obstacle towards the realization of GaN p-channel transistors is its low hole mobility. In a recent paper entitled “Route to High Hole Mobility in GaN via Reversal of Crystal-Field Splitting” [143], the authors have demonstrated that the hole mobility can be significantly enhanced if the split off hole band can be raised above the light hole and heavy hole bands by reversing the sign of the crystal-field splitting via biaxial tensile strain which will be further discussed in this subsection. Tailing of band edge states (Urbach tails) and band gap narrowing effect affecting the subthreshold slope in III-V TFETs and optical frequency in III-V LEDs and solar cells have been discussed with their underlying physics in connection to a recent paper by an author group. The authors are also directed to references on Ge p-type FETs with integration in III-V n-type FETs to boost hole mobility required for optimized CMOS performance. Several high-quality papers relevant to this topic with underlying device physical concepts along with technology development of germanium n and p type MOSFETs are analyzed in this subsection. A difficult problem with III-V MOSFETs results from the unpassivated dangling bonds on III-V free surfaces. The energy locations of the dangling bonds are directly related to the bulk band structures of different III-V semiconductors. Using charge neutrality level model (CNL), Peide Ye at al. demonstrated by using device physics the experimental observations on III-V MOSFET drive current improvement by ALD Ga_2O_3 (Gd_2O_3) gate dielectric. For this discussion, the author of the above article using simplified band diagram showed that the narrower separation between CNL and conduction band minima (CBM) of the parabolic Dit in In-rich InGaAs compared to pure GaAs was ultimately responsible for higher inversion drive current in InGaAs using Al_2O_3 as gate dielectrics but this experimental finding can be extended for all other gate dielectrics including Ga_2O_3 (Gd_2O_3). The author series found some important supporting observations to explain the functional device physical aspects of this band diagram as discussed by Peide Ye and additionally confirm the In-rich current boost of $In_xGa_{1-x}As$ MOSFETs a property significantly arising out of Dit profile shape and its occupancy of charge carriers between CNL and CBM energy levels resulting in threshold voltage shift which is minimum for the highest In content. Other Dit passivation techniques to improve high-k heterojunction interface of III-V HEMTs are referred to the readers from references covered for Chapter 2.

Bulk CMOS technology has become the dominant technology for very large scale integration (VLSI) circuits for manufacturing processes that were prevalent 10–15 years back but for today’s emergent giga-scale integration (GSI) of circuits in chips, bulk CMOS technology is facing severe limitations in optimal performance due to extreme short channel effects, gate-to-channel potential integrity compromising effects and 2D quantization effects due to extremely

scaled channel length and gate width. To scale bulk MOSFETs further into the nanometer regime without being vulnerable to excessive short channel effects (SCE), both the front gate dielectric and the depletion width in silicon must be reduced in proportion to the channel length. Narrowing down the depletion width in silicon requires simultaneously the shallow junction technology and an increase of substrate doping concentration, the latter in turn creates a high field region in the vicinity of drain to substrate junction. In such an elevated high field region, quantum mechanical (QM) band-to-ban tunneling (BTBT) takes place which causes a considerable junction leakage current in the form of gate-induced-drain-leakage (GIDL) current when the gate bias is 0 V or negative and the device is in switched off mode that eventually results in significant stand by power consumption for a CMOS logic switch based gigantic network systems. Therefore, Chapter 3 of this book comprehensively enunciates the structural and device parametric modeling and device physics based aspects of silicon-on-insulator MOSFETs [144–153] and its derivatives as potent alternatives to conventional bulk MOSFETs suited substrates made of Si, Ge and III-V materials. In contrast to bulk MOSFETs, the thin silicon film of a silicon-on-insulator (SOI) MOSFET is electrically isolated from the underlying substrate by a thick buried oxide (BOX) grown layer. Therefore, an isolation is created from source and drain side junction from the silicon film to the substrate contact resulting in near absent junction capacitance, reduced source/drain junction surface periphery and lower leakage current as the silicon film thickness is thin enough to cut-off sub-surface flow of subthreshold current and allowing proximity of source and drain junction as per scaling requirement without allowing adverse penetration of drain field towards the source and offsetting gate-to-source barrier control. Other benefits of SOI MOSFETs are increased immunity to radiation induced photocurrents with associated single event upset (SEU) phenomenon and usually prominent latch up effects a burning issue of back-end-of-line (BEOL) faults located for manufactured CMOS devices. Additionally, the presence of buried layer isolating the thin silicon film from the substrate reduces the sensitivity to body effect induced by substrate contact leading to more stringent threshold voltage control. The reduction of junction capacitance and sub-surface leakage current flow path enables the SOI-based modern CMOS circuits providing higher intrinsic speed and lower power consumption. Additional impact on drive current, threshold voltage and channel mobility can be engineered by back gate bias in accumulation, depletion or inversion mode. Wide-band gap material is the next driving technology as alternatives to silicon and germanium and in these short overview of different subsections analyzing the device physical and performance aspects of various wide band gap semiconductors cannot be discussed issue by issue because of space content of the book but only brief discussion will follow.

The readers are further directed to discussion from the above references on SOI based MOSFETs and their derivatives will first elucidate the important scaling theory of silicon film, front gate insulator and buried oxide (BOX) thickness in short channel ultrathin SOI MOSFETs that generate the optimal 2D field contour in the silicon film essential to combat short channel ef-

fects. Thinning down the BOX thickness aggressively is expected to lead toward double gate (DG) structures on SOI substrate. The scaling limits of bulk and DG MOSFETs are not applicable to ultrathin body SOI MOSFETs because the BOX is sufficiently thick to define an enclosed region with well specified boundary conditions in solving the 2D boundary value problem for Poisson's equation. The classic paper by Yuan Taur et al. on ultrathin SOI MOSFETs scaling limit will be analyzed in the first sub-section of this chapter, particularly the general scaling principle presented in the paper by Yuan Taur et al. taking assistance of extensive 2D numerical simulation based design study in their paper. The context of the impact of device parameters optimization on SCE, e.g., high-k dielectrics, channel doping, thickness of silicon film and front gate insulator is addressed in this Chapter through reference to this paper in view of practical considerations. The important undesirable threshold voltage shift due to quantum effect arising out of excessively thinned silicon film in ultrathin SOI MOSFETs is further assessed from the scaling limit stipulation.

The readers are directed from references above discussing mainstream SOI MOSFETs, detailed performance attributes of partially depleted (PD) SOI MOSFETs, fully depleted (FD) SOI MOSFETs, ultra thin body (UTB) SOI MOSFETs, extremely thin (ET) SOI MOSFETs, and their associated gate architectures.

The readers are further directed from these references on research outcome from potential self-heating effects which is unavoidable in ultrathin SOI MOSFET with thin silicon film and buried oxide (BOX) layer. The self-heating effect (SHE) is directly proportional to the thermal resistance built up in the nm dimensioned silicon film affecting device transconductance (g_m), channel mobility degradation, maximum saturated drift velocity, and drive current reduction. Self-heating dominates in ultrathin SOI MOSFETs due to the increased current density out of scaling and progressively reduced silicon volume available for heat removal. The latter effect is due to the significantly smaller thermal conductivity of silicon dioxide of BOX layer compared to that of bulk silicon at room temperature: κ (SiO_2) = 1.40 W/m-°K while κ (Si) = 148 W/m-°K. The classic paper by Guido Groeseneken et al. providing experimental validation of self-heating simulations and projections for transistors in deeply scaled node. In an important paper by Jea-Gun Park et al., the authors showed that when the ultrathin body SOI film thickness is less than 3 nm, the apparent increase of mobility due to quantum mechanically induced greater electron occupancy in a two-fold valley is actually neutralized when proper self-heating in thin silicon film is considered. The impact of self-heating on phonon limited electron mobility as a function of silicon film thickness as analyzed in this paper is referred to the readers. In another important paper by A. Cresti et al.. the authors performed a full quantum study to extract the increase of self-heating effects in nano SOI devices induced by surface roughness. Nanotransistors fabricated out of 3D fins structures and overlapping nanometer architectures are more sensitive to spatial fluctuation induced by interface imperfections such as surface roughness present at the semiconductor film and buried oxide interfaces. Surface roughness can naturally result in deterioration of the electrical performance due to

carrier back scattering, especially in the strong inversion region of transport when the electron wave function is pushed towards the interfaces by the electrostatic attraction of the gate voltage-unlike volume inversion where the intrinsic doping of the Si film counters this effect. Another possible drawback of surface roughness present at different gate-to-silicon channel and silicon film-BOX boundary interfaces is the enhanced self-heating caused by the surrounding oxide layers to provide optimal capacitance coupling with the gate voltages. Unlike bulk silicon film which has a relatively high thermal conductivity of 149 W/m-°K, both experiments and theoretical calculations have reported that this quantity is strongly reduced in very thin layers by quantum confinement and heat dissipation impediment by the underlying BOX layer and also by the spatial fluctuations such as roughness at the interface adding to phonon limited various boundary layer scattering components. The impact of surface roughness on self-heating effects for a tri-gate nanowire (NW) Si FET is studied by the authors of this paper and their findings can be transferred to similar understanding of tri-gate NW SOI MOSFETs. The readers are directed to the references which includes an important article "Analysis Of Self-Heating Effects In Ultrathin Body SOI MOSFETs by Device Simulation" by the authors Anthony G. O' Neill et al. In their paper the authors through 2D electrothermal device simulation with a transport model tuned to the results of Monte Carlo (MC) transport analysis, systematically discussed the effects of heating on the I–V characteristics of ultrashort MOSFETs and to analyze the impact of available technology options on the self-heating effect (SHE) occurring in the UTB SOI MOSFETs, taking into consideration both digital and analog circuit applications. The authors reported particularly significant finding of degradation of backscattering coefficient as the device temperature in silicon film is increased due to SHE when ballistic transport for 25 nm UTB SOI MOSFET is considered. The author after analysis of these various papers where self-heating, a reliability phenomenon is addressed, considers lowering the substrate temperature operation may align the carrier temperature closer to device temperature in operation when the gate and drain bias voltage are properly scaled to a lower value and this observation will be also highlighted in this chapter.

The readers are further directed to promising steep switching devices that are discussed in the above references. A band modulation device in advanced FDSOI technology sharp switching Z^2-FETdevice characteristics is studied by Sorin Cristoloveanu et al. in a recent paper. In this paper the authors demonstrate that integrated with 28 nm FDSOI technology, the Z^2-FET (Zero impact ionization and zero subthreshold slope) is equipped with excellent properties as sharp switch and tunable triggering voltage. Other sharp switching device that employs positive feedback similar to field effect diode (FED) and feedback FET (FB-FET) is discussed by S. Cristoloveanu and A. Zaslavsky et al. These two papers provide important contributions on steep switching devices implemented in buried oxide technology.

The readers are directed to reference article "Multiple Gate SOI MOSFETs and their Device Design Guidelines" is authored by J. P. Collinge et al. and is analyzed in this subsection. In this

article, the concept of a triple gate device with sidewalls extending into the buried oxide with the name Pi-gate (Π) MOSFET is introduced. The Pi-gate device is simple to manufacture and offers electrical characteristics similar to harder to fabricate gate all round MOSFET. In this paper the authors provide important illustration comparing four gate architectures (double gate, triple gate, quadruple gate, and newly conceived Pi-gate) built in SOI technology to determine the efficiency of different gate structures impacted by four variable device parameters, e.g., gate length, channel width, doping concentration and silicon film thickness. Due to shortage of space, the readers are [illegible]ugh reference articles that could not be discussed in brief in Chapter 3.

[illegible] the continued downscaling of conventional MOSFETs, the constant field scaling greatly reduces the gate and drain supply voltage and shifts the threshold voltage (V_{th}) upward due to increasing channel dopant requirement to prevent source and drain punch through and also to reduce subthreshold leakage current exponentially dependent on threshold voltage reduction. As a result of this V_{th} increase, the gate voltage overdrive (V_{gs}-V_{th}) is decreasing at a faster rate for every scaled technology node which limits the maximum attainable drain current for digital logic based circuits applications. One ingenious alternative to maintain the drive current improvement in commensurate with ITRS roadmap even when scaling forces minute gate overdrive is enhancement of n-channel mobility for n-MOSFET and p-channel mobility for p-MOSFET paving the path for strained n-channel and p-channel MOSFET with leverages on SOI, III-V architectures with different gate structures, e.g., single gate, double gate, tri-gate, and FinFET. Chapter 4 therefore initiates the overall strained Si outlook for n-channel and p-channel MOSFETs by directing the readers to the references [154–174], such as "Mobility Enhancement-the Next Vector to Extend Moore's Law" by Scott E. Thompson et al. [159], where the authors discussed the technology option that starting with the 90 nm technology generation, mobility enhancement through uni-axial process induced strained Si has emerged as the next scaling vector being widely adopted in logic devices. The above referenced article by the authors is targeted to abreast the readers with an introduction to the physics of strained silicon and current state of the art in uniaxial strained Si MOSFET. The first part of the article explains how strain alters the valence and conduction band of Si as well as scattering rates. The important review useful for the readers' knowledge is then centered upon state-of-the-art strained techniques being implemented in 90 nm and 65 nm process technologies. The authors of this first referenced paper also provide important concluding remarks on the future scalability of strained Si MOSFETs in the ballistic regime and nanoscale CMOS.

Chapter 4 is related to the classic book "Strain Effect in Semiconductors–Theory and Device Applications" by three authors Yongke Sun, Scott E. Thompson, and Toshikazu Nishida that the readers are referred to peruse. Some of the expositional excerpts from the introductory part of this second referenced book are restated here for the readers' attention. Following the promising $Si/Si_{1-x}Ge_x$ heteroepitaxy results, wafer-based substrate strain was experimentally and theoretically studied by a large number of researchers. In the 90s, two process induced strain sources were investigated,

high stress capping layers deposited on MOSFETs and embedded SiGe source and drain. The embedded SiGe literature prompted Intel to evaluate the technology, which resulted in larger than expected device performance enhancement which was later attributed to compressive channel stress. However, neither biaxial nor uniaxial stress was immediately adopted in CMOS logic technologies for several reasons. For biaxial stress, issues included defects in the substrate and performance loss at high vertical electric fields. Process induced uniaxial channel stress was not initially adopted since different stress types (compressive and tensile for n and p-channel MOSFET, respectively) were needed. After careful analysis of the hole mobility enhancement at high vertical electric fields and the potential for continued effectiveness at nanoscale dimensions, process induced uniaxial strain was adopted over biaxial stress. Uniaxial stress provided significantly larger hole mobility enhancement (a key requirement for Si CMOS symmetric transfer curve with near equal SNM window with the fact that hole mobility in silicon is almost 3 times lower than electron mobility) at both low strain and high vertical field. Since application of high strain can lead to strain relaxation via defect formation, large mobility enhancement at low strain is critical for yield. Uniaxial stress also provided larger drive current improvement for nanoscale short channel devices since the enhanced electron and hole mobility arise mostly from higher rate of reduction of conductivity effective mass of Si instead of primarily from subtle reduction of scattering rates for biaxial stress. Another important consideration was the strain's effect on the threshold voltage. Process-induced uniaxial stress resulted in an approximately five times smaller n-channel threshold voltage shift out of 2D confinement effects. The smaller threshold voltage shift was manifested in a smaller penalty for overall adjusted threshold voltage shift retargeting the adjustment of the channel doping. Alternatively, the larger threshold voltage shift observed for the substrate induced biaxial tensile stress causes approximately half of the stress enhanced electron mobility to be lost. Based on the merits of uniaxial stress and the necessity for opposite stress types for n and p-channel MOSFETs, two process flows were developed that independently targeted the stress magnitude and direction. The first involved embedded and raised SiGe in the p-channel source and drain and a tensile capping layer on the n-channel device. The second method used dual stress liners: compressive and tensile silicon nitride (SiN) for p- and n-channel devices, respectively. As a feature enhancement for CMOS, process induced stress was employed in nearly all high performance logic technologies at the 90 nm, 65 nm, and 45 nm technology nodes for both microprocessor and consumer products. This reference book will be further analyzed as per its organization of contents where part I elaborates band structures of strained semiconductors, part II discusses how changes in band structure coupled with changes in carrier scattering caused by strain affect carrier transport, part III discusses strain applications on feature enhancement of electronic devices such as Si and SiGe planar and nonplanar MOSFETs.

The readers are also directed to the paper by D.A. Antoniadis and J. L. Hoyt et al. entitled "Continuous MOSFET Performance Increase with Device Scaling: The Role of Strain and

Channel Material Innovations" [174] will be analyzed. In this article the authors developed an analytical model that quantitatively illustrates the relationship among carrier velocity, MOSFET drain current and switching time. From their analysis it is projected that additional improvements in channel velocity and therefore in mobility will be required maintaining compatibility with the commensurate scaling and thus research efforts are directed to improve channel electron and hole mobility using heterostructure of strained Si and strained SiGe on both bulk and on-insulator substrates. The authors then provide insightful analysis of some of the remaining technology challenge associated with these heterostructure MOSFETs. The readers are directed to another classic paper by Roland Stenzel et al. on "Understanding Strain-Induced Drive Current Enhancement in Strained-Silicon n-MOSFET and p-MOSFET." The authors of this paper start the discussion justifying that strain boosts the effective carrier velocity in the channel which compensates for the detrimental impact of parasitic resistances of source and drain contact and access regions and accompanied reduction of overdrive factor due to supply voltage scaling of ultrashort channel devices. Further discussion in this article is centered on strain techniques that are classified as local process induced such as stressed overlayers, embedded source/drain stressors, stress memorization techniques (SMTs) or stressed contact and metal gates. Strain can also be built directly into the substrate and is thus classified as globally strained. The authors provide important observation that drive current of short channel p-MOSFETs in the low lateral field regime corresponding to low drain to source voltage (mostly in the linear or triode region of operation) is enhanced comparatively to a higher value by local strain techniques than that in the high lateral field regime corresponding to high drain to source voltage and mostly saturated region operation of the drive current. In contrast, the n-channel MOSFET shows the opposite trend when drive current is computed by local strain techniques for low lateral field regime and high lateral field regime leading to the ratio of linear to saturation current benefit for n-MOSFET tending to wards unity or even less than unity (this ratio is greater than unity as explained above). The authors rightfully claim that there is a lack of in-depth understanding of the physical factors that contribute to the difference in the transport characteristics between n and p-channel MOSFET when effective strain technology is independently employed. In contrast to reported data that are typically limited to specific saturation and linear current metrics at particular voltage conditions or studies with wafer bending techniques that impart strain independent of the strain technique, the authors in this paper focus on the drive current enhancement response subjected to each strain technique as a function of applied vertical and lateral electric fields with associated modulation of the electronic band structure and interaction with strain induced drive current alternatives. Locally and globally strained n-MOSFETs exhibit similar behavior under changing vertical and lateral electric fields. Whereas strained p-MOSFETs show a different sensitivity resulting from the differing electronic band structure modulation where band warpage and distortion are disproportionate caused by different strain processes included. The paper is further analyzed from the viewpoint of the relationship between carrier mobility

enhancement and drain current enhancement and the physics behind the field dependence of the drain current enhancements under strain from overlayer stressors and a comparative analysis of the induced stress patterns resulting from the various strain techniques and their influence on the drive current enhancements. In-line with the generic analysis of the application of strain technology, another classic paper by S. Takagi on "Strained-Si CMOS Technology" [171] is referred to the readers for perusal. In this review paper by the author, the author insightfully conjectures that simple device scaling encounters a trade-off relationship among the current drive, power consumption and the short channel effects that prompted the device technologies to adopt new channel structures and materials that are collectively termed "Technology Boosters" in ITRS 2004 edition and include strained Si channels, ultrathin SOI, metal gate electrode, multigate structures, ballistic transport channels, metal source/drain junctions and so on. By using strained Si channels by local strain techniques, the mobility enhancement ratio can be obtained as a percentage of applied strain resulting in higher current drive under scaled supply voltage and slightly relaxed scaling of gate oxide thickness. The author in this article provides insightful reviews on the principle and the device application of this strained Si CMOS technology with the emphasis on the physical mechanism of mobility enhancement due to strain. Next a series of classic papers by Scott E. Thompson et al. are analyzed that blended effective use of strain in n-channel and p-channel MOSFET from industry batch manufacturing perspective.

AU: This sentence needs a verb

The author takes references from important articles that outscore research effort on hole mobility improvement by different strain technology. In the first of this paper by Dimitri A. Antoniadis et al., by growth of: (i) single strained Si channel on relaxed Si_{1-x} Ge_x substrate; (ii) dual channel heterostructures on Si_{1-x} Ge_x substrate; and (iii) thin strained Si on relaxed Si_{1-x} Ge_x and digital alloy channels, electron mobility enhancement as high as 1.8 and hole mobility enhancements over 8 have been achieved. The ability of the hole to hybridize different band structures in nanometer scale layers explains the unexpected improvements in hole mobility. Dual channel heterostructures display large hole mobility enhancements and hole mobility increases with buried channel Ge content. The authors of this referenced paper also reported symmetric mobility in n- and p-MOSFETs on a single substrate the ultimate dual channel. Low temperature operation is an effective means to boost low vertical field phonon related mobility and the effective hole mobility at low temperatures can provide the important boost in mobility values both for n- and p-MOSFETs device technologists are looking for. The author in this context of low temperature strain technology further takes references from one article by Katsuyoshi Washio et al. titled "Low-Temperature Electrical Characteristics of Strained-Si MOSFETs" and a thesis by Xiaodong Yang titled "Strain Effects on the Performance of Silicon MOSFETs." In the article by Katsuyoshi Washio et al., the Figure 5 of their paper shows electron mobility between 2,000 cm^2/V-s and 1,000 cm^2/V-s from low vertical electric field to high vertical field at the operating temperature of 83 K for a strained Si n-MOSFET which is almost twice for the device unstrained at T = 83 K. From Figure 7 of their article

the experimental finding is hole mobility for p-MOSFETs at T = 83 K when the device is strained reaches a peak value of 150 cm^2/V-s at a vertical electric field nearing 0.8 MV/cm and almost 1.5 times enhancement is noted for the same device at these operating conditions but unstrained. In the thesis by Xiaodong Yang, his research theme centered around application of controlled external uniaxial mechanical stress to n- and p-MOSFETs to measure the strain-altered electron and hole mobility versus temperature and he used k•p method and a modified surface roughness scattering to simulate the low temperature mobility as low as 77 K. From Figure 4-2 of this thesis, the measured hole mobility in p-channel MOSFETs can peak as high as 500 cm^2/V-s for 60 MPa longitudinal compressive stress for (100) Si substrate along <110> transport direction at a temperature of T = 100 K. From Figure 4-4 the 60 MPa uniaxially compressive stress results in peak phonon limited mobility around 900 cm^2/V-s for the above device in the specified transport direction at a temperature of 100 K. Compared to large increase of phonon limited mobility at low temperatures surface roughness (SR) scattering is the dominant process at high enough vertical electric field and SR limited mobility for 60 MPa uniaxial longitudinal stress gradually increases for T < 300 K and around T = 100 K, the SR limited mobility is around 1,000 cm^2/V-s. Further information will be discussed from this thesis about findings on n-channel mobility enhancement at low temperatures where the benefit is certainly more visible compared to p-channel mobility enhancement. Next, in a classic paper by J. C. Sturm et al. titled "Alloy Scattering Limited Transport of Two Dimensional Carriers in Strained Si_{1-X} Ge_X Quantum Wells", the authors investigated the electron and hole mobility of 2D systems in strained Si_{1-x} Ge_x quantum wells at low temperatures as a function of Ge content in the well. Alloy scattering with V_{alloy} = 0.8 eV for electron and V_{alloy} = 0.6 eV for hole reduces both electron and hole mobility at 10 K from their maximum peak value in the vicinity of 10,000 cm^2/V-s from the Figure 4 of this paper by the authors. In another paper by P. K. Bhattacharya et al. titled "Carrier Velocity Field Characteristics and Alloy Scattering Potential in Si_{1-x} Ge_x/Si," the authors provide experimental measurements of hole mobility as a function of Ge content up to 100% for relaxed and coherently strained p-Si_{1-x} Ge_x alloys. At a field of 8,000 V/cm with Ge mole fraction 1, hole velocity in relaxed Si_{1-x} Ge_x reaches a value of 3 × 10^6 cm/s whereas for n-Si1-x Gex alloy with Ge mole fraction 0.33, electron velocity reaches 5 × 10^6 cm/s at a field of 3,000 V/cm. Final paper that sheds light on improving hole mobility in strained-Si channels is by the authors T. E. Whall et al., where using Si/$Si_{0.5}$ $Ge_{0.5}$/Si p-MOSFET the authors showed through physical characterization and analysis of the electrical properties that at a total sheet hole carrier concentration of 10^{12}/cm^2, the 300 K mobility is 220 cm^2/V-s being double that of a Si control device and at 4 K, a peak hole mobility value of 1,800 cm^2/V-s is reported that is extended up to a sheet inversion hole carrier density of 10^{13}/cm^2. Contrary to the previous paper, the hole mobility in this experiment is found to be affected by interface roughness scattering rather than alloy scattering.

The readers are also referred in Chapter 4 a classic comprehensive paper by Scott E. Thompson et al. on the physics of strain effects in bulk semiconductors and surface Si, Ge and III-V channel MOSFETs which are presented. To supplement the strain effects on electrons as carriers in these devices, another paper on physics of hole transport in strained silicon MOSFET inversion layers by Martin D. Giles et al. is discussed.

A perfect switch has zero current flow when it is open and zero resistance when it is closed and the switch is capable of switching instantly from "off" to "on" state and vice versa. MOSFETs are, unfortunately, imperfect switches. The off state current is not ideally zero but limited by the minimum value of unceasing subthreshold leakage current and on the other hand, the on state current is restricted by residual on-state resistance and sub-proportionate current ceiling as the lateral drift field increase does not result in proportionate drift velocity increase but the velocity saturates beyond a particular lateral field and this on-current-drift field characteristics is more conspicuous as the channel dimension gets shrunk to promote packing density of transistors per chip and economize chip price. Moreover, since the subthreshold charge in the channel region has to be removed from the channel from the source contact, switching from "on" to "off" state shows a gradual degree of latency for conventional MOSFET implying some small time is needed from on-to-off state switching. In order to expedite fast and smooth switching behavior of MOSFETs as transistors are being shrunk in dimensions, 2D and 3D gate architectures are proposed as alternatives to planar configuration where instead of single top-gate electrode, the gate electrode now wraps around several sides of the conducting channel taking channel height into account and as a result of this, electrostatic control over the channel or gate to channel integrity is improved. Such multigate architectures will allow a further shrinking in transistor size without facing downgrade of transistor performance. In Chapter 5 of this book, the author therefore starts the discussion on multigate transistors providing superior switching and transport performance compared to scaled planar MOSFETs [175–193] by first referencing an excellent review article on this topic titled "Multi-gate Transistors as the Future Of Classical Metal-Oxide-Semiconductor Field-Effect Transistors" [175] by Jean-Pierre Colinge et al. The authors in their review starts the analysis on defining the subthreshold slope of planar MOSFET where the injection is by thermionic emission over source to channel barrier by first describing $SS = n\,{}^{kT}/_{q}\ln(10)$ where "n" the body factor represents the efficiency or rather, the inefficiency with which the gate voltage electrostatically controls the channel region. The body factor, as the authors described in this review with an expression different that derivation quoted in textbooks, is proportional to the change in gate voltage with a change in channel potential, that is $n = {}^{dV_G}/_{d\varphi_{ch}}$. For excessively scaled MOSFET with close proximity between source and drain and comparatively larger voltage on the drain contact, mid region to drain region channel potential drop is more asymmetrically steeper compared to source region to mid region channel potential rise (both referenced from the middle of the channel position) resulting in 2D channel potential across the length of the channel and depth of the channel being highly asymmetric for a

planar MOSFET. Thus, a non-linear high sensitivity of the channel potential at spatially different locations between source and drain is found to exist for small gate voltage changes making the n factor worse and significantly above than unity which consequently makes SS very much above than ideal thermodynamic limit at T = 300 K or 60 mV/decade a direct consequence of aggressive channel length scaling. The authors of this review article thus comments that in the best possible case, if the electrostatic coupling between the gate and channel region remains at 100%, then n becomes almost 1 and subthreshold slope is then very much the ideal 60 mV/decade at T = 300 K. Multistacked gate structure imposes control on the channel through the gate to channel vertical field that is now more systematically applied through all sides of the channel or channel volume minimizing the asperity in the spatial variation of the 2D channel potential with additional effect of considerable reduction of the impact of the drain to channel lateral 2D field penetration leading to improvement of the scaling length parameter which is the metric gauging how fast the channel can be turned on and off and how the short channel effects (SCE) can be controlled by optimized gate structures. The authors then in their review article discusses various multigate structures with process compatibility similar to planar MOSFETs and the previous technology boosting MOSFET devices discussed in this book from Chapter 1–4 can be seamlessly applied to multigate structures preserving the beneficial transport characteristics improvement. The authors further state an important parameter "natural length λ" of a MOSFET which represents the extension of the electric field lines from the source and the drain into the channel region being modulated by the applied gate and drain voltages. This is an important parameter which will be further scrutinized and analyzed by the author from the review analysis of this first referenced article. The authors show, in their review paper in Figure 8, that when drain-induced-barrier-lowering (DIBL) effect in mV/V is measured or plotted against normalized channel length L/λ_N where λN is the natural length parameter for different number of gates (N), DIBL shows a sharp rise when L/λ_N value is lowered than 6 with important conclusion that for lowest DIBL or short channel effects, the multigate structures show the benefit with increasing number of gates compared to planar MOSFETs to effectively keep the L/λ_N in the vicinity of 6. The authors also projected in the perspectives section of their paper that the smallest silicon transistor ever reported with ab initio simulations is the use of multigate architectures extending Moore's Law down to 3 nm node.

The readers are referred to the natural scaling length λ an important indicator about how effectively short channel effects (SCE) can be controlled by muligate number based MOSFET devices will be explored by referencing a series of excellent articles presented in Chapter 5. Some of these papers discuss on the natural length of a multigate structure with a square cross section was found equal to a double gate device λ value reduced by a factor of √2. The authors in this paper rightly opines that unlike DG MOSFET, the natural length for the multigate (MG) FET which has a 3D structure cannot be directly obtained from the solution of 2D Poisson's equation. In this paper, the author proposes a simple scaling theory for the MG-MOSFET based on the concept

of effective number of gates (ENGs). By avoidance of modified 2D Poisson's equation required for precise computation of natural length parameter, the author extracts the ENG equation based on the perimeter-weighted-sum (PWS) method. The authors in this paper illustrated that through this simple and feasible formula, potent SCE effect on DIBL can be analyzed for MG-MOSFET. The readers are also directed to a paper by James D. Plummer et al., "Cylindrical Fully Depleted Surrounding Gate MOSFETs" exert greater gate-to-channel integrity providing tighter packing density and the authors developed scaling theory for this device structure. Other reference articles such as, the natural scaling length of Tri-gate MOSFET by the authors Tsu-Jae-King Liu et al., where the authors used the exact 3D Poisson's equation with parabolic potential variation. For the given value of gate width W and Si thickness t_{si}, the authors through their analytical derivations showed that the scale length is smallest for trigate bulk MOSFET design because it combines the benefit of a multigate structure with the benefit of a ground plane structure (used for leakage suppression). The trigate bulk MOSFET as verified by simulation results reported in this paper has the least amount of drain-induced-barrier lowering and therefore is most scalable.

In the manufacturing perspective, quasi-planar fin field effect transistors (FinFETs) appears attractive due to the higher on-current, while double gate FinFET offers some advantages through less pronounced corner effects [191–193]. In-line with previous discussion of scaling theory of multigate MOSFET structures in this Chapter 5, scaling thory for FinFEts based on 3D effects or the effect of fin-height is analyzed by reference of the paper by authors Liliu Tian et al. The authors start the discussion by noting that since the vertical fin (channel) introduces additional parameter, the analysis of FinFETs electrical characteristics should include the 3D effects. The authors then continued that due to the difficulties in conducting 3D simulation and development of analytical model, the research on FinFETs currently focuses on the process simplicity and compatibility with the planar CMOS technology. The paper's significant contribution as per authors' claim is that the scaling theory for FinFETs is still not available in literatures due to the complexity to study the 3D effects analytically. In order to solve the 3D Poisson's equation in the channel region, the authors employ the method of minimum channel potential derived from the evanescent mode and a function expression of subthreshold swing is proposed. The extracted scaling length could be optimized for the structural optimization on fin height over silicon thickness (fin width). In section 2 of this very important paper, analytical solution of 3D Poisson's equation in the channel region is derived. For 3D Poisson's equation formulation, the bias regime is in the subthreshold with the channel is considered fully depleted allowing the simplification of the Poisson's equation without the free carriers density. The choice of superposition method and evanescent mode analysis by the authors originate from the observation that although some of previous scaling theory derivation considered parabolic channel potential, it is impractical for FinFETs because of the coupling between the surrounding gates. The author will provide more elaborate description on the associated boundary interface conditions definition and simplifications that the authors in this reference article utilized.

In the discussion section of this paper by using the analytically developed equations for channel potential ψ(x,y,z) with the lowest order mode and further approximation, the minimum channel potential $\psi_{min}(x,y)$ and the lowest order scaling length λ_{11} are constituted which shows that in short channel, this minimum potential $\psi_{min}(x,y)$ at the source or virtual cathode is impacted by $exp\,\frac{-L_{eff}}{2\lambda_{11}}$ where $\psi_{min}(x,y)$ is lowered at the source end by the DIBL effect and additionally by $exp\,\frac{-L_{eff}}{2\lambda_{11}}$ noting that $exp\,\frac{-L_{eff}}{2\lambda_{11}}$ is <1 for $\lambda_{11} < L_{eff}$. For $\frac{-L_{eff}}{\lambda_{11}} < 1$, there exists severe SCEs, including high output conductance and threshold voltage roll off. The authors in this paper then develops a general subthreshold swing model S for FinFET. From 3D simulations the authors explain the subthreshold swing as a function silicon film doping density. From the citation of general S parameter, the authors develop a modified and simpler expression of S for design considerations after some manipulations of parameters in general expression S and a lower limit of subthreshold swing is thus derived. Then the authors provide three illustrative plots of: (i) analaytically derived subthreshold parameter compared with experimental data; (ii) with 3D simulation; and (iii) with different fin height as a function scale factor $\alpha = \frac{-L_{eff}}{2\lambda_{11}}$. Important observation is also found from the simulation results that the thinner the fin structure comprising triple gate provides distinctive advantage over double gate in decreasing the scaling length for low fin height values of H_{fin} . From the authors' simulations it is also evident that as long as the geometrical parameters (i.e., H_{fin} and T_{fin}) are the same, the scaling length of triple-gate structure is always less than that of double gate structure keeping the function scale factor α at an optimized value not affected by the scaled decrease of L_{eff}. The authors have also studied the optimum gate oxide thickness as a function of fin thickness at fixed scale ratio but with different L_{eff} and a lowest value of 3 nm gate oxide thickness is found for L_{eff} = 20 nm, H_{fin} = 50 nm and t_{fin} = 5 nm. The high-k dielectric constant effect on scaling length is compared for DGMOSFET and FinFET as the final illustrative outcome. Use of higher-k dielectric materials significantly benefits FinFET structure reducing the scaling length twofold with reduced fin thickness T_{fin} = 5 nm, H_{fin} = 50 nm and $t_{ox1} = t_{ox2}$ = 2 nm. Fringing field induced gate-to-source barrier loweing (FIBL), an effect that gets accentuated for higher k gate dielectrics which may impact the scaling length by coupling with DIBL was not considered in this paper by the authors. The author finds the possible explanation for FIBL exclusion lying with the extremely reduced oxide thickness used for study (2 nm where FIBL effect can be neglected).

In addition to providing conduction channels that will otherwise bring conventional FETs to the uttermost scaling limits, the authors find that the physics of 1D group IV nanowire endows new device principles. The review is based upon systematic recapitulation and analysis of the physics of silicon and germanium nanowires and CMOS like devices built from Si and Ge nanowires. A phenomenon quantum confinement present in 1D nanowire will be discussed from this paper. Confinement in ultra thin nanowires leads to a strong alteration of the band structure in the transport

direction in-plane and out-of-plane modifying band gaps, effective electron and hole masses and effective channel mobilities due to these. Additionally, 1D nanowire channel geometry for thinner nanowire radius or pore structure imposes challenge on electron transport mostly occurring in close proximity to the interface and thus affected by charged surfaces and surface roughness scattering. The nanometer geometry can create unique features of longitudinal and radial heterointerfaces by its enhanced lattice relaxation properties. Under electronic transport in silicon and germanium nanowires, the authors discuss: (i) nanowire mobility in the drift-diffusion regime; (ii) quantum transport in nanowire systems; (iii) ballistic nanowire MOSFET; and (iv) quantum capacitance. Under conventional silicon and Ge nanowire FETs, (i) inversion and accumulation mode nanowire FETs will be discussed. Under unconventional nanowire transistors, 1D quantum well and high mobility nanowire FETs are referred to the readers for perusal.. The ballistic nanowire MOSFET discussion in relation to important papers authored by Kenji Natori are also referred to the readers. [illegible] progress in the computing and information technologies over the past four decades has been the enabler of a countless number of applications involving largest data transmission over a much larger bandwidth permitting gazillions of FLOPs (floating-point operations). In a foreseeable future, nanoelectronics will deliver self-powered, energy saving autonomous families of sensing, computing and communicating devices delivering ultra-low power consumption and steepest slope based switching amenable to today's peta to exa Hz switching frequencies of processors integrated with these devices in either digital signal processor format or ASIC based mixed signal format for many scenarios in the frameworks of the internet-of-things and internet-of-humans. Naturally as stated, power consumption during off-state or subthreshold state or during near-threshold computing scenarios is the main hindrance to the progress of the computing technologies laden with protracted idle states in operational cycles for a certain part of core processors circuit blocks.

Indeed, integrated circuits simply do not have the energy budget necessary for the full exploitation of their potential performance. This utilization hindrance led to the so called dark silicon age where at any point of time, optimization and logic scheduling involving massive parallelism and multilevel deeper pipelining warrant that a significant fractions of the gates available on a chip are idle or significantly underclocked. The origin of the utilization bottleneck is that, in the CMOS technology generations after about the beginning of the years 2000, it has been impossible to scale the power supply voltage V_{DD} as propounded by the Dennardian scaling theory so that V_{DD} scaling was discontinued from its plateaued value of about 1 V to contain the exponential power consumption growth due to subthreshold current increase. The dark silicon age officially marks the transition from the Dennardian scaling, where the progress in CMOS technologies was measured in terms of improvements in transistor speed and number leading to drive current enhancement, to a new era where the progress will be mainly measured in terms of energy efficiency, i.e., switching energy consumption reduction per processing cycle or time. A number of measures and design techniques against the power crisis as stated in the foregoing discussion have been devised at circuit and system

AU: This introduction isn't broken up into Chapters. Please revise.

level, which in view of CMOS technology nodes from 130 nm to 90 nm, allowed the consequent improvement in circuit delay while maintaining scaling of V_{DD} down to 1 V. After the 90 nm technology node, however, it has been impossible to further scale V_{DD} and reduce delay. Circuit delay time is directly proportional to supply voltage V_{DD} and inversely proportional to gate overdrive (V_{gs}-V_T) where V_{gs} is as per V_{DD} value and V_T is transistor threshold voltage. When V_{gs} and V_T both are reduced, the overdrive is a very small value compared to the reduced supply or gate voltage shifting the delay on an upward curve as transistor scaling node is reduced which is detrimental for giga-scale-integration (GSI) logic circuits with their analog counterparts. To ameliorate this scenario, designers had to envisage new schemes to convert geometrical scaling into performance enhancements, e.g., features such as lightly doped drain, ultra-shallow junction technology, retrograde body doping, applying strain to increase mobility—these are some of the technologies that have been laid out in various preceding Chapters of this Introductory section of the book. In addition, increased parallelism and pipelining in CMOS microprocessor based core logic scheduling gradually ceased to maintain its effectiveness in terms of FLOPs when each hardware unit approaches the minimum energy per operation and in fact the number of cores in microprocessors has already started to saturate. The above discussion and the following paragraph serve the prologue to Chapter 6 of this book which discusses the Tunneling FET device architectures as energy efficient switches. ~~Due to lack of space of the book, issue by issue as Tunnel FET device physics and transport are addressed here, cannot be discussed but only discussed in bief.~~

Although conventional CMOS based devices faced barrier in improving the delay time as the power supply is scaled, energy efficient computing is best achieved when supply voltage as low as few kT/q tending to be 0.1 V at room temperature in the vicinity of near threshold or subthreshold operation leads to orders of magnitude reduction in dynamic or switching power. A serious challenge to this aggressive V_{DD} scaling is the requirement to maintain the ratio I_{on}/I_{off} close to 10^6 for steepest switching requirement and which is ultimately set by the subthreshold swing SS of the transistors. Consequently several novel devices have been recently investigated to overcome the fundamental 60 mV/decade thermionic emission based Boltzmann's limit of the SS in CMOS transistors at room temperature where the carrier injection mechanism into the channel is controlled by gate-to-source voltage barrier induced thermal energy driven over-the-top emission. When the carrier injection into the channel is altered from thermionic emission to steep band-to-band (BTBT) tunneling at the source end where the tunneling window defined by tunneling width and height is efficiently controlled by gate-to-source potential, a novel energy efficient and highly manufacturable tunneling field effect transistor (TFET) architecture comes into being. The Chapter 6 of this book introductory section is thus devoted to expatiation of the review and performance analysis of TFET devices (n-type and p-type) through survey of some of the most classic articles from literature references [194–212]. Excellent review papers that the author discusses such as , by Alan C. Seabaugh and Qin Zhang. The physics of operation of tunnel FET is exquisitely narrated

AU: Once again I'm confused about your use of "introductory section"...please reword each instance to let the reader know if you are talking about the current Introduction or about other chapters in this book.

by the authors of the above reference paper where for an ultrathin body sub 10 nm semiconductor transistor with p+-n+ abrupt junction, a tenth of voltage bias at the drain making p^+ - n^+ junction reverse biased with steepest energy profile. Induces large Zener or BTBT tunneling from p^+ valence band to nn^+ conduction band. With a gate aligned to the p^+ - n^+ junction and a metal work function selected to fully deplete the channel when the device threshold voltage is adjusted to operate in enhancement mode or normally off state, application of positive gate voltage favors the Zener tunneling by control of tunneling window as stated earlier and the saturation current extractable from [illegible]s set by the maximum value of BTBT injection. With an intrinsic channel associated [illegible]t to steep p+-n+ source junction, these BTBT injected carriers travel with faster effec[illegible]y enabled by vertical field reduction accompanied by BTBT efficiency than the usually [illegible]bility observed in conventional n-MOSFET with doping induced Coulomb scattering [illegible] field induced mobility reduction. The authors of this paper further state that unlike the [illegible] the n-type TFET and p-type TFET devices can be potentially designed to carry the same current if they use the same tunneling junction, therefore equal gate widths give equal on-currents and symmetric layouts are possible. The authors of this referenced paper then illustrated through a highly precise figure based on measured n-channel and p-channel TFET drain current per micron gate width versus gate to source voltage V_{gs} at a fixed V_{ds} for 32 nm node CMOS on projected state-of-the-art developments with TFETs against CMOS technology. The authors then embark on nicely enunciating the TFET analytic theory to elaborate on Zener tunneling also known as band-to-band tunneling (BTBT) and the primary transport mechanism in tunnel transistors. Zener tunneling current is determined by integrating the product of charge flux and the tunneling probability from the energy states on the p^+ side to those on the n^+ side where the tunneling probability is calculated by applying the Wentzel-Kramers-Brillouin (WKB) approximation to the triangular potential at the p^+ - n^+ junction. The authors systematically developed analytical equations first for Zener tunneling, then 1-D Zener tunneling, Zener tunneling versus material and dimensionality where tunneling current per unit width as a function of internal field for Si, Ge, InAs, InSb and graphene nanoribbons (GNRs) lateral tunnel junction are compared using material properties such as energy gap and tunneling mass. Illustrated in these figures, the authors also showed through simulation the benefits in going from bulk 3D to 1D nanowire (NW) tunneling device taking the case of Si and Ge. The very important subthreshold swing is discussed next in their paper by the authors. From the defined tunneling current relationship as a function maximum electric field at the $p^{=}$ - n^+ junction, the derivative of this tunneling current value with respect to gate-to-source voltage Vgs can be utilized to determine an expression for the TFET subthreshold swing. Next the authors of this referenced paper systematically shed light on the design considerations and trade offs for TFETs with an overview of the geometrical considerations from the perspectives of dependences on supply voltage, gate electrostatics, source doping, and mechanisms limiting off-current. Progress in circuit developmental factors that can degrade subthreshold swing,

like interface trap density and phonon assisted tunneling near turn-off will be supplemented in this subsection 1 by taking note some other reference papers cited next in-line with this subsection to round up the di[illegible]f this one of the key referenced papers on tunnel FET. As the discussion of this first paper is continued, the authors next show the pictorial structures of two approaches for nTFETs: (i) single gate lateral; and (ii) double gate vertical structures in which the gate field originates from the surface, perpendicular to the orientation of the tunnel junction internal field increasing BTBT rate by compacting the lateral dimension of the tunneling window reducing the TFET size and enhancing volume integration. An approach laid out in this reference paper to increasing on-current inserts a degenerate pocket under the gate in both lateral and vertical n-TFET configurations. The pocket increases the area of the tunneling junction in the on-state increasing the on-current. It also assists a lower subthreshold swing by aligning the gate field with the internal tunnel junction field. The effects of supply voltage, electrostatics, source doping, off-state characteristics and circuit status as discussed by the authors in this reference article are directed to the readers' attention. In another classic paper by David Esseni and Tommaso Rollo et al., a comprehensive review analysis is performed in various subsections in this second reference paper on a number of analytical models for BTBT ranging from bulk materials to nano-structured devices where previously omitted (paper 1 in this subsection) quantum confinement results in significant changes of the band structure and thus the reconfigured energy states involved in tunneling process compared to bulk materials. Analytical models are supplemented by the authors of this paper 2 with TCAD oriented models and more computationally intensive full quantum transport models. Several simulation case studies and comparisons to experimentally fabricated devices concerning TFETs with group IV and III-V semiconductors and promising III-V based heterostructures. The reference paper with which Chapter 6 discussion starts is by authors Ian A. Young et al. who started their discussion by noting that although Si and Ge TFETs are studied mostly due to these materials' lower fabrication cost and complexity in spite of their limited on-current drive due to their indirect band gap nature with larger effective mass, III-V materials for TFET can provide the required on current with a compromise of Ioff rise because of their low band gap carrier effective masses and additionally supported by configurations such as broken gap or staggered heterojunctions that have to-date produced the highest TFET drive current. The authors used an atomistic quantum mechanical device simulator to characterize I-V and C-V behavior of both the Si MOSFET and the GaSb/InAs TFET for L_g = 13 nm node. The band structures are calculated using the sp^3s^* tight binding model with spin-orbit coupling and the transport is assumed to be ballistic. The authors later in their paper projected that specifying a square NW (nanowire) geometry device with a body thickness 5 nm gives a heterojunction TFET SS significantly below 60 mV/decade and I_{off} ≈ 10 pA/µm. To achieve reasonable short channel behavior at L_G = 13 nm node TFET requires a 10 nm undoped drain region between the gate edge and doped drain region. The authors in this paper proposed a wrapped around drain TFET structure by use of undoped drain epi growth under

[illegible] and another 5 nm in vertical direction before continuing with doped [illegible] authors provided pictorial sketch that due to the additional gate control through the spacer, the on-state carrier transmission energy profile does not change but the off-state tunneling energy profile has a longer tunneling path leading to lower leakage current. The authors then provide the power performance (minimum energy vs. delay curves) comparisons of CMOS and TFET logic circuits. Next the authors in this paper simulated the effect of device variations on circuit power performance by Monte Carlo simulations of circuits with random V_T modeling. For the device design and challenges section in this paper, the authors provided insights [illegible]dation due to non-idealities, scaling requirements, TFET circuit layout issues and consequences for unidirectional conduction. Another reference paper from the tunnel FET listed papers, described in this subsection by Massimo V. Fischetti et al. incorporates the very important quantum mechanical treatment including multiple electron and hole valleys and interactions with phonons revealing a big shift in the onset of tunneling or decrease of tunneling probability. The readers are referred to Dimitri A. Antoniadis et al. paper that focuses on parasitic leakage current due to Auger generation out of degenerately doped source region which forms an intrinsic link with BTBT tunneling rates and therefore imposes a fun[illegible]mit on ultimate TFET performance. The readers are also referred to paper by Kaustav Banerjee et al. probed further into the hidden physics and design rules leading to small subthreshold swing and the choice of structures and materials based on which both the theoretical and experimental results could be well explained and a general design rule for TFETs with small SS can be formulated. The [illegible]e then directed to another comprehensive review is provided by the authors Guido Groeseneken et al. in their paper titled “The Tunneling Field Effect Transistor” in co[illegible]ction to Chapter 6 contents.

The readers are also directed to Chapter 6 on some discussion focusing on the essential physics of TFET structures as explained by a group of classic reference papers and another set of classic reference papers are concentrated on reducing ambipolar leakage current that enhances I_{off} and reduces I_{on}/I_{off} ratio and studying potential off-state leakage current components, i.e., trap-induced tunneling currents and their temperature dependence and BTBT current as a function of substrate temperature.

The readers from TFET based reference listing are directed to various TFET device structures that boost the device on-current of TFET devices which is generally less than their CMOS counterparts at that specific device node. Different compact modeling of multigate architectures with additional NW geometry on TFET devices will be also highlighted in this subsection taking reference of relevant articles. The use of Ge mole fraction to boost drive current and reduce substhreshold swing in a strained SiGe source TFET architectures along with SiGe channel natural TFET structures will be addressed as important performance tuning scheme laid out in certain reference papers. This

In the final Conclusions of this book, salient features and advantageous impact from device physics based parametric engineering of conventional MOSFETs architectures in Chapter 1, III–V MOSFETs in Chapter 2, Silicon-on-Insulator MOSFETs in Chapter 3, Strain induced MOSFETs in Chapter 4, Multigate engineered MOSFETs in Chapter 5, and Tunnel FETs in Chapter 6 will be analytically compared for best performance as per ITRS roadmap. The historical scaling trendline naturally gave way from conventional Si MOSFETs upto 90 nm node with their silicon-on-insulator counterparts and then from 90 nm to 45 nm strain induced MOSFET architectures were employed with SiO_2 gate dielectric being replaced by high-k gate dielectric and both source, drain and gate contacts being formed by metal gates with mid gap work functions to reduce the flatband voltage shift and enable very low positive turn on threshold voltage. Then from 45 nm to around 22 nm, multigate architectures were introduced with their attributed effects on scaling length parameter that measures the short channel effect aggravation. One of the key features of multigate architectures at less than 30 nm node is the potential to intrinsically dope the channel and use of ultrathin body Si material so that high body doping which gives rise to high V_T shift, increased random dopant fluctuations, increased Coulomb scattering and surface roughness scattering associated drift mobility reduction can be avoided and ultra thin body Si allows the body to be fully depleted and body potential uniformly distributed by improved gate integrity by multi stacking from all sides of the channel resulting in superior scaling length parameter diminishing short channel effects. Another reason of using ultrathin body in multigate architectures is the source-to-drain sub surface leakage current paths are eliminated by entrenching the body thickness. From 22 nm to 10 nm node, the architectures mostly employed are FD-SOI (fully depleted SOI), FinFET, gate-all-around (GAA) FET with transformation to nanowire architectures employing gate all around wrapping and tunnel FET architectures. Therefore, Chapter 1–6 are strategically aligned in sequence to reflect this transition of adapted architectures along the trend line of gate length node evolution. The architectures discussed in different chapters shed light on rather low hole mobility of p-MOSFETs built with these device architectures and important technologies are discussed such as III–V materials based p-MOSFETs and Ge based p-MOSFETs. On top that application of strain, fully depleted silicon-on-insulator, multigate stack and source end tunneling efficiency in Tunnel FETs all can add or impart substantial mobility increments when P-MOSFETs of a basic CMOS inverter logic block are supplanted from silicon to III–V and Ge p-MOSFETs inclusive of all applied device engineering method stated in these Chapter 1–6. Low substrate temperature operation of all engineered device architectures has also shown to boost both n-channel mobility and specially p-channel mobility multi folds with the application of strain and III–V heterostructure mole fraction control. For TFET structures although low substrate temperature operation reduces BTBT tunneling rate from source, certain currents like trap-assisted-tunneling (TAT) and ambipolar leakage currents are reduced at low temperatures than 300 K consequently reducing I_{off} and I_{on}/I_{off} ratio remains still high as per ITRS projected roadmap. The author notes that the best

analytical modeling based device performance analysis of the most evolved state-of-the-art device architectures must be attuned or correlated with the best production or yield data of manufactured devices from the semiconductor companies making design for manufacturing a key ingredient for the purpose of device physical analysis. The concluding section will provide insightful discussions on these topics from the implicated knowledge of various referenced papers covered in Chapter 1–6 of this book much to the interests to device physicists and engineering professionals.

In the Introduction part of this book, the readers are referred to some classical descriptions from references with regards to silicon MOSFET device architecture and its variants. Although the introductory overview has been made very detailed with novel unpursued features before, the various sub sections and their discussions as hinted in these overview discussion cannot be analyzed due to lack of book contents and space and only short overview will be provided which will address the key features narrated the short nature of this book, therefore only key features will be highlighted going forward. ucture. This limitation arises from the size of the book, so each chapter will not have sub section as mentioned in introductory overview but will narrate the key physical attributes of devices that contribute to their performance. The readers will gain their impetus from the introductory overview and if needed, identify the papers from references to augment their understanding on the issue. But as the chapters are detailed here to worth, the contents will be seriously needed adjustment for the fixed book content. The author therefore suggests to the readers that as per introductory overview narrated in the Introductory section, if each chapter and sub section of the chapter could be composed, it would have been more than the limit of the pages of the book but the author has already given direction the readers can pursue to understand scaling based device physics of each device and how even today these devices are maintaining their performance despite the detractors suggesting otherwise referring to an eventual performance bottleneck. Some of the papers referred in this Introduction for Chapter 1–6 are not included in references and the interested readers are encouraged to detect these papers from online reference hub. The reference list appended at the end of Conclusions will cover almost all contents where these referred articles are applicable in description of Chapter 1–6.

AU: I don't think it's good to talk about what you can't include, but rather focus on what you did.

AU: All sources mentioned in the text really should be provided for the reader. They can be added at the end of the Reference list and numbered accordingly.

CHAPTER 1

Device Physics-Based Scaling Insights on Conventional Short Channel n-type MOSFETs

[illegible]timate scaled MOSFET production from design phase, the principal bottleneck that springs [illegible]orizon is the concept of limits of lithography or the utmost level of accuracy of critical dimension (CD) features on the mask being transferred on the wafer by multi-stage lithographic process and precisely calibrated instrumentation set-up. There are three main constituents of the technology improvements that have kept the industry in-line with Moore's Law based trajectory of technology node evolution for a sustained time period up until now. They are: (i) lithography; (ii) increased wafer size; and (iii) design. Roughly half of the density improvements have been derived from innovation in lithographic process and techniques. With the cost to fabricate a wafer remaining roughly constant being independent of its size or contents, this has resulted in 30% reduction in cost per function per year over this sustained period. The ultimate limits of lithography therefore depend on the interplay of science, technology and economics. The author has given a preview of Chapter 1 in the Introduction of this book where the classic reference paper of Lloyd R. Hariott in *Proceedings of the IEEE*, March 2001 titled "Limits of Lithography" [82] was introduced to the readers' attention about the cutting-edge lithography techniques overview articulated in this narrative. The author is at this point summarizing the key discussions in this reference paper along with other reference papers to be discussed in sequence which will be immensely helpful in prognosis of future lithographic techniques and instrumentation design that can be successfully deployed for manufacture-ready 10 nm, 7 nm, and 5 nm technology node of future MOSFET device architectures. Leading-edge production lithography employs optical projection printing operating at the conventional Rayleigh diffraction limit. The image of the master pattern or mask (usually reduced by 4–5 times) is projected onto the wafer substrate that has been coated with a photosensitive material (resist). The solubility of the resist is changed by exposure to light so that a pattern emerges upon development. The remaining resist pattern is then used for subsequent process steps such as etching or implantation doping. The optical projection systems used today have very complex multielement lenses that correct for virtually all of the common aberrations and operate at the diffraction limit. The resolution of a lithography system is usually expressed in terms of its wavelength λ and numerical aperture (NA) of the projection lenses leading to resolution = $k_1 \cdot \lambda/_{NA}$, where the constant k_1 is dependent of number of process sequences used in final pattern

transfer to the die on the wafer. In IC manufacturing, typical values of k_1 range from 0.5–0.8 with a higher number reflecting a less stringent process or reduced number of process sequence. The NA of optical lithography tool ranges from about 0.5–0.6 today. Therefore, in order to maintain the CD resolution on wafer as per scaled node evolution, the smallest resolved wavelength λ of the optical illumination source is necessary. It leads to the typical rule of thumb that the smallest features that can be printed on the die are about equal to the wavelength of the beam being used restricted by Rayleigh diffraction limit. Historically, the improvements in IC lithography resolution have been driven by the decrease in the printing wavelength. The illumination sources were initially based on mercury arc lamps filtered for different spectral lines with progression along feature size evolution from 435 G-line to 365 nm I-line. This was followed by a switch to excimer laser sources with KrF at 248 nm and more recently ArF at 193 nm. The scaling evolution as per technology nodes suggests that the progress warranted in IC minimum feature size is on a much steeper slope than the trendline of reduction factor of lithographic wavelength. This stipulates resolution improvement to be on a more stringent path dictated by IC minimum feature size reduction over successive generations and thus with every lithographic technology improvement, attentions have been directed towards numerical aperture or NA adaptation or engineering in maintaining the resolution of fabricated feature size on the semiconductor wafer.

The ability to print features significantly less than the wavelength of the exposure radiation can be largely attributed to improvements in the imaging resist materials. Modern resists exhibit very high imaging contrast and act as thresholding function on the aerial image produced by the optical system. In other words, even though the light intensity image turns out to be less than full modulation for the small features in traversal of overall contours, the combination of high-contrast imaging material and good process (exposure dose) control can reliably produce sub-wavelength features. In this way, the improvements in imaging resists have lowered the value for k_1. One of the key functional ability of any highly evolved lithographic technique is the image size reduction capability. As has been aforementioned, the image of the mask is generally reduced by a factor of four to five when it is printed on the wafer and equal CD size on mask and die are prohibitive from the point of high-contrast reliable pattern transfer. The main reason for this is due to the mask making process. Masks are patterned by a scanned electron or laser beam primary pattern generator. The original resolution preservation and placement accuracy of the pattern generator are the basis for the optical printing system. Reduction imaging relaxes the requirements on the pattern generators' beam exposure which makes the CD resolution on the pattern generator's exposure beam to be 4–5 times larger than the CD specifications on the wafer. In the regime where feature sizes printed on the wafer are less than the exposure wavelength, the process is deemed highly non-linear. In terms of the mask, this introduces a complication referred to as mask error enhancement factor (MEEF). This regime usually puts the k_1 factor to be less than 0.5 where less of the light diffracted by the mask is accepted by the entrance pupil of the imaging system and contrast is lost. One of

the consequences of lithography in this nonlinear imaging regime is that the enhancement of the printing of the linewidth control errors on the mask. As a result of this printing error, although the overall demagnification of the feature size from the mask to the wafer may still be four to five, small errors on the mask are not suppressed by the magnification ratio. The MEEF is defined as the derivative of the CD on the wafer to that of the mask (corrected for the magnification ratio) or $MEEF = \frac{\partial CD_{wafer}}{\partial \mathrm{CD}_{mask}}$ and is usually < 1. But for a highly stringent or aggressively scaled process node, k_1 factor can be below 0.5 and about 0.35 and for equal lines and spaces on the mask pattern, the mask errors are enhanced by about 2.5 times and MEEF will also significantly increase. To counter this, the requirements for feature size control on the mask are 2.5 times more stringent than one would otherwise expect with reduction lithography performed at larger k_1. Thus, the demands on mask making costs are becoming more intensive and overpowering in relation to successive feature size reduction as optical lithography is pushed below the wavelength of exposure.

Some compensation for the image degradation from diffraction are possible by pre-distorting the mask features. A simple example is a correction for corner rounding by using serifs. The addition of sub-resolution features does not enhance the quality of the image on the wafer somewhat, but requires the addition of these correction features on the mask, increasing its complexity and cost. This kind of approach on the mask pattern is referred as optimal proximity effect correction (OPC) and will be discussed in a while. Increased NA of light concentrating lenses is another route to improved resolution in optical lithography. Improved optical designs aided by sophisticated computer modeling are enabling larger NA lenses to be designed. KrF systems with NA = 0.7 will be available shortly with NA = 0.85 systems being designed. The penalties for these very high NA systems are primarily in cost and reduced depth of focus factor affecting image quality and contrast transferred on the wafer. The cost of the lens and thus the lithography system scales roughly with the cube of the NA (volume of the lens material). At these large NAs, the weight and size of the lenses also present many practical issues. The depth of focus (DOF) of a system can also be characterized by the wavelength and NA as DOF = $k_2 \frac{\lambda}{NA^2}$ where k_2 is also a process dependent parameter generally taken to have the same value as k_1. Since DOF $\propto$ 1/NA^2, this clearly shows the penalty in DOF for high NA systems. Reduced DOF requires extremely tight control and planarity in the wafer process. For even a modest 0.6 NA system, the DOF is only a few hundred nanometers. Usable NA will be limited to some value less than theoretical limit of 1 by DOF considerations as well as effects caused by the refraction of the high angle light and polarization effects in the resist film. A practical maximum is currently thought to be NA = 0.85. Improvements in illumination methods have also been adopted to improve lithographic performance. The resolution limit equation reflects the fact that the first diffracted order must be captured in the lens as complete as possible for the image information to be transferred by the optical system. Off-axis illumination (OAI) uses a tilted illumination to capture one of the first diffracted orders while allowing the zero order also to pass. In

other words, the resolution performance can be doubled in principle since the aperture of the lens now must cover the zero and one of the first diffracted orders, whereas for a conventional system the aperture covers the span of the -1 and +1 orders. In practice, however, this method is limited by the fact that the illumination must be tailored to the mask pattern since the diffraction pattern will be different for different mask pattern. Specific illumination patterns (annular or quadrapole symmetry) are chosen for specific types of circuit patterns to emphasize the performance of specific features such as line-space gratings.

Other strategies to improve resolution include phase modulation in addition to the image amplitude modulation introduced by the mask pattern in the imaging system. Most photomasks are made using a transparent quartz (synthetic fused silica) substrate with a patterned metal layer (usually Cr) as an absorber. The mask and thus the image are then binary, full or zero intensity. The optical system then creates an aerial image at the wafer plane that is a convolution of the binary intensity pattern and the diffraction point-spread function of the lens. Phase modulation or phase shifting can improve the effective resolution of the system through constructive and/or destructive interference in addition to the amplitude modulation. There are a great many implementations of phase shift masks (PSM) but the earliest form was the alternate aperture PSM. For want of space of this Chapter 1, the author s suggests the readers to refer to this reference article discussion on alternate aperture PSM. A variant of this PSM using two masks for gate levels utilizes the first mask which is a binary mask to print the gate level of a circuit with conventional design rules. Then a second mask is used to re-expose the same resist pattern with a phase-shift pattern. This "trim" mask acts to reduce the size of the gates and can, in principle, achieve linewidths as small as half that achieved by conventional masks. In the time of publication of this reference article and as per the knowledge of the author of this reference article, this method has been applied to the gate level of a three million transistor digital signal processor (DSP) chip yielding gate lengths of 120 nm using 248 nm KrF lithography. An important aspect of the two mask PSM approach [illegible] other resolution enhancement technologies (RET) is that while they improve the minimu[illegible] size that can be printed, they do not improve the density of features over conventional bi[illegible] lithography. The phase shift approaches improve the printing of isolated features such as gates but do not improve the minimum distance (half pitch) for the patterns. In terms of IC performance, smaller gates will improve the circuit speed and power consumption but they do not increase the packing density or number of circuit elements per chip. This has led to a bifurcation of the roadmap for the IC industry over the past few years. The progress in the minimum half pitch or density of circuit elements has been driven mainly by memory applications and has continued to follow the Moore's Law trend with some amount of acceleration due to improvements in resist and NA (numerical aperture). Meanwhile, minimum feature sizes are driven by the speed requirements of the microprocessors and have been following a more aggressive trend. In general, gate sizes are almost one generation ahead of circuit density or minimum half pitch as a result of limitations of RET.

In a very recent paper published in 2018 entitled "Promising Lithography Techniques for Next Generation Logic Devices," Rashed M. Hasan et al. provides an excellent overview of next generation lithography supplanting the optical lithography for resolution enhancement. The author has chosen this second reference article full of insightful discussion of potent next generation lithography tools and techniques that have already been deployed by semiconductor manufacturing plants and the discussion that follows next has excerpts and implications of this second reference article. As the conventional optical photolithography has approached its ultimate limit dictated by Rayleigh wavelength diffraction, considerable efforts have been devoted to next generation lithography enabling techniques by various research laboratories and industries around the globe. These techniques are extreme ultraviolet lithography (EUVL), electron-beam lithography (EBL), focused ion beam lithography (FIBL), nanoimprint lithography (NIL) and directed self-assembly (DSA) with the order of listing of these techniques is guided towards increasing favorability to highest degree of resolution enhancement of sub-nm feature sizes. From the resist material functionality engineering perspective in terms of its structure, the next generation potential ULSI techniques warrant the selection or development of novel engineered resist materials capable of providing high resolution in pattern transfer, high selectivity enabling selective etching of the pattern and low linewidth roughness. However, at the limits of lithography, it has been proven difficult to achieve high resolution, high selectivity and low line edge roughness simultaneously due to an inherent trade off relationships co-existing between each other. Therefore, frontier research on the development of advanced resist materials such as molecular resists of chemically amplifiable nature will be required to favorably impact the trade off between resolution, line edge roughness and selectivity. In addition, the next generation resists must have the ability to mitigate stochastic barriers. The advancement of novel resist materials is entering a new research realm with the accompanied challenges and opportunities to fulfill the stringent requirements for the future patterning techniques with minimum undercutting or near vertical etch profile capabilities. Figure 1.1 is a picture from author Lloyd R. Hariott's article "Limits of Lithography" [82].

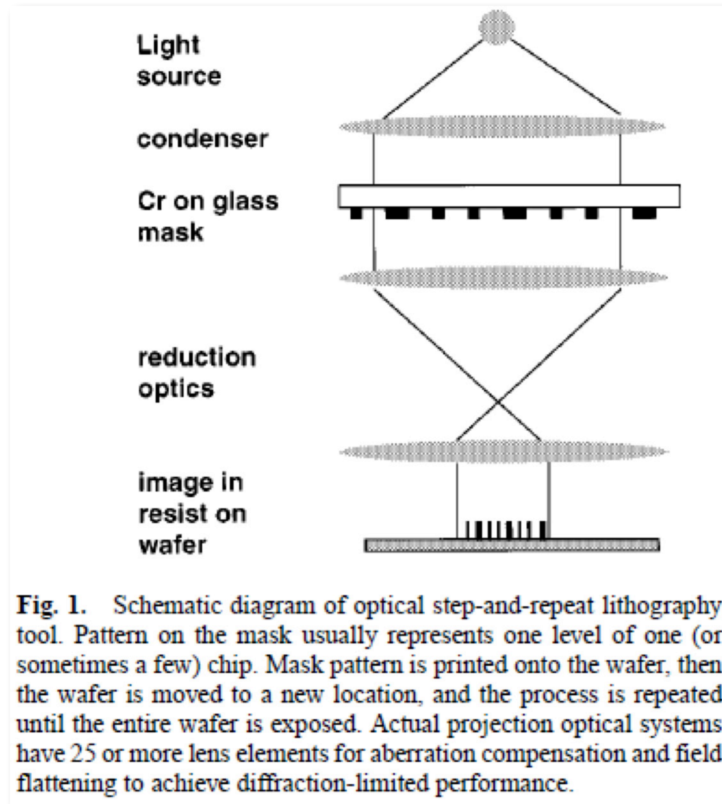


Fig. 1. Schematic diagram of optical step-and-repeat lithography tool. Pattern on the mask usually represents one level of one (or sometimes a few) chip. Mask pattern is printed onto the wafer, then the wafer is moved to a new location, and the process is repeated until the entire wafer is exposed. Actual projection optical systems have 25 or more lens elements for aberration compensation and field flattening to achieve diffraction-limited performance.

Figure 1.1: This picture is taken from the article "Limits of Lithography" by Lloyd R. Hariott [82]. Copyright permission is required for this picture.

At the forefront of next generation lithography enabling techniques, mainstream instrumentational development and research were directed on extreme ultraviolet lithography (EUVL). In the last decade, researchers put extensive interest in EUVL as a "next wavelength" replacement for 193 nm dense ultraviolet (UV) lithography. For the EUVL image transfer process description, for want of space in this Chapter 1, the readers are referred to consult this reference article. Over the last few years, considerable progress has been made to move EUVL towards increasing high volume manufacturing viability. Most remarkably, there have been substantial developments to exposure throughput, reliability variance control, and patterning materials in line with high resolution requirements. Currently, EUVL is projected to be used in manufacturing at the 7 nm node or beyond. The source power, masks and resist materials still pose critical issues to be resolved for mass production. For the future technology at the 5 nm node and beyond, source powers of 500–1000 W at a reduced operational cost per wafer may be required. Laser produced plasma (LPP) and discharge produced plasma (DPP) are the two main techniques to produce EUV sources. The source power has been maintained to improve ten times in last five years. It has been reported by reference that

more than 250 W LPP-EUV power could be generated by using plasma generation schemes. Using free electron laser source, many tens of kilowatts power can be produced. CO_2 driver laser system is capable of providing high power high voltage manufacturing (HVM) LPP-EUV source with more than 25 kW output power. A key factor for adoption of EUVL in HVM is the choice of a typical set of EUV resist with high sensitivity, high resolution, low LER, low LWR, and better contact hole CD uniformity features. Chemically amplified resists (CARs) have effectively [illegible]ed the scaling requirements of the semiconductor industry. RLS (Resolution-LER-Sensitivity) performances and stochastic variations developed from inhomogeneity of the structure of the resist are the key issues for the CARs as well as for other resist materials. High sensitivity (<20 mJ/cm^2) resist materials are required to reduce the development of high power exposure sources that in turn lead to larger LER values. Acid diffusion in CARs influences these performances. It has been experimentally demonstrated that by selecting an appropriate ratio of quencher to PAG (Photo-acid generator), an EUV dose reduction of up to 12% can be achieved with 240 s post exposure bake time while keeping LWR and resolution constant. For better optimization of RLS parameters, non-chemically amplified resists (non-CARs) have been used since they have no acid diffusion issues resulting in high resolution capability, high sensitivity and low LER. Some researchers have reported the development of the metal tin (Sn) containing photoresist that shows high sensitivity performance giving way to enablement of low energy power source to realize EUVL. Some other new techniques including nanoparticles photoresists with high sensitivity have been reported. Lately at the *2016 SPIE Advanced Lithography Conference*, a good numb[illegible] was presented demonstrating substantial research on photosensitized CARs. However, it is urgent need to mitigate the stochastic failures such as broken line, nano-bridge, merging holes and closing holes. These nano failures are influenced by many factors including aerial image quality, photon absorption, acid shot noise, and acid diffusion. The probable solution lies with co-optimization of variety of different aspects (materials used, hardware and metrology tools, etc). Adjusting exposure dose can be an effective knob to drive down failures. Higher dose absorber materials can reduce the stochastics. Moreover, the substrate underneath the resist influences the exposure dose and the LWR and therefore optimization of the substrate, such as, pre-treatment before resist coating, adhesion promoter, resist layer thickness tolerant stress and strain in the substrate leading to coating variation and variation in resist exposure, etch, and bake time, could be a potential improvement knob to the exposure dose selectivity and LWR reduction. Mask blank defects and yield limit the applicability of EUVL. Extensive researches are still needed to improve mask materials, fabrication process, defect inspection, and disposition metrology and mask protection. Highly transmitted and long-lasting pellicle marks are desirable to improve overall process defectivity. The interaction between the oblique incident EUV light and the patterned absorber may cause the mask 3D effects at the wafer level. For the defect free mask manufacturing, an EUV aerial image metrology system has been developed by many noted lithography metrology and tool providers to actinic review of EUV mask. These actinic

tools are very useful for blank inspection, pattern mask inspection, and defect repair verification. In support of EUV roadmap, 13.5 nm 0.5 NA micro-field exposure tool has been designed. According to ITRS 2015 report, ASML is going to produce a 0.55 NA EUV scanner with different magnification in both X and Y directions and it could be available for use in manufacturing in 2021. Advantages of EUVL are high throughput, wide process windows and extensibility to future nodes. It uses a smaller wavelength allowing more densely packed components on the microchip creating faster processing power. This technique has the potential to provide economic sustainability with its applications in nearly every field including engineering and medical electronics. Reduced power consumption and a lesser number of exposures make the EUVL more cost effective in most nanometric patterning processes. EUVL does face some challenges from higher start-up costs, complexity, reliability, and relative infrastructural immaturity.

After highlighting the features of EUVL, the authors of this second referenced article then discusses the maskless lithographic techniques such as electron beam lithography (EBL) and Focused ion beam lithography (FIBL) that are widely used in nanostructure patterning and IC fabrications with the capability to form arbitrary two dimensional patterns down to the nanometer scale. EBL uses an accelerated electron beam to dramatically modify the solubility of a resist material during a subsequent development step. The electron beam is focused on the resist and then scanned on the surface of the resist with the diameter of the e-beam as small as a couple of nm in a dot by dot fashion. Then the patterns can be transferred to the substrate material by etching like other lithographic methods. Similarly, FIBL involves the exposure by an accelerated ion-beam to directly hit the sample surface. When the high speed ions hit the sample surface, energy is transmitted to the atoms on the surface which leads to five possible reactions: (1) sputtering of neutral ionized and excited surface atoms; (2) electron emission; (3) displacement of atoms in the solid; (4) emission of photons; and (5) chemical reactions. However, both these methods suffer from low throughput that limits their applications within research and mask/mold fabrication. Multiple e-beam direct write (MEBDW) can increase the throughput of EBL with nanometer resolution using > 10,000 e-beams writing in parallel. In recent years some progresses have been reported including MAPPER (a 5 kV raster wafer writer), IMS (50 kV raster mask writer, single source, many spots in single lens field) and multi-beam wafer writer. These are the promising solutions in exposure cost reduction for 20 nm half pitch and beyond. MAPPER's 3rd generation platform (FLX) using 650,000 beamlets has been introduced with a target of 40 wafers per hour throughput. For the case of EBL, PMMA is one of the high resolution resists that is commonly used for having the properties of high selectivity, sharp contrast and better roughness at the edge profile. Some CARs and Hydrogen silsesquioxane (HSQ) which is a commercially successful non-CAR resist are also used to minimize RLS trade off and HSQ resist for EBL has high resolution (sub-5 nm) capability, high etch resistance, and small local CD uniformity. However, these resist display relatively low sensitivity. In the last decade, less than 10 nm resolution capability by maskless lithography has been repeatedly

reported. Electron microscope equipped with pattern generation modules enables nanoscale patterning within desired areas. The nanometer pattern generating system (NPGS) is one of the popular SEM (scanning electron microscope) lithography system that provides a powerful, versatile and user-friendly system for doing advanced EBL or ion beam lithography using a commercial SEM, scanning transmission electron microscope (STEM) or helium ion microscope (HIM). According to ITRS roadmap, the key challenges for these maskless technologies is to build a pilot tool for patterning entire wafers with chip like patterns and overlay control. The earliest deployment of such kind of technology is expected in 2021 and the target would be the 5 nm node. As these techniques are maskless, mask defects and inspection requirements are relaxed and they are the ideal tools for flexible generation for low volume applications. Due to their intrinsically high resolution, excellent pattern generation can be achieved. These instrumentation operability are highly automated and equipped with very accurate control of pattern with direct writing. EBL has greater depth of focus naturally not achievable by optical lithography systems for their twin requirements of ultrashort λ and smaller NA for resolution enhancement. Because depth of focus can be extended by EBL, they provide low diffraction for thicker mask samples and hence also a great choice for the formation of masks and templates that are used in optical lithography and nanoimprint lithography. On the other hand, both EBL and FIBL have the drawback of low speed and low throughput. They also suffer from scattering and over exposure problems.

As an important and versatile next generation lithography technology outlaid in this second reference article is the nanoimprint lithography (NIL) that is capable of high throughput patterning of nanostructures with high resolution (down to the 5 nm regime). Because of the low cost, reduced process steps, and high fidelity, NIL has become an attractive technique for a wide range of applications. Nanoimprint lithography methods can be classified into four categories: thermal NIL; UV-NIL; laser assisted NIL; and electrochemical nanoimprints. Due to lack of space in this Chapter, the author directs the readers to consult the reference article for a short description of each of the above NIL method. The adaptation of NIL on a larger scale is fraught with issues such as defectivity, contamination, throughput, and overlay accuracy. Again, for lack of space the author directs the readers to consult the reference articles where each of these issues and their remedy is covered in detail. NIL is a rapid prototyping enabling process with ability to fabricate nanopatterns of high aspect ratio in a short time span allowing higher throughput for certain imprint applications. It has also low cost of ownership and high resolution extendibility. NIL thus could offer its 3D patterning capability for the advancement of emerging 3D chip technology with heterogenous integration. However, low overlay registration accuracy and thermal expansion effects during mold formation and curing are the disadvantages of the NIL. One of the drawbacks of the NIL over other nanofabrication techniques is the flexibility of ULSI integration based highly complex pattern formation which requires very thin template like mold pattern transfer through embossing

on the substrate could be very challenging for NIL. The mold must be remanufactured when the revision of the designed pattern takes place.

The most recently emergent versatile next generation lithography method as outlaid in this second reference article by the authors is the directed self assembly (DSA) for high volume low cost manufacturing at a sub-nm lithographic resolution. DSA enables finer resolution that attracted a great deal of interests from major semiconductor manufacturers. Besides, recent developments in DSA materials and processing make it a compelling next generation patterning technique. There are two types of DSA processes: "epitaxial self-assembly" (chemoepitaxy) and "graphoepitaxy." In epitaxial self assembly, dense chemical patterns are employed a direct block copolymer (BCP) self assembly. Highly ordered nanopatterns can be achieved if the period of the surface chemical pattern is proportionate with the equilibrium period of the BCP self-assembled nanostructure. Graphoepitaxy guides patterning by topological geometry for DSA. The selective wetting of a BCP component at the trench side walls enforces the lateral ordering of the self-assembled BCP nano-domains along the trenches. Thus, it improves the pattern density by subdividing the topographical pre-pattern. DSA pattern defects, pattern uniformity, pattern placement accuracy, material quality control, cost, and ease of integration into manufacturing flows are the critical issues to adapt DSA technology into semiconductor manufacturing. A very useful discussion is available from the authors of this reference article for the interested readers who wish to probe DSA further. DSA can also integrate bottom up self assembly with the top down conventional lithography. The probable integration of DSA with ArF immersion photolithography for 16 nm line/space DRAM process has been scheduled in 2018. Researchers are investigating the possibility of hybrid DSA processes (combination of both chemo and grapho epitaxy) which can be possible alternatives for sub 5 nm process nodes. Recently several application fields for DSA other than semiconductor device process such as flexible/transferrable DSA technology utilizing chemically modified graphene (CMG) have been demonstrated. Several DSA consortiums initiated by CEA-LETI, IBM, and IMEC are involved in systematic investigation to integrate DSA effectively into commercial semiconductor process. However, further research has to be invested on process optimization, perfect defect control, effective pattern transfer and relevant material development to make DSA a forefront next generation lithography solution. By employing DSA in IC manufacturing, the overall resolution can be increased to a level that is compatible with the future 7 nm and 5 nm logic nodes. DSA could simplify and reduce the process steps of other lithographic process like EUVL. It can ease process integration and provide low cost processing in HVM. Unlike EBL, FIBL, and NIL, DSA does not suffer from low throughput. Moreover, DSA techniques can transfer complex patterns on the substrate with efficiency matching that of EUVL and MAPPER. The block copolymer of DSA have the defect repairing tendency of its own in the patterns manufactured by other photolithographic techniques. By DSA, it is possible to define accurately the orientational and angular features containing landscape, structural dimensions, and pattern density or half pitch that becomes of the

order of λ/2 at which lithography limits of optical lithography systems are encountered. However, in terms of LER and CD control, DSA seems still lagging behind EUVL.

Two of the most versatile lithographic schemes after direct self assembly (DSA) and nano imprint lithography (NIL) are scanning probe lithography (SPL) and thermal-scanning probe lithography (t-SPL) nicely propounded and expatiated in the references titled *Advanced Scanning Probe Lithography* by Ricardo Garcia et al. [84] and *Sub-10 Nanometer Feature Size in Silicon Using Thermal Scanning Probe Lithography* by Yu Kyoung Ryu Cho et al. [85]. The author is providing a condensed overview of these two lithographic techniques and interested readers are suggested to consult these two references for the device manufacturing applications where these two technologies compete with EUVL, Nanoimprint, Maskless lithography, and DSA. As has been detailed in the reference article by the authors Ricardo Garcia et al., since their invention, scanning probe microscopes have been used to image, modify and manipulate surfaces at the nanometer and atomic scales. Recent developments with techniques operating in ambient atmosphere have shown that some scanning probe nanolithography approaches could also be competitive in terms of resolution, throughput and versatility of the materials that can be patterned. For example, thermal SPL has achieved a resolution of 10 nm while throughput is in the 10^4–10^5 μm^2 h^{-1} range. In general, the ability of SPL to image the surface of a material, to fabricate complex patterns in situ with sub 10 nm precision in size and single nanometer accuracy in positioning and to allow post patterning in situ metrology is rather unique. Applications of SPL to pattern silicon, graphene, piezoelectric/ ferroelectric ceramics, polymers, and proteins have been demonstrated. The capability of using the same SPL set-up to pattern different materials at the same time is also very appealing. Besides, thermal scanning probe lithography as discussed by the reference article of authors Yu Kyoung Ryu Cho et al. has demonstrated an overlay accuracy of less than 5 nm and the capability to fabricate 3D depth profiles with nanometer scale precision which is highly promising from IC manufacturing standpoint. There is other exciting lithographic technique on the horizon using quantum interferometric lithography that enables sub-nm resolution breaking the Rayleigh limit of optical lithography. In the reference article titled "Quantum Interferometric Optical Lithography: Exploiting Entanglement to Beat the Diffraction Limit" [93] by Jonathan P. Dowling et al., it has been postulated that using non-classical photon number states, entangled N at a time, it is possible to write feature size of resolution λ/2N in an N-photon absorbing substrate. This allows one to write a factor of N^2 more elements on a semiconductor chip. The N-photon absorption cross section with N entangled photons scales as I (input flux) but not I^N as is expected from classically uncorrelated N-photons. This illustrates the point that although classical N-photon lithography requires unrealistically high optical powers, entangled N-photon lithography requires the same level of power as classical one-photon device whereas enabling CD accuracy down to N-th order. Finally computational lithography and computational metrology are additional versatile techniques to optimal proximity corrections of CD image on the wafer as described by two papers titled "Com-

putational Lithography and Computational Metrology for Nanomanufacturing" [86] by Shiyuan Liu and "Computational Lithography: Exhausting the Lithography Limits of 193-nm Projection Lithography Systems" by David O. S. Melville et al. For lack of space, the author describes some observation from the first article by Shiyuan Liu in connection to computational lithography and directs the readers to peruse the contents of these two articles for more detailed information. As has been described by Liu in his published article, the main physical components of a typical lithographic imaging system contains four elements, namely: the illumination source; the mask pattern; the projection lens; and the aerial image or the resist image. According to the particular nanomanufacturing requirements including predictive simulation of lithographic processes, optimal design of mask patterns and in-line characterization of lens properties, the aerial image intensity, the mask function, and the wavefront aberration need to be fast and accurately calculated. The computational lithography involves three key issues and techniques, namely: (1) the aerial image simulation; (2) the mask pattern optimization; and (3) the pupil phase retrieval. The aerial image simulation for a specific lithographic imaging system is a traditional forward process in computational lithography. The mask pattern optimization, which is well known as the inverse lithography technique, aims at deriving the optimal mask pattern for a desired image. The pupil phase retrieval can be treated as an extended inverse problem in computational lithography. From the computational metrology section, the author of this reference article focuses on scatterometry which is based upon conventional reflectrometry or spectroscopic ellipsometry, in which a collimated beam that is polarized in a known state and projected onto the sample, the state of polarization of the reflected wave is analyzed. From the incident and reflected states of polarization, the intensity of the reflected light under different states of polarization or the ratio of complex coefficients for the incident orthogonal linear polarizations parallel and perpendicular to the plane of incidence are determined. These output parameters are subsequently related to the structural and optical properties of the ambient-sample interface region by invoking an appropriate model and the electromagnetic theory of reflection. Eventually, the model parameters of interest can be determined by solving the corresponding inverse problem. Further details can be gathered from the article with very useful technical detail.

1.1 DEVICE PHYSICAL ATTRIBUTES OF CONTACT RESISTANCE FOR SCALED N-CHANNEL CONVENTIONAL MOSFET

lc

Contact Resistance impact on device performance and the physics of contact resistance formation that allows a device performance with boost in drive current in terms of source and drain reduced contact resistance. The first paper that fundamentally extracts the key physical aspect of observed contact resistance variation among various metals and their salicides are Schotty Barrier Height or (SBH). As the author opines, the lack of a coherent explanation of a wide spectrum of experimen-

tally observed SBH data and then lack of control over the magnitude of SBH speak volumes about the difficult and complex problem of SBH nature itself. The main sticking points can be traced to a sharp dependence of the SBH on the atomic structure of metal-semiconductor (MS) interface. Presently, short of full-fledged numerical calculations, there exists no comprehensive and quantitative theories on the formation of interface dipoles to explicitly address the most crucial part, the structural dependence of the SBH. As the author in this classic review points out that through an examination of experimental and theoretical results on atomically controlled MS interfaces and known fundamental principles, a comprehensive quantum mechanics based picture of SBH formation can precisely calculated. However, as it is the problem of quantum description of any realistic and practical phenomena, the formation of SBH shows distinct structural setting intricately dependent on the specifics of MS interface under study. Therefore, quantum description SBH in general cannot be formulated into simple analytical equations that are applicable to all MS systems, especially since there are two many important details about the chemical nature of many MS interfaces materials that are not experimentally reliably extractable. Now, while delving into SBH further, the author of this classic paper termed that why SBH is so mysterious and unpredictable is due to its interface chemistry. Experimental results confirmed the absence of a strong dependence of SBH on metal work function and correspondingly resulting in Fermi level pinning fixing the barrier potential with an unchanged Fermi level and making the SBH's effect on contact resistivity worse. The author's physical explanation to this nature is that the charge distribution in any real MS interface turns out to be significantly different from a simple superposition of the charge distribution on the original surfaces. In reality, during formation of an MS interface, the metal and semiconductor come within close range of one another. From quantum mechanics we know that when an atom is brought so close to another atom, some orbitals from the atoms significantly overlapped, these isolated orbitals are no longer eigen states of the system. Instead, molecular orbitals are formed which will lead to chemical bonds. The chemistry at the MS interface should not be any different because metals and semiconductors are made of atoms. Electronic states that belonged to only the metal or only the semiconductor have to be modified so that they can be incorporated in the larger quantum mechanical system. States that are related to the surfaces of the two crystals will not survive the formation of MS interface needing serious modifications. New interface states that are localized in the interface region and are interacting with interface atomic structure may form. Because of chemistry, charge distribution at an MS interface will be quite different from a linear superposition of the charge distribution in two individual surfaces. This difference of charge rearrangement at the interface can be attributed to formation of additional interface dipole. The interface dipole can be largely viewed as the result of transfer of charge between the metal and semiconductor.

AU: What paper??

In the next paper, advanced source and drain technologies for the parasitic source drain resistance reduction is discussed. In the source drain region of a MOSFET, contact resistivity ρ_c is exponentially dependent on $\frac{\phi B}{\sqrt{N}}$ where N is the active doping concentration and in the source and drain

contact and hence is made extremely degenerate and ϕ_B is the Schottky Barrier Height between and source and drain region and metal silicided layer. Without Fermi Level pinning, surface states and metal induced gap states φ_B is generally engineering tunable but in all fabricated MS interface, φ_B seems to be fixed in nature, highly degenerate values of N is required. Nevertheless, fabrication and process techniques have been designed that aim at reducing φ_B by selection of metal silicide work function. Silicides of low and high work function are needed for contacting n-type and p-ype semiconductors respectively. Dopant segregation technique allows the contact resistivity to improve in presence of high doping and when the degenerate limit is high, some of these dopants may also enhance contact resistivity through incomplete ionization although very small. A third approach involves the interface Schottky Barrier Height engineering, for instace, introduction of atomic species like Sulfur (S) or Selenium (Se) for n^+ Si or Aluminium for p^+ silicon. The segregation of Sulphur at the metal semiconductor interfaces with NiSi revealed that electron SBH values tend to converge towards the conduction band of the semiconductor. This also opens up the possibility of implementing a single metal silicide to independently control contact resistances of n and p-type MOSFETs. For electron barrier height ϕ_B^nadjustment for the n-FETs, rare-earth silicides generally have low work functions and may be used as salicides in S/D contacts Other notable silicides are ytterbium silicide and erbium silicide. Metals such as Yb, Er, Ti, and Al were cosputtered with Nickel and annealed to form Ni alloy silicides. However, the metal introduced may not necessarily distribute uniformly in NiSi and the extent of ϕ_B^ntuning depends on whether the added metal can be found at high concentration near the interface with the heavily doped S/D. Ti, Dy, and Yb tend to redistribute at the top surface of the silicide thereby limiting their effectiveness in reducing ϕ_B^n. In addition to tuning the work function of the silicide contact though material selection, another approach as just mentioned priorly is to exploit dopant segregation effects to boost the active dopant concentration in the S/D region beneath the silicide. In this approach, traditional dopants such as As+ ions are implanted at very shallow depth prior to silicidation. For As segregation, does of As^+ implanted may range from 1×10^{14} cm^{-2} to1×10^{15} cm^{-2}. As segregates at the NiSi/Si interface during the nickel silicidation process. The heavily doped region of segregated As^+ at the contact interface effectively reduces R_c. Furthermore, it has been shown that the dopant segregation can be combined with nickel alloy silicidation or work function tuning method to yield additive effects. NiAl silicidation when combined with As^+ segregation gives greater reduction in ϕ_B^nthan either of the method alone.

AU: Do you mean 'dose' or 'dope'?

AU: Please provide exact reference information here for the reader.

A classic paper that discusses study on the intrinsic limitations on contact resistances will be discussed here. In this paper the authors opine that recently cumulative efforts of the scientific and engineering communities have substantially extended the physical limits of the source and drain contact resistance thanks to the increase of activated dopants in the S/D contact region and to the Engineering of Schottky barrier which have been discussed in previous paragraph. As a result, contact resistivity as low as 2 × 10-9 Ω cm^2 and 8.4 × 10^{-10} Ω cm^2 n-Si:a-TiSi and p-Si$_{30}$Ge$_{70}$:

a-TiSi with dopant concentrations as high as 6 × 10^{20} cm^{-3} to 1 × 10^{21} cm^{-3}. Aside from pushing the doping limits in conventional semiconductors, new materials are also being investigated which may allow better scaling and transport by virtue of their properties. Two dimensional materials (2D) specially transition metal dichalcogenides have emerged as possible candidates in this regard for both MOSFET and tunnel based FET applications. Electrostatic doping in 2D materials also promises higher carrier concentrations allowing for better transport. Although these accomplishments are impressive, little is known on the intrinsic limits that will ultimately determine the contact resistivity of conventional 3D and 2D semiconductors and it is unclear at the moment while higher bound of dopant species will keep the contact region to perform equilibrium kinetics due to many body or atomic concentrations. The authors in this paper have shown that through first principles calculations with Non-Equilibrium Green Functions transport simulations, whereas metal/semiconductor intrinsic contact resistivity initially being scaled doping concentration, it is found to saturate at 2 × 10^{-10} Ω cm^2 at a doping level of 5 × 10^{21} cm^{-3} that ma make the contact resistivity calculation deviating from equilibrium at high degeneracy being required. The authors through their simulation found beyond this doping concentration, the electron injection is governed by Ohm's Law. It turns out that at high doping regions, the electron injection does no longer depend on barrier height set by the interface potential to be crossed. In turn it also becomes less sensitive to the chemical composition of the interface or degree of disorder as suggestive by surface states or metal induced gap states, but the carrier injection is limited by intrinsic transmission probability of electrons being injected from metal into the semiconductor and imposes an intrinsic physical limit to the contact resistance since neither ϕ_B^N tuning nor dopant activation magnitude can lower the contact resistivity set by this limit. This intrinsic transmission probability depends on the value of the Fermi energy level of the metal, on the effective masses of the metal and semiconductor and imposes an intrinsic physical limit to contact resistance. A very important conclusion from this paper is that in high doping regime where intrinsic transmission probability is the dominant process, contacting metals with a heavy electron effective mass S/D region help increasing the transmission probability and hence contribute to semiconductor metal intrinsic contact resistivity. Contact resistivity related other conceptual findings as narrated in the Introductory section of this book are not possible to discuss here for lack of space but the readers can be familiar with the concepts and whenever possible gather the related information from references. In this context, the author strongly infers that for metal salicidation of source and drain contacts, only those metals which have degenerate carrier density in the vicinity of maximum number of atoms per unit cell for silicon, i.e., 5 × $10^{22}/cm^3$ should be identified for reduced contact resistivity. Metal silicides with metal degenerate carriers above 5 × $10^{22}/cm^3$ will not be stable and there will be a departure from equilibrium kinetics enabling us to derive the contact resistivity, Schottky barrier height, etc. So, with the drive to ensure intimate salicidation technology with metal silicides, we must study metals with their crystal structure and determine number of atoms per unit cell and only retain those

AU: Referring to sub-sections is very confusing as there are no headings within each chapter. I suggest that you break the chapter up using section numbers and titles and refer to those in text for better reader understanding.

metals for silicidation formation for which free carriers for the metal is still below $5 \times 10^{22}/\text{cm}^3$. Therefore, from the above paper, the increasing transmission probability contributing to silicide material's intrinsic contact resistivity is only valid when equilibrium kinetics can still be applied and hence the silicide forming metal's maximum free carrier density derived from its number of atoms per unit cell is very important to include in study and research when determining intrinsic contact resistivity reduction limit for all metal silicides in silicon.

The third sub section of Chapter 1 as introduced in Introduction part of this book is not analyzed further as the stationary and non-stationary transport insights referred there can be related to the relevant reference materials for the readers. To me as author of this book, the most fundamental technology that will enable these tiniest device to be reliably operable is the advancements in lithography with batch manufacturing techniques and smallest possible CD (critical dimension) definition. This is the reason I maintained coherence with Chapter 1 discussion of different lithographic techniques with the suggestions provided in the Introduction part of this book. The fourth subsection of Chapter 1 where ultra-shallow junctions using mono-layer doping technique was introduced are also omitted for further discussion for lack of space as the book not only discusses device physical aspects of conventional MOSFETs but also III-V MOSFETs, strained silicon, SOI MOSFET, Multi gate and Tunnel FETs and each device itself is worth detailing a comprehensive book. Therefore, in the subsequent chapters the composition part will be rather brief compared to the summary overview the readers detailed in Introduction of this book which were being made familiar with. As I stressed, it is because of lack of space to discuss all these different device architectures in a single book.

referred to

COMP: Link

AU: This sentence is confusing as you talk about what is omitted due to space but then you outline what is included. Please revise.

CHAPTER 2

Device Physics-Based Scaling Insights on III-V Short Channel n-Type MOSFETs and its Derivatives

In continuation toward Moore's Law based scaling evolution, fundamental device architectures built in silicon devices are reaching their performance limits in terms of key device benchmark factors like drive current I_{on}, maximum value of subthreshold leakage current I_{off}, I_{on} to I_{off} ratio, device transconductance g_m, allowable lowest subthreshold slope at T = 300 K, maximum high-field drift velocity and high-field mobility. Therefore, researchers and academicians have pursued alternative approach in replacing silicon material to build these different logic devices and much to the boon to these efforts, alternative device structures compatible with silicon based manufacturing technology evolved with III-V and II-VI materials that have significant higher velocity, lower effective mass, and higher mobility up to ballistic limit with which most of the benchmark metrics in regard to a logic device's performance are evaluated. Logic devices that were built in silicon 10–15 years before are now readily and reliably fabricated in n-type III-V materials and p-type III-V materials. The author now systematically discusses some key review papers on III-V materials based FETs with emphasis on their device physics based performance features.

The first paper that the author finds most informative and foundational is titled "Nanometre-Scale Electronics with III-V Compound Semiconductors" [128] by Jesús A. del Alamo. In this paper, the author first refers to the incompatibility of the scalability of performance of silicon based logic devices by first noting that MOSFET scaling entered power constrained scaling due to their ultra high scale integration or packing density. Power density in these logic devices cannot increase much further without incurring substantial packaging and cooling costs that make these chips impractical for most applications. Continued progress in transistor power density scaling will require a reduction in operating voltage but this will compromise switching speed. The author noted that this scaling issue is partly why the operating voltage of CMOS devices has bottomed out at 1 V for some time without any further reduction so that switching speed is not severely compromised. This drove the author to postulate that one possible solution is to introduce a new channel material which delivers much higher saturation drift velocity than silicon at room temperature operation. This will simultaneously allow a reduction of voltage without loss in performance, i,e., drive current and switching speed. Plethora of III-V compound materials such as GaAs, AlAs, InAs, GaN, GaSb, InP, and their ternary and quaternary alloys combine elements in columns III and V of the Periodic

Table. Some III-V compound materials have unique optical and electrical properties. Being direct band gap materials, their ability to efficiently emit and detect light means they are often used in lasers, light emitting diodes and photodetectors for optical communication, instrumentation and sensing. A few notably GaAs, InAs, GaN, and InGaAs exhibit outstanding electron transport properties. In fact, there is a large and mature industry manufacturing for III-V integrated circuits in great volumes for applications as diverse as smart phones, cellular base stations, fiber optical systems, wireless local area network, satellite communications, radar, radioastronomy, and defense systems. The recent widespread use of handheld devices and their enormous consumption of data has been a boon to the III-V integrated circuit industry which is now characterized by highly automated and rigorous large scale manufacturing, relatively large area wafers where device performance benefits outperform the cost of these wafers, sophisticated device and circuit design tools, well-established device reliability and a rich and competitive industrial ecosystem. Although 2D dichalcogenides are the recently developed materials that deliver the needed performance features of silicon logic devices, the author of this paper has substantial research findings to demonstrate that no other family of materials currently being considered to replace the silicon channel in a MOSFET has such an impressive list of attributes. Next, the author of this paper focuses on the rationale behind the use of III-V compound materials in logic devices by first referring to the extraordinarily high mobility in these devices even in high lateral drift field. The author noted that experimental evidence confirmed that devices built with InGaAs or InAs show more than 10 times electron mobility compared to silicon at comparable channel sheet density. The outstanding high frequency response of III-V FETs is also frequently invoked. For example, current-gain and power-gain cut off frequencies of InGaAs based HEMTs exceed 600 GHz and 1 THz respectively. Then the author focuses attention on the fundamental operation of a logic switch by first noting that for faster switching, a high on current is desired and to limit standby power consumption off current I_{off} must be minimized. In an NMOS transistor in saturation, I_{on} is determined by the product of the sheet electron concentration and the electron injection velocity V_{inj} at the virtual source which is the location in the channel adjacent to the source contact that provides the highest barrier for injection of carriers from source. This is the bottleneck to electron flow. We can learn about the injection velocity of future III-V transistors by first examining III-V HEMTs. In this regard, HEMTs provide an excellent model system to study issues of importance in future III-V FET devices. The author then provides important information that measurements of V_{inj} for InGaAs and InAs HEMTs have revealed values that approach 4×10^7 cm/s at 0.5 V. It is readily observable from this data that how future voltage scaling lower than 1 V can be sustained by fabricating logic devices in these HEMT devices. Also in devices built as HEMT with III-V materials, V_{inj} is more than twice that of silicon at less than half the voltage. For devices shorter than 50 nm, the injection velocity becomes independent of gate length. Monte Carlo simulations indicate that for these ultra short devices electron transport through the channel is governed by ballistic

AU: Not sure how this highlighted piece makes sense. Please revise.

transport that is almost with no collisions. The author revealed important finding that in these devices V_{inj} is determined by the band structure of the channel material and using stoichiometry such as in InGaAs, increasing the mole fraction of InAs increases V_{inj} due to lower effective mass of InAs. Sheet carrier concentration also affects I_{on} and a reasonable question emerges that since most III-V materials possess low electron effective mass, how can they impart the required high sheet carrier density for sustainable I_{on}. But the recent measurements carried on InGaAs and InAs HEMTs suggested that the electron effective mass in these devices are greater than theoretically reported bulk value This is explained by the presence of strong non-parabolicity in the vicinity of conduction maxima from where these effective masses are extracted. In addition, in HEMTs high degree of electron quantization in thin channel exists and biaxial compressive strain is also encountered. Therefore, the combination of superior V_{inj} and reasonable channel density of states confers that InGaAs and InAs quantum well FETs have the potential to boost the drive current I_{on} substantially higher than CMOS logic devices in silicon in the vicinity of operating voltage in 1 V range. But I_{off} for every technology node is becoming more and more important from power constrained scaling perspective, the author shows that for HEMT devices without source and drain junctions, I_{off} is set by the subthreshold swing S at the temperature of operation which quantifies the sharpness of drop of drain current below a minimum threshold. In InAs and InGaAs HEMTs, the 2D quantum confinement present yields a steep subthreshold behavior competitive with silicon. The thinner the channel, the closest the subthreshold swing to its ideal value 60 mV/decade. However, when the channel material is thinned down, boundary layer scattering tends to degrade the carrier transport. When the thickness of the channel of InAs HEMT is reduced from 10 nm to 5 nm, electron mobility degrades from 13,000 cm^2/V-s to 10,000 cm^2/V-s. However, the injection velocity V_{inj} of short gate length material is affected much less showing ballistic transport is present in these short channel and thin channel HEMTs. A thin quantum well FET structure has the potential to scale to very small dimensions. The author then states that underpinning the phenomenal electrical characteristics of III-V FETs is the heterostructure growth technology with mono-atomic layer precision and an ability to synthesize perfectly specular interfaces. Molecular beam epitaxy (MBE) and increasingly, metal organic vapor phase epitaxy (MOVPE) are at the heart of band gap engineering in IIII-V heterostructures. The author brought a scenario very much interesting to the device researchers that perhaps the most dramatic testament to these technologies is that the achievable electron mobility of the order of 36 million cm^2/V-s obtained at low temperature for AlGaAs/GaAs system. The III-V HEMTs have made the case for III-V CMOS technology. Although HEMT devices possess higher gate leakage current posing barrier for their ready acceptance as logic devices, they contain valuable device features including a junctionless design with a thin undoped InAs rich quantum well that extends under the extrinsic portion of the device on which the raised source and drain junctions are fabricated. Figure 2.1 is an illustrative picture from this reference.

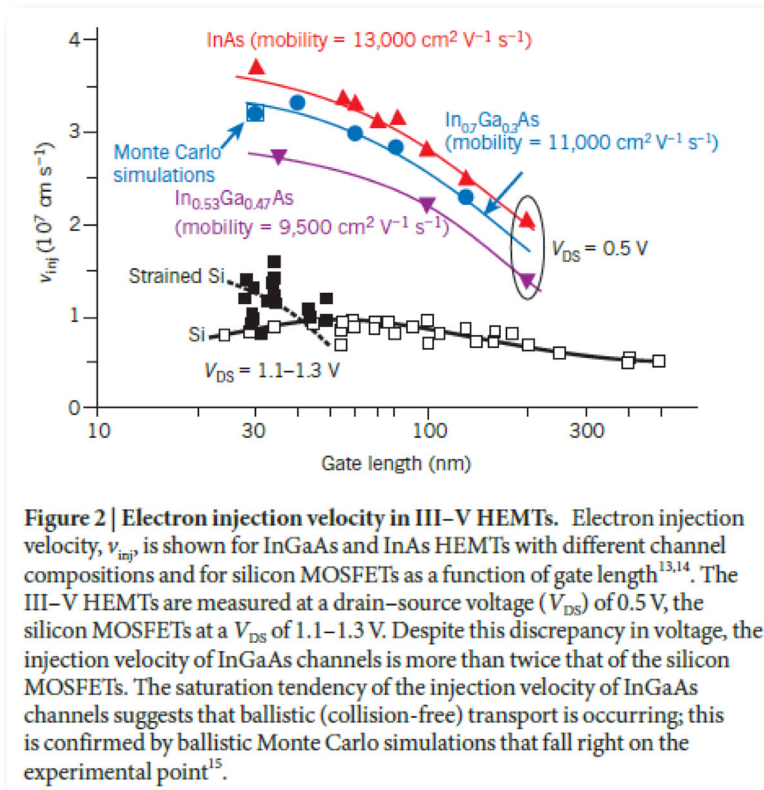


Figure 2 | Electron injection velocity in III–V HEMTs. Electron injection velocity, v_{inj}, is shown for InGaAs and InAs HEMTs with different channel compositions and for silicon MOSFETs as a function of gate length[13,14]. The III–V HEMTs are measured at a drain–source voltage (V_{DS}) of 0.5 V, the silicon MOSFETs at a V_{DS} of 1.1–1.3 V. Despite this discrepancy in voltage, the injection velocity of InGaAs channels is more than twice that of the silicon MOSFETs. The saturation tendency of the injection velocity of InGaAs channels suggests that ballistic (collision-free) transport is occurring; this is confirmed by ballistic Monte Carlo simulations that fall right on the experimental point[15].

Figure 2.1: This picture is taken from article “Nanometre-Scale Electronics With III-V Compound Semiconductors” [128] by Jesús A. del Alamo. doi:10.1038/nature10677, page 318. Copyright permission of this figure is required.

The author is further discussing this classic paper with the contents in a new paragraph as these contents are pretty important where critical issues of III-V materials are addressed before these materials can be completely suitable as alternative to silicon logic devices as the silicon logic devices are still most cost efficient in terms of manufacturing where the SiO_2 is the most native oxide material when grown on silicon device with fewest of defects. From the gate stack perspective, at the heart of a MOSFET is the gate stack. It is composed of a metal gate, a high-k gate dielectric and the semiconductor channel. As silicon already established this property that the device gate stack must have a dielectric free of defects or interface states, a smooth interface with the underlying silicon channel with few interfacial imperfections and high stability. As we know the key for silicon devices to be the mainstream logic devices for decades long time is the existence of native SiO_2 that naturally meets these lowest trap requirements. No such native oxides can be reliably and defect-free state built on III-V materials. In fact, exposure of a III-V surface to oxygen results in Fermi level pinning which is an inability to modulate the electrostatic potential at the channel

interface. This makes it impossible to use in a MOSFET where precise surface potential value may not be extractable due to this inherent problem. In GaAs, the most advanced III-V compound, dielectric oxidation formation creates a rich mixture of Ga and As oxides and sub-oxides, elemental As, As-As dimers and Ga dangling bonds among other defects. Associated with this is a high density of interface states that prevents the effective modulation of surface Fermi level. Because of the difficulty of avoiding surface oxidation as by products rather than maintaining seamless integration and nativity, early attempts to build GaAs FET yielded devices with poor performance and low stability. Later in 1995, Ga_2O_3 deposited on GaAs yielded an interface quality which was quite improved for AlGaAs/GaAs system. This led to fabrication of both n and p-channel GaAs MOSFETs and suggested that dielectric and underlying channel interface with unpinned Fermi level is possible now. A major step forward was taken in 2003 when a GaAs MOSFET with Al_2O_3 gate dielectric was deposited by atomic layer deposition (ALD) technique. The ALD technique is ex-situ, robust, and highly scalable and is widely used in modern semiconductor manufacturing. So a high quality oxide growth options with ALD technique with III-V interface made it more promising for III-V CMOS devices to be realizable. Transmission electron microscopy (TEM) and and X-ray photoelectron spectroscopy (XPS) soon showed that during ALD growth, a kind of clean up effect naturally occurs in which previously existing surface oxides are largely eliminated. This happens at very early stage of ALD and with subsequent exposure to ALD oxidants does not regrow the III-V oxides. Using ALD soon led to MOSFET demonstrations on other III-V compounds such as InGaAs, InAs and InP which show high promising prospects as these compounds are contenders for today's logic devices. Progress in the electrical characteristics of III-V MOSFET accrues from reductions in interface state density D_{it}. As their name suggests, interface states are electronic states that arise from disruption to ideal bonding structure of a semiconductor at its interface with a dielectric. Interface states affect device performance in several ways. Interface states situated closer to conduction band edge increase the subthreshold swing and increase threshold voltage shift whereas those inside the conduction band trap carrier electrons in a temporal way. Both effects reduce I_{on} for a given I_{off}. Interface states additionally distort the surface potential at the dielectric semiconductor interface and cause mobility degradation. There are several ways to deal with these problems. The first is to engineer the interface through predeposition cleaning treatments, the use of interfacial layer underneath the actual dielectric, the deposition chemistry, post deposition treatments or alternative dielectrics that generate less D_{it}. The second involves changing the substrate crystalline orientation of growth direction and transport direction now are decoupled from one another. The most common orientation is (100), but better device quality has been reported on GaAs (111)A surface. This is also the case for InGaAs and InP two other prominent III-V logic device substrates. The third approach is to use compounds containing indium. The device characteristics of the MOSFET improve significantly when the InAs mole fraction in InGaAs channel is increased. InP has also yielded good results. According to DFT (Density

Functional Theory calculations, the interface states associated with group V dimers and group III dangling bonds are predicted to lie well below the conduction band edge. Another important aspect of gate stacking where high k-channel material are involved is the channel mobility. Interface roughness scattering, Coulomb scattering associated with charged interface states and remote soft optical phonon scattering associated with high-k gate dielectric severely degrade the mobility. The author cites that in reality for InGaAs MOS structures with scaled gate stacks at realistic sheet carrier concentrations, a mobility value of 1,000 cm^2/V-s is difficult to obtain. A possible solution to reduce interface roughness related scattering is a buried-channel device where a thin wide band gap material is placed between the channel and high-k gate stack. In addition to interface roughness scattering, Coulomb scattering emanating from charged interface states and bulk oxide traps are also reduced and remote phonon scattering effect is also less pronounced as the carriers are effectively pushed away from the high-k channel interface.

The author then explores various III-V MOSFET device architectures from the perspectives of self-aligned transistor design. The challenge in making small transistor that sustains scaling is twofold. First it is important to maintain adequate electrostatic integrity so the vertical fields do not penetrate the channel region and compromise device performance by DIBL (Drian induced barrier lowering) effect, gate to source barrier reduction and allowing a pathway for direct flow of carrier from source to drain end. This means that the gate exerts a greater degree of electrostatic control over the electron concentration in the channel than the in the drain which calls for a high geometric aspect ratio for the channel if multigate or surrounding gate architecture to be discussed later are not integrated with the device. The second challenge is maintaining low parasitic capacitance and resistances at all transistor contact regions and overlap regions. One of the contribution of parasitic capacitance comes from fringing field out of high-k gate stack and its thickness. Contact resistance and other channel resistance enhancement with scaling can be circumvented with self-aligned device structure. This means that the contacts are placed without requiring an optical alignment to the gate as commonly done in HEMT. Self-alignment results in a gate to contact distance of only a few nanometers. A very low ohmic contact resistance is also required. The author then showed that fundamentally InGaAs should not be at a disadvantage when contact resistance of source and drain are considered. Doping levels of n-type silicon in InGaAs easily reach mid 10^{19} cm^{-3} and combining with electron mobility in excess of 1,000 cm^2/V-s in InGaAs result in a contact resistivity that is still lower than As doped n-type silicon material with doping in $10^{20}/cm^3$ but electron mobility few 10th fraction of InGaAs. The electrostatic integrity in III-V FETs can best be achieved by planar quantum well design which require a very thin channel and extremely thin gate barrier. Higher electrostatic integrity and scaling potential can be realized from nanowire FETs built on III-V substrates. These consist of a long array of thin and short nanowires with the gate wrapped around them. For III-V compounds, horizontal and vertical nanowire FETs with impressive characteristics are demonstrated with InAs system. To improve electrostatic integrity

and scaling length of the device, silicon multigate architectures such as FinFETs and TriGate are also selectable for III-V devices with commensurate level of performance.

The author, after careful and investigative survey on device modeling of silicon and III-V material based logic devices, found that there are shortage of precise modeling computations for incomplete ionization of donors in semiconductors at T = 300 K when the doping is in the range 10^{17}/cm^3 to a few multiples of 10^{18}/cm^3. The efforts by ETH Zurich researchers for the first time extracted analytically computable activation percentage in both n-type and p-type silicon at T = 300 K for donor like Phosphorous and acceptor like Boron. This ionization percentage profile reveals important fact that the usual consideration of 100% activation of dopants are not precise when the dopants introduced are in the range of 10^{17}/cm^3 to a few multiples of 10^{18}/cm^3. For low value of doping the activation percentage starts from 100 % and show a gradual decrease up to 80% for the above doping range and then contrary to accepted equations for ionization percentage computation which hints at monotonically lowered value of ionization percentage, the ionization percentage actually starts rising and reaches to 100% again for doping in the multiples of 10^{19}/cm^3. The other device physical explanation that can be attached to this ionization percentage profile is that as the doping is enhanced in the substrate, individual discrete donor or acceptor states become so close to each other that they start forming indistinguishable band and start to overlap with conduction band leading to the experimental finding that at higher doping instead of less activation, the activation actually stays close to 100% which is also consistent with the approach of source and drain contact resistance reduction in MOSFETs where both source and drain are doped in high degenerate level and only 100% ionization will retain their actual doped value and contribute to contact resistivity reduction. This incomplete ionization effect in the mentioned doping range will be more pronounced in III-V materials FET when compared with Si and Ge MOSFETs owing to the reason that III-V devices such as HEMTs and quantum well structures use lower body or substrate doping generally in the few multiples of 10^{17}/cm^3 range and as a result of which dopants will be not fully activated and slight to moderate percentage reduction of activation will impact the sheet channel carrier density and their screening of charged donors or acceptors impacting Coulomb scattering and vertical field induced interface roughness scattering where the surface band bending is a factor of activated dopants in the channel which is slightly reduced when the doping range stated above is considered for III-V materials. Therefore, when precise incomplete ionization effect is modeled for III-V materials, the sheet carrier density, Coulomb mobility, interface roughness mobility and saturated drift velocity all contributing to device drive current I_{on} will be affected and need to be constantly reevaluated for a selected set of medium range dopant values and this incomplete ionization effect coming out of dopant activation needs to be calculated at lower substrate temperature values and the generally higher reported intrinsic and ballistic mobility values at lower substrate temperatures than 300 K for III-V materials that are reported in literature references ideally still consider the reference doping value at T = 300 K to be fully ionized which

is now proved to be incorrect and a new directive on modeling of incomplete ionization of dopants now is more imperative for precise calculation of device transport parameters such as mobility and drift velocity and other parameters like inversion sheet carrier density and device drive current I_{on}.

Since one dimensional nanowires seem to be on the horizon as the scaling is pushed to sustain Moore's Law, III-V nanowire FETs like InAs or InGaAs may provide the required I_{on} (mA/μm) as stipulated by ITRS roadmap specifically when the gate length is less than 10 nm. Therefore, the author feels to critically analyze and document a highly enriched content-based article on III-V nanowire performance analysis entitled "III-V Nanowires-Extending a Narrower Road." The analysis of the paper is more centered on its device physics-based performance rather than the various nanometer processing approach where vertical nanowires have been experimentally demonstrated. As the author of this paper rightly justifies that semiconductor nanowire MOSFETs are attractive for implementation beyond the 22 nm node as improved electrostatic control within the channel can be achieved by use of wrapped-around gate in cylindrical geometry. Without the wrapped-around gate nanowire assembly as developed and fabricated, the short channel immunity in line with aggressive device scaling can only be achieved with continued reduction of gate oxide thickness and ultra-thin body, both seem to work against control of gate tunneling leakage current and two dimensional quantum confinement effect on threshold voltage shift and also contributes to device channel mobility reduction through confinement related [illegible]odification in the current transport direction. In the high performance devices sub-section of this paper, the author states that for successful nanowire device implementation, it is necessary to consider the details in the device operation and architecture. Four important DC device metrics need to be considered for a FET: (1) the extrinsic transconductance $g_m = \frac{dI_{ds}}{dV_{gs}}$; (2) the on current I_{on}; (3) the off current I_{off}; and (4) intrinsic g[illegible]tance C_{gs}. For nanowires, also the diameters of the wires need to be considered along w[illegible]e area to volume ratio. For a cylindrical shaped nanowire surface area to volume ratio is $= \frac{2\pi r^2 + 2\pi r h}{\pi r^2 h} = \frac{2}{r + h}$, so with reduction in nanowire diameter, nanowire length or height has to be proportionally adjusted imparting larger surface area to volume ratio which is advantageous for larger volume inversion of carrier density at a particular gate voltage. But as the gate length or perimeter in the case of surrounding gate is scaled, the body thickness or diameter needs to be decreased in order to preserve good electrostatic control of the channel to avoid the short channel effects or drain field encroachment toward the source. The author then discusses that for a sufficiently small nanowire diameter, the inter sub-band separation can become large in the order of a few hundred meV. For instance, the author quotes that for an InAs nanowire with diameter 10 nm, the sub-band separation is around 200 meV. The gate length to nanowire diameter relationship can be determined as $L_g = 1.2\,(t_{ox} + t_{NW})$ and for $L_g = 10$ nm and t_{ox} at 1 nm, the nanowire diameter is > 7 nm and for materials with small effective masses, this leads to single sub-band operation as the sub-bands are further apart both for narrower diameter and lower effective mass and this

happens for a gate overdrive value of V_g-V_T = 0. 3 V. It is thus a matter of interest to analyze the device performance in the single sub-band limit. A single mode ballistic 1 D nanowire transistor operating in the quantum capacitance regime has a transconductance $g_m = 2q^2/h$ = 77 µS and drive current $I_{on} = (2q^2/h)\,(V_{gs}\text{-}V_T)$. Normalized to the circumference of the wire, this gives $g_{m,norm} = \frac{77\times10^{-6}}{\pi t_{NW}}$ and for t_{NW} = 10 nm, $g_{m,norm}$ = 2.5 mS/µm, which compares favorably to planar 2D HEMTs. Nanowires with lower diameter will result in higher normalized $g_{m,norm}$. For wires with larger diameter and larger effective masses, several modes will contribute to the transport which can lead towards higher I_{on} and larger transconductance. However, multimode devices will also operate away from the quantum capacitance limit which will reduce the gate coupling. Numerical calculations indicate that InAs multimode ballistic wires will have a transconductance around 2–3 mS/µm. The author systematically arranged the performance parameter values of today's best reported III-V nanowire FETs in a Table 1. From the listed values in the Table 1, it is clear that nanowire FET already shows good drive current I_{on} (mA/µm) possibly best matched with the corresponding ITRS requirement and good device transconductance For a vertical 50 nm diameter InAs nanowire with L_g = 10 nm and 5 nm HfO_2 deposited on n-type InAs substrate, the FEt shows g_m value 0.8 mS/µm coupled with a subthreshold slope of 100 mV/decade, although when ITRS benchmark specifications are targeted, the device engineers focus on designing architectures that best delivers I_{on} (mA/µm) although subthreshold slope may be in the 90 mV/decade to 100 mV/decade range. The total intrinsic gate capacitance of a nanowire FET is given by $C_{gg} = L_g\left(\frac{C_q\,C_{geo}}{C_q + C_{geo}}\right)$ where C_q is the quantum or semiconductor capacitance and C_{geo} is the geometrical oxide capacitance. It is generally found that C_q is << C_{geo} and hence C_{gg} is dominated by C_q. A 5 nm diameter InAs nanowire has a C_q 0.1 aF/nm and the intrinsic gate capacitance C_{gg} will be thus 1 aF when 10 nm gate length is considered. This small value of intrinsic C_{gg} makes a very high f_T achievable for analog and RF device applications However, for nanoscale FETs, parasitic capacitance present in the gate to source, substrate and drain junctions can dominate when the nanowire is scaled and should be incorporated in calculation of C_{gg}. For many applications, it is required to have high drive current and many nanowires may need to be fabricated in parallel with vertical architecture. For vertical nanowires built in III-V substrates, the wire height or length must not be too long when the nanowire radius is reduced to sustain scaling in order that nanowire surface area to volume ratio remains high enough to enable high enough volume inversion, high drive current I_{on} and high gate to channel capacitance C_{gg} or C_{gs}. As we already discussed that for analog and RF high-frequency device applications for enhancement of f_T C_{gg} acts in the quantum capacitance limit or in its lowest limit but in the case of digital applications, we require inversion charge conversion efficiency per incremental increase of gate voltage and for that C_{gg} or C_{gs} of nanowire device should be high.

CHAPTER 3

Device Physics-Based Scaling Insights on Silicon-on-Insulator (SOI)-Based Short Channel n-Type MOSFETs

Silicon-on-insulator device architecture has been developed with isolation of thin silicon film, where the channel is formed, from the substrate by a thick buried oxide (BOX) that entrenches the source and drain junction capacitances to the substrate. As a result of reduced source and drain to substrate capacitance, intrinsic device speed of the device becomes faster which is extremely beneficial not only for silicon logic applications but also for high frequency analog and RF applications. An added advantage of SOI impacted reduction of source and drain to substrate capacitances is that they affect the silicon thin film depletion capacitance reducing it further which directly goes into reduction of subthreshold slope and this benefit of subthreshold slope reduction and short channel control becomes more evident for extremely thin (ET) fully depleted silicon-on-insulator (FDSOI) devices or ultrathin body SOI devices. A relative disadvantage of SOI devices that will be comprehensively represented in this Chapter is the presence of self-heating in ultrathin body full depleted SOI devices where the self-heating affects the reliable performance of the device by impacting reduction and degradation of channel mobility and also introduction of non-linearity in the drain current-drain voltage (I-V) characteristics of the SOI output or transfer curve. Scaling theory of SOI devices with respect to control of ultra-thin body thickness, top and bottom gate oxide thickness and thickness of buried oxide in controlling the 2D scaling length to effectively subdue short channel effects (SCE) will be also discussed in this Chapter.

The author first analyzes in this Chapter the classic review paper "Frontiers of Silicon-on-Insulator" [144] by G. K. Celler and Sorin Cristoloveanu. In this paper, first the authors state that monolithic integration as we all know, has revolutionized electronics and changed the world around us. Concomitant with the rapid progress and evolution of microelectronics to todays giga scale integration, it has become increasingly clear that junction isolation is not the best approach to achieving monolithic integration. These junctions introduce extra capacitance and reduce the achievable density of transistor in the circuits. If the ambient temperature is high enough, leakage currents diminish the goal of isolation between various circuit components. From the time of publication of this paper by the authors found in their survey that in the last 30 years, a growing body

AU: This paper was published in 2003--only 18 years ago. Can you please clarify the timing you are referring to here?

of research and some niche applications had demonstrated that it is possible and often advantageous to build monolithic semiconductor circuits employing dielectric isolation instead of reverse biased p-n junction isolation. This is accomplished by utilizing silicon-on-insulator wafers. Since approximately 1998, commercial applications of SOI devices have grown exponentially and entered the mainstream device applications of ultra large scale integration (ULSI) of logic circuits. As the authors continued their analysis in this paper, SOI structures consist of a film of single crystalline silicon separated by a layer of SiO_2 from the bulk substrate. The authors add very important observation that the top Si film must be monocrystalline, yet needs to be separated from the crystalline substrate by the amorphous dielectric SiO_2 which pose a fabrication challenge. There is no deposition method that would result in a single crystalline film grown without some kind of a template below it. There have been many attempts to utilize various localized templates and then extend epitaxial growth from these templates to other regions but these approaches although possessing scientific interests lacked practical solutions. Out of the broad range of pursuits, two technologies emerged as dominant industrial methods for SOI device fabrication with reliable performance. They both rely on ion implantation but one of the two methods also utilizes wafer bonding. Although ion implantation is essential to both approaches, the implanted species are different and the goal of the implantation is also different. In the first method, known as the separation by implanted oxygen (SIMOX) process, an oxide layer is synthesized directly from oxide ions that become buried under a superficial silicon film. The second approach, known as the Smart Cut™ , utilizes ions, most commonly hydrogen ion implantation as an atomic scalpel that cuts through the crystalline lattice and permits a clean and uniform transfer of a thin layer of silicon to another substrate. Other SOI device-fabrication processes that offer unique advantages for specific device applications include epitaxial lateral overgrowth and zone melting recrystallization—these approaches may permit the building of multiple stacked layers of active devices and contribute to the development of 3D integrated circuits. Currently, performance enhancements have motivated many integrated circuit manufacturing companies to use SOI wafers albeit at a higher wafer price than silicon wafer. As have been explained by the author in the introductory summary of this Chapter, for the same supply voltage, digital logic circuits such as microprocessors run much faster in SOI MOSFETs compared to silicon MOSFETs. Alternatively, it is possible to run the SOI MOSFET with much lowered operating voltages reducing the power consumption while maintaining the clock speed consistent with their widespread usage and adoption for as alternatives to power hungry bulk circuits. As we approach near the end of Roadmap of scaling evolution of silicon MOSFETs, SOI is needed to extend the traditional reliable performance of silicon based devices. As has been noticed, transistors with gate length 25 nm or less do not perform reliably in bulk Si MOSFETs. The 2D scaling theory confirms that at these aggressively scaled operations, the vertical field induced in the channel through the gate has to compete and compromise the deleterious lateral encroachment of fields from source and drain which mainly reduce the gate to source energy barrier at the source

enhancing carrier transport at relatively low supply voltage and subthreshold leakage current value. These short channel effects (SCE) can be reduced by going into thin SOI technology. The authors of this paper then discuss the fabrication t of this book's space limitations, e been developed over the years in detail and because of lack of space for this Chapter in this book, the author omits the analysis of these fabrication methods of Celler and Cristoloveanu, ice physics based scaling impacted performance analysis. In section IV of this paper by the authors, the authors addresses the motivation behind these SOI circuits which are very relevant for analysis by the author. SOI chips consist of millions of single transistor islands dielectrically isolated from each other and from underlying silicon substrate. On one hand, the vertical isolation protects the thin active silicon layer from most parasitic effects produced by the bulky substrate: leakage currents, radiation induced photo currents, and latch up effects. On the other hand, the lateral isolation makes the interdevice separation in SOI circuits free of complicated schemes of trench and well formation. The overall technology and circuit design in this respect for SOI, turns out to be highly simplified and results in more compact VLSI circuits. In SOI built configuration, source and drain junctions extend only up to the buried oxide and there is no path to substrate contact yielding reduction of junction area, lower leakage current and source and drain junction capacitance. This enables the operation of high density logic circuits built in CMOS with lower stand-by and operational power consumption, improved circuit speed and wider temperature range of operation. More innovative devices (multiple gate MOSFETs, power transistors, sensors, MEMS, etc.) can be conceived and combined in SOI that intrinsically is a flexible structure allowing adjustable thickness of film and buried oxide. As the authors importantly further describe in this paper, a second class of advantage is the superior capability of SOI devices to face scalability challenges. The key feature is that unlike the case of bulk silicon, the SOI film thickness poses as an important tunable parameter to aid in device shrinking. It has been found from modeling analysis that the thinner the film, the lower is the drain to body field penetration which causes drain induced barrier lowering (DIBL) effect by additional lowering of gate to source energy barrier for source ended transport. Moreover, the limited extensions of source an drain regions makes SOI less vulnerable to short channel effects originating from charge sharing between gate to other junctions. Back gate biasing as available in SOI-based MOSFET architectures can also be effectively used to control the threshold voltage of the device and for increasing drive current with positive gate bias at the substrate side and decreasing the subthreshold leakage current with negative gate bias at the substrate side. High performance CMOS circuits integrated on SOI compatible with low voltage/low power and higher speed ULSI applications have been frequently demonstrated in deep submicron logic devices. For example, ring oscillators with 50 nm gate length enable a delay time below 10 ps/stage. Frequencies in the range of 150 GHz has also been achieved for SOI MOSFETs. It is in the highly competitive domain of circuits operated with a single cell battery supply in the range 0.5–1 V that SOI can fully express its potential. A small gate voltage swing is suited to switch a transistor from off to on state. CMOS/SOI devices exhibit

superior performance in terms of small gate voltage swing capability enabling low leakage current and quasi ideal subthreshold slope of 60 mV/decade at room temperature 300 K. Hence, in line with supply voltage reduction and lower gate to source voltage swing, threshold voltage can be simultaneously lowered with SOI CMOS circuits in the range of 0. 3 V for a 0.5 V to 0. 6 V supply voltage setting. Complex circuits with a major impact on mainstream microelectronics as realized by SOI FETs include high performance 2 GHz as well as low power 0.5 V 500 MHz microprocessors, Gbit range DRAM and SRAM memories and various RF circuits. Repeated comparisons show that operation at similar gate voltage offers a performance advantage of 20% to 30% when compared to bulk silicon MOSFETs whereas the performance gain nearly doubles when compared to bulk silicon MOSFETs for similar low power dissipation values. Fully depleted CMOS SOI circuits are still operational at temperatures beyond 300°C exemplifying the case of high power electronics applications very attractive for oil, aeronautics, and automobile industry applications. The leakage current is much smaller and temperature sensitivity of threshold voltage is also lower 0. 5 mV/°C when these circuits are operated by SOI compared to silicon bulk. SOI devices can also be tailored to extreme environments able to sustain doses in excess of 10 Mrad for space applications. SOI devices with dynamic threshold (DT) characteristics enable very low-voltage low-power logic applications. The gate and body are interconnected so that increasing the gate voltage in weak inversion simultaneously raises the body potential and causes gradual V_T decrease which is a case of dynamic threshold voltage. This also imparts a positive feedback effect on the drain current. The coupling between gate voltage and inversion charge is excellent and results in improved drive current, subthreshold slope, transconductance and short channel behavior. The DTCMOS in SOI is also suited for ultra low power applications focusing in subthreshold current conduction when the lower gate voltage simultaneously gets coupled to the body and lowers the body potential enhancing the changing V_T and additionally reducing the subthreshold current. In DT SOI n-MOSFET, the substrate being at the same high gate potential drives the back Si film/BOX interface in inversion. As a result of which drain fields are attracted towards the back Si film/BOX inversion sheet density and some of these fields reach the source end underneath the channel in Si film causing gate to source barrier reduction and V_T reduction but DIBL effect becomes prominent through this case. Also the front channel in the Si film and back inversion layer in the back Si film/BOX interface screen or shield the depletion layer impact on subthreshold slope and that is why subthreshold slope is also reduced for DT SOI MOSFET. The authors of this paper then discuss typical mechanisms in SOI transistors where for fully depleted MOSFETs and partially depleted MOSFETs are analyzed. In SOI MOSFETs, two inversion channels can be activated, one at the front Si/SiO_2 interface and the other at the back silicon film-buried oxide (BOX) interface. Full depletion happens when the depletion region fully covers and extends the entire silicon body thickness. The depletion charge remains constant and cannot grow further as the gate voltage is further increased allowing lower threshold voltage or V_T. The excellent coupling between the gate bias and inversion charge

in the channel offers improved current and subthreshold slope. The front and back surface potential remains interrelated. Interface coupling means that the electrical characteristics of channel formed in silicon film is dependent on the gate bias condition of the opposite gate where the back Si/BOX interface can be in accumulation, depletion or inversion. In practice, the front gate measurements may include surface potential effect from BOX and from the BOX/bulk Si interface and highly dependent on back gate bias. The characteristics are complex and totally new $I_{ds}(V_{g1})$ relations controlled by both gate voltages $V_{g1,2}$ apply to fully depleted SOI MOSFETs and the authors provide these experimental transfer characteristics when the device $I_{ds}(V_{g1})$ at strong inversion, log $I_{ds}(V_{g1})$ at the weak inversion and transconductance gm (V_{g1}) are measured and plotted when the back interface is biased in accumulation, depletion and inversion. Each curve can be explained by variation of a dominant parameter. Figure 3.1 is provided below from this reference article covering the following discussion.

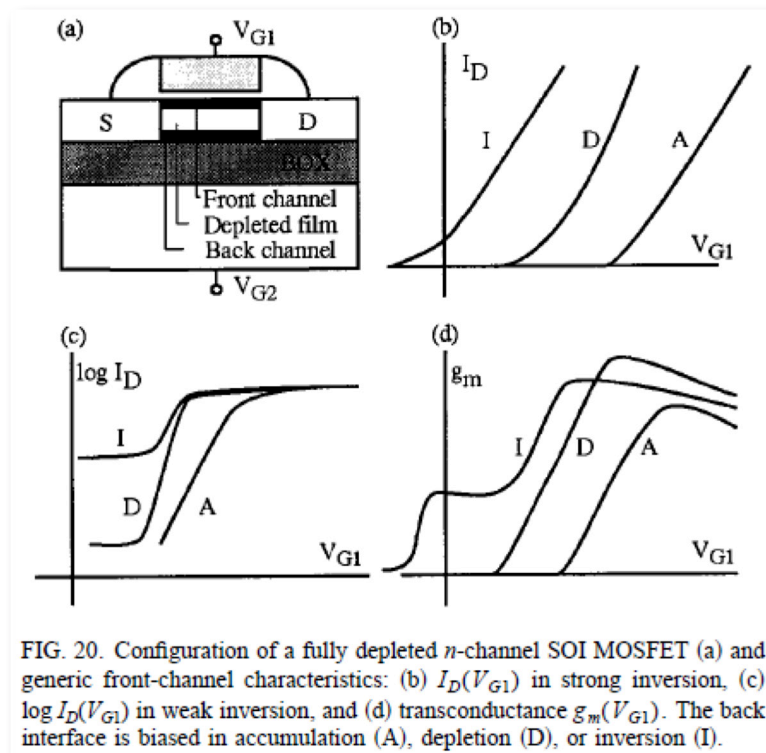


FIG. 20. Configuration of a fully depleted *n*-channel SOI MOSFET (a) and generic front-channel characteristics: (b) $I_D(V_{G1})$ in strong inversion, (c) log $I_D(V_{G1})$ in weak inversion, and (d) transconductance $g_m(V_{G1})$. The back interface is biased in accumulation (A), depletion (D), or inversion (I).

Figure 3.1: The above Figure is taken from the article "Frontiers of Silicon-on-Insulator" [144] by G. K. Celler and Sorin Cristoloveanu, *Journal of Applied Physics*. Copyright permission for this figure is required.

For the $I_{ds}(V_{g1})$ curve, the threshold voltage decreases linearly with increasing V_{G2} between two plateaus corresponding to accumulation and inversion at the back interface. Inversion condition at the back interfaces lowers the vertical field through the front gate and front gate coupling of channel charge and hence lower V_T is suitable. The coupling factor is approximately equal to the thickness ratio of front gate oxide and BOX. For the log $I_{ds}(V_{g1})$ curve, the subthreshold slope reflects the contributions of front and back interface traps. The slope is steepest for depletion condition at the back interface. The ideal value (60 mV/decade) of subthreshold slope when the trap densities are low and the BOX is much thicker than the gate oxide and silicon film. For the $g_m(V_{g1})$ curve, the transconductance hits a plateau when the back channel is activated. The effective mobility and series resistance depend on the back gate bias and back Si film/BOX interface bias conditions. The worst case happens for accumulation at the opposite interface. Not only the vertical field is large enough for front gate to thin silicon film terminations for mobility degradation but also the source/drain extensions can get depleted which increase the access resistance. The qualitative features presented in the above analysis are well captured by conventional 2-interface models. Also available are 3-interface models which match better the modern trend of using thinner film and BOX. For this improved model, the influence of defects located at the bottom of the BOX as well as possible formation of a depletion region underneath the BOX are accounted for. Another feature the authors in this paper in connection to FDSOI MOSFET operation cited is the defect coupling meaning the carriers flowing through the channel at the front gate interface are impacted by the defects present at the BOX or BOX/Si film interface. In practice, an apparent degradation of the front channel properties can be induced by remote damage that is actually located at the back interface or in the buried-oxide. This situation is frequently observed after back interface degradation by radiation or hot carrier injection. Finally, another conspicuous feature of FDSOI devices is the presence of self-heating conveyed by the low thermal conductivity of the BOX and mostly happens in the silicon film. Self-heating in FDSOI devices causes mobility degradation and drive current reduction and initiates negative output conductance in the saturation region. Self-heating raises the carrier temperature from the lattice or substrate temperature and higher carrier temperature not only causes mobility reduction but also the threshold voltage, saturation drift velocity, bipolar gain and interconnect temperature are dependent on this carrier to substrate temperature mismatch. As the silicon film thickness becomes thinner, self-heating becomes more prominent—a viable case for FDSOI MOSFET devices. The channel or carrier temperature is also raised with increasing BOX thickness and channel to contact separation. The author will separately analyze the self-heating in SOI devices from other reference sources. Fortunately, self-heating is highly reduced under low voltage or dynamic operation. The authors of this paper then goes on to discuss the partially depleted SOI MOSFETs (PDSOI) where the thin film is much thicker so that depletion region does not extend all the way through the film thickness and a neutral point or potential exists near the bottom of the film thickness. Interface coupling effects are cancelled and instead a floating body

effect due to the neutral potential is observed. The floating body effect produces a kink in the I_{ds}(-V_{ds}) curve and the kink is triggered by majority carriers generated by impact ionization which get collected at the neutral region. The body potential is raised at the neutral point and threshold voltage is subsequently lowered. The kink effect is materialized in excess current and low frequency noise in saturation. In weak inversion and for high drain bias, a similar positive feedback is responsible for negative resistance regions, hysteresis in log I_{ds}-V_{gs1} curve and eventually for transistor latch. The floating body may also induce transient variations of body potential, threshold voltage and current. When the gate voltage is switched on, majority carriers are expelled from the depletion region (instantly formed by capacitance coupling) and get collected at the neutral body potential resulting in drain current overshoot. The overshoot decreases over time as the accumulated excess carriers at the neutral point recombine with thermally generated electrons in the depletion region. A reciprocal undershoot when the gate is switched from strong to weak inversion. The undershoot gradually disappears as the current is increased with time and excess majority carriers are generated in the depletion region causing the depletion depth to shrink. As a general rule, the amplitude or current overshoot or undershoot is proportional to the difference between the final and initial body charges and the transient duration depends on the generation-recombination rate in the film volume, at interfaces and on the edges. Although carrier lifetime has remarkably been improved, in state-of-the-art MOSFETs the transient time is dramatically reduced for several reasons: (i) in short channel, the source and drain junctions effectively remove the majority carriers; (ii) in narrow channels, the edges which are more defective play a dominant role for majority carrier recombination; and (iii) in ultrathin oxides, the gate tunneling current becomes large enough to compensate for body charge. Transient effects are of particular concern at high drain bias, where impact ionization becomes an additional source of majority carriers. During high-frequency switching of integrated circuits built in PDSOI, the transistor body does not always reach the equilibrium. Designers know how to use the extra overshoot current for improving the circuit speed. However, the charging and discharging of the body is an iterative process that may cause history and memory effects as well as dynamic instabilities. The solution of alleviating floating body effects is to design the body contact at the expense of an increased die size. In ultrathin films with large sheet resistance, the body contact is far from being ideal (intrinsic resistance, potential barrier, excess noise, etc.). That is why body contact is provided to only a selected set of PDSOI circuits that play a critical role in the circuit. The authors in this paper then discusses another important section on new directions in SOI devices where they shed insights on: (i) short channel effects (SCE); (ii) scaling trends; (iii) possible high-performance device architectures like (1) ultimately small MOSFET and (2) double-gate MOSFETs. In both fully and partially depleted MOSFETs with submicron length, the lateral bipolar transistor through source to body to drain coupling can easily be turned on. The basic mechanism is the rise of body potential by impact ionization which causes a forward bias of source to body junction. The activation of parasitic bipolar effect has advantage in

the form of extra drain current or detrimental effect causing premature breakdown, latch-up which is a condition reflected by inordinate rise in drain current with very low drain to source voltage. Parasitic bipolar effects are enhanced in n-type channels, shorter devices, thinner and wider silicon film and at higher temperatures. The reliability of short-channel MOSFETs is affected by hot carrier injection into the oxide. The degradation mechanisms are very complex in SOI where two oxides and two channels are involved. In n channels, most defects are created at the interface where the electrons flow. Exceptionally when the transistor is biased in the breakdown region, injection into the opposite interface occurs and causes defect coupling. The device aging is accelerated when the gate voltage is low and the back Si/BOX interface is accumulated. In p-channels, electrons generated by front channel impact ionization become trapped in the buried oxide. An apparent degradation of the front interface again happens through coupling. A surprising short channel effect in SOI transistors is the metamorphosis of partial depletion into full depletion. The lateral depletion width governed by source and drain junctions not only covers a large portion of the body but also reduces the effective doping of the body, thereby enabling full depletion by the gate voltage. In addition, the lateral profile of back interface potential can be highly inhomogeneous: from depletion in the middle of the channel to weak inversion near the channel ends. This localized weak inversion region explains the degradation of the swing. More familiar short channel effects resulting in threshold voltage V_{th} roll off are charge sharing between front gate and terminals and drain induced barrier lowering (DIBL) a 2D effect where the reverse biased drain fields encroach towards the source terminal and reduce gate to source barrier enhancing current transport which is specially a problem in subthreshold and weak inversion conduction. An extra DIBL effect is the lateral penetration of drain fields into the BOX and underlying substrate. The fringing fields terminated on the BOX increase the potential at the thin film-BOX interface, very much as if the back gate is positively biased and this effect is termed by the authors in this paper as drain induced virtual substrate biasing (DIVSB). Due to DIVSB and interface coupling, both the threshold voltage and subthreshold slope are lowered. For the scaling trends insightful discussion in this paper, the authors noted that the scaling of bulk silicon MOSFET requires a reduction of junction thickness or ultrashallow junction and an increase of body doping such that source and drain depletion regions are isolated and do not merge. The need for ultrashallow junction to control threshold voltage V_{th} roll off unfortunately raises the junction capacitance and high enough body doping reduces channel mobility. According to the authors' projection, doping related difficulties will probably terminate the scaling of bulk silicon or PDSOI MOSFETs at about 35–50 nm gate length. Fortunately, the scaling rules and design rules are more relaxed for fully depleted SOI transistor (FDSOI) because additional tunable parameters like silicon film thickness, front gate oxide thickness, buried oxide (BOX) thickness, possibility of lower doping for full depletion and front and back gate biasing, are all available for device optimization. The minimum channel length that can be envisioned is given by L = 3 λ where λ is a 2D scaling length essential for short channel effect control. λ is solved using

Poisson's equation from 2D potential contours emanating from source and drain at a selected gate and drain voltage. Several expressions have been proposed:

$$\lambda FD = \sqrt{\left(0.5\,\frac{\varepsilon_{si}}{\varepsilon_{ox}} t_{si}\, t_{ox}\right)} \text{ or } \lambda_{FD} = \frac{\left(t_{si} + \frac{\varepsilon_{si}}{\varepsilon_{ox}}\, t_{ox}\right)}{\pi}$$

and both these equations deliver a clear message that 2D scaling length is much reduced which makes SCE better by using ultra thin silicon film in conjunction with scaling induced reduction of front gate oxide whereas for this condition of operation, the doping effect becomes irrelevant. Actually, below a critical film thickness 15 nm for 80 nm long MOSFETs, no doping is needed at all. This is extremely beneficial in terms of carrier mobility but implies use of metal gate to set the device threshold voltage. The impact of fringing fields (DIVSB) by using a thin and low-k buried oxide or BOX at the expense of increased parasitic capacitance, larger influence of the depletion region underneath the BOX and enhanced interface coupling effects. Moreover, a ground plane located below the BOX is able to suppress the field penetration and depletion region in the substrate. DIBL is also significantly improved by a ground plane insertion. Combining 50 nm BOX with a ground plane (GP) allows relaxed film thickness 20–30 nm which match the capability of current SOI processes. A GP can be achieved by ion implantation underneath the BOX or by bonding two wafers, one of which has a highly doped surface region or a metal overlay. The optimized architecture of sub 50 nm long FDSOI MOSFETs will combine several ingredients: very thin silicon film with low doping, moderately thin or lower-k BOX, mid-gap work function enabled gate for low V_T and use of ground plane (GP). The authors project using the second expression of λ_{FD} that FDSOI will reach its limit with 20 nm gate length and 5 nm body thickness. Elevated source and drain will be required to decrease the series resistance in the ultrathin films of FDSOI MOSFETs. A crucial challenge for wafer suppliers is to provide soon enough ultrathin films with excellent quality and uniformity. Otherwise, thickness fluctuation will be responsible for intolerable SCE induced by DIBL and field penetration. The depletion charge variation out of variation in film thickness will not affect the device performance or SCE effect. It is worth noting that a small geometry device also implies a narrow channel. The parasitic currents flowing on the sidewalls of the transistor depend on the isolation technology, local defects, stress, doping segregation and lifetime degradation. Floating body effects tend to vanish whereas new coupling effects are triggered. A strong interdependence develops between the three dimensions of the transistor, i.e., the scaling of the length and width is made easier when the body thickness is simultaneously reduced. When analyzing ultimately small MOSFETs, the authors noted that the absolute minimum physical thickness of a SOI transistor is one atomic plane. However, the transport properties of a monolayer of silicon will be very different than crystalline silicon. Experimental results exist for 1 nm thick MOSFET where four monolayers of silicon do maintain MOS like functionality. Thickness related effects on the carrier transport, interface coupling and electrostatics become considerable and re-

quire full quantum mechanical treatment. A film thinner than 10 nm behaves as a vertical quantum well, the parameters of which can be modulated by front and back gate voltages. When the film thickness is below 10 nm, the carrier and energy confinement by 2D quantization effect are significant leading to an increase of threshold voltage even though the depletion charge is reduced. The authors projected that minimum channel width found is 1 nm. Such a transistor is a quantum wire where the carriers confinement develops in the lateral direction leading to an increase of threshold voltage. Imagining the case of an extremely miniaturized transistor, the study associated with short channel effect so far gets transferred to minimum volume effect. Considering a volume of dimensions 10 nm × 10 nm × 10 nm and $10^{18}/cm^3$ body doping gives a doping number of just 1. A doping level of $10^{16}/cm^3$ will give a doping number of 0.01 as if there is no detectable doping particle for carrier transport in this extremely miniaturized transistor. Therefore, when dealing with a single or fractional dopant, defect or trap, macroscopic notions such as interface traps, doping concentrations are useless. Finally the authors in this paper [illegible]s another very potential device architecture built with FDSOI that is double gate or DG. Double gate SOI transistors have, in principle, two symmetrical gates interconnected or a unique gate that surrounds the body or gate-all-around GAA configuration. DG MOSFETs can be planar, vertical or mixed mode (vertical film with side gates and horizontal transport). Fabrication is achieved with delta process, epitaxial lateral overgrowth, wafer bonding, Fin process, tunnel epitaxy, SON (silicon-on-nothing) etc. In the GAA process, a suspended Si membrane is formed by tunnel epitaxy, by etching a cavity in the BOX or by growing a stack of sacrificial layers. The membrane is thus oxidized and the gate is deposited around. The DG concept was initially demonstrated in 1987 on conventional SOI MOSFETs by simultaneously biasing the front and back gate of an SOI transistor. The formation of front and back inversion channels causes, by continuity, volume inversion in thin SOI film. Thus the carriers responsible for current flow in the middle of the film and experience less surface roughness scattering, hence the channel mobility, transconductance, drive current and 1/f noise all are improved. The two gates exert ideal control on the potential and inversion charge constituting a superior gate integrity so that major short channel effects (SCE), for instance, charge sharing, DIBL, fringing fields and punchthrough. But DG MOSFET built on SOI is different from conventional Si MOSFET with the difference being the back gate creates channel in the Si film through the BOX thickness instead of much thinner second gate oxide of conventional silicon DG MOSFET. So for DG FDSOI MOSFET, the general voltage requirement on the back gate will be larger to cause wider uniform channel thickness in the film with the goal being volume inversion of the film. Also for efficient DG SOI MOSFET performance, BOX thickness should be reduced and the substrate thickness beneath should be reduced with higher substrate doping. This will reduce the substrate resistance underneath the BOX so that back gate potential drop through the substrate is minimal and also thin BOX will enable sufficient inversion charge density at the back Si film/BOX interface possible with a slightly higher back gate voltage compared to the front gate voltage. Since the in-

trinsic scaling length parameter λ_{FD} is very much reduced in DG FDSOI MOSFET compared to single gate FDSOI FET, they are now considered the final metamorphosis of the MOS transistor especially beyond the 10 nm barrier. Fin technology has produced already 15–20 nm long DG MOSFET in SOI with promising characteristics. Numerical simulations including quantum effects, band-to-band tunneling and direct source-to-drain tunneling tend to prove that transistors with 2 nm gate length still perform appreciably. Keeping in mind that the body thickness should be roughly 1/3 of its length, the authors thus voice in this paper that the ultimate frontier of a DG MOSFET gate length is 1 nm which would mean the use of one monolayer of silicon. Measurements in a 3 nm thick transistor show a large increase of transconductance in the DG mode, which can be explained by the much uniform carrier distribution and higher carrier mobility avoiding surface scattering. Since the potential profile is symmetrical, the vertical films cancel in the middle of the film. The self-consistent solution of Poisson and Schrödinger's wave equation reinforces the volume inversion concept indicating most carriers flow far from the interfaces. However, the total inversion charge in DG SOI MOSFET is twice the single gate MOSFET. The gain in transconductance is beyond 200%. Only in DG MOSFETs can the carriers avoid the rough interface regions. These empirical arguments have been validated by rigorous Monte Carlo simulations that reveal an increase of channel mobility in DG MOSFET with 3–5 nm silicon film thickness. In the asymmetrical configuration of the DG MOSFET, two different work functions of front and back gate provide enough leverage in tuning the threshold voltage of the device. In addition, a longer back gate and overlapped source and drain extensions can also be beneficial. The back gate biasing induces not only inversion channel but also accumulation in the extensions (electrical junctions). This field effect management of series resistance allows larger drive current and greater transconductance while tolerating slight misalignment of the two gates.

3.1 SELF-HEATING IN FDSOI MOSFETS, ITS ORIGIN AND IMPACT ON DEVICE PERFORMANCE

The author considers device physical attributes and analysis of self heating in FDSOI MOSFETs are essential from long term device reliability perspective. In this context, the author will analyze the classic paper entitled "Analysis of Self-Heating Effects in Ultrathin Body SOI MOSFETs by Device Simulation" by Anthony G. O'Neill et al. The adoption of SiGe source on thick buried oxide (BOX) in silicon wafers significantly impacts the thermal properties of the device as these materials possess a much lower thermal conductivity compared with the bulk silicon. Ultrathin body (UTB) FDSOI MOSFETs with silicon film thickness on the order of few tens of nanometers present a degraded thermal conductivity due to phonon boundary scattering that reduces the phonon mean free path substantially. In their paper, the authors apply 2D electrothermal device simulation, with a transport model tuned to the results of Monte Carlo (MC) transport, to investigate the effect of

self heating (SHE) on UTB SOI MOSFET both for digital and analog circuits taken into consideration. The authors explain heat generation in short devices by first noting that heat generation in MOSFET through the emission of phonons by carriers heated by electric field in the channel region. Since the electric field is maximum at the drain end of the channel and source-to-drain distance is comparable to the mean free path for phonon scattering, each carrier suffers a few scattering events or no scattering in the case of quasi ballistic transport. Hence most of the heating effects take place inside the drain region. More accurate calculations of phonon emission performed by Monte Carlo (MC) transport simulation predict a much broader heat generation region with a lower peak value displaced inside the drain junction. Nonlocal effects on carrier heating and phonon emission become more relevant as the device channel length approaches the mean free path for phonon emission. The precise quantification of the impact of local-heating approximation is still an open issue because the analysis lacks a detailed treatment of short- and long-range electron-electron interactions that influences the carrier heating and the subsequent cooling process or thermalization happening at the drain region. In spite of its inherent local approximation, the conventional model for Joule heating is still largely adopted in consideration of the ease of implementation in the framework of device simulators, allowing an efficient electrothermal simulation by the self-consistent coupling of carrier transport, heat generation and heat transport thus providing the possibility to investigate the impact of technology options on device performance including the SHE effects. The authors further explain in this paper that when studying heat transport in nanoscale devices with dimensions of the order of or smaller than the phonon mean free path, sub-continuum transport effects occur as follows.

1. Hot-spot ballistic phonon-emission effect: heat generation due to optical phonon emission by heated electrons, taking place in a region of limited extension (hot spot) compared with the phonon mean free path, leads to higher temperature rise because a significant change in temperature may occur only over a distance comparable or larger than phonon mean free path.

2. Hot-spot far from equilibrium effects: SHE occurs mainly through the emission of optical phonons characterized by low propagation velocity. On the other hand, heat transport involves much faster acoustic phonon modes. Due to the larger rate of optical phonon emission in the drain channel end and the difference between propagation velocity of optical and acoustic phonons, a localized out-of-equilibrium condition is realized in which the ratio of optical phonon concentration to acoustic phonon concentration is much larger than at equilibrium, leading to a reduced heat spreading capability.

3. Phonon boundary scattering in thin films: thermal conductivity in thin films is substantially reduced with respect to bulk crystals due to the enhanced scattering of phonons at the film boundaries causing a large reduction of phonon mean free path.

This also reduces the effects of hot-spot ballistic phonon emission and making the limitations of (1) less critical. In the important section analysis of SHE in the UTB SOI MOSFETs, the authors report the results of self consistent electrothermal simulations of the SOI MOSFETs and highlights the main issue related to SHE in these devices. The dependence of SHE effects on the temperature parameters is analyzed in order to point out a need for device design explicitly taking into account SHE. First the authors show that device thermal resistance R_{TH} increases as the device is changed from PDSOI to UTB FDSOI with low doped channel and T_{si} = 6 nm. This is largely due to the phonon boundary scattering present in these ultrathin layers. By neglecting the cooling effects through the drain, source and gate interconnects, lead to a significant increase of device thermal resistance. The SHE is larger in SOI MOSFETs compared with bulk silicon MOSFETs as the latter one is effectively cooled, owing to the high thermal conductivity path along the silicon bulk toward the ideal sink placed at the substrate contact. R_{TH} is also affected by UTB SOI gate length. Shorter gate length translates into a more compact device with smaller cross section of the plane at the silicon/BOX interface and therefore less efficient cooling from the buried oxide. R_{TH} is also dependent on BOX thickness Increasing the BOX thickness results in lower thermal conductivity of SiO_2 more manifest increasing R_{TH} . The authors then venture into digital and analog circuit impact of SHE by first noting that digital circuits require tight device packing whereas analog devices require multi-finger layout where adjacent devices/fingers do not thermally interact due to the lateral heat spreading, leading to an enhanced SHE in analog UTB FDSOI configurations. When plotted, device thermal resistance R_{TH} shows as a function of number of fingers for PD and FD 25 nm SOI MOSFETs and different packing densities, a rapid increase with the number of fingers in the layout. The impact of multi-finger layout becomes larger for larger packing density or smaller spacing. The UTB SOI MOSFET is not sensitive to multifinger layout impact on R_{TH} if the spacing is larger than 200 nm. Because of degraded thermal conductivity of UTB SOI, the thermal healing length is shorter than W_{spac} = 200 nm. When the spacing is reduced below the healing length, the thermal interactions between adjacent fingers become very significant due to the much higher carrier heating that takes place in these ultrathin devices. Devices of digital circuits undergo successive switching transients contributing to dynamic power dissipation interleaved by off-state period contributing to static power dissipation and device heating due to the gate tunneling and subthreshold leakage current. Under SHE device heating is delayed and hence contributes to lower power dissipation compared to DC I_{ON} only. In the case of dynamic power dissipation for FDSOI MOSFETs unless enough cooling through the contact's interconnect occurs, SHE induced temperature rise can be as high as 90°C compared to 27°C ambient temperature pose reliability

concerns. In fact, both the lifetime of metallization and time to breakdown (TBD) of the gate dielectric are degraded at large temperatures, with the temperature dependence of TBD is becoming more prominent as the thickness of the gate dielectric is reduced to nanometer regime. For SHE effect on drain current, ON current degradation is mostly in the 10% range and would not seriously harm the performance of logic circuits although compared to bulk silicon MOSFET, I_{ON} degradation is on a larger scale when SHE effect is taken into account for UTB FDSOI MOSFETs. The impact of SHE on thermally activated O_{FF} current is significant. The temperature rise subsequently results in larger I_{off} leading to an increased static power consumption.

CHAPTER 4

Device Physics-Based Scaling Insights on Strained Si-Based Short Channel n-Type MOSFETs

By alloying with other group IV materials such as Ge by forming Si:Ge source, due to Si and Ge's lattice mismatch, large uniaxial compressive strain or biaxial tensile strain can be induced at the source to channel point. This strain operates on the conduction band band structure by changing the concavity near the conduction band minima and also for valence band maxima. As a result, conductivity effective mass of electron in n-MOSFET and hole in p-MOSFET are reduced leading to higher carrier transport from source to channel and this feature can be utilized specially for p-MOSFET for which group IV and III-V materials generally exhibit very low mobility values. For using strain in the source point or channel, the inversion charge density remains unaltered negligible change in V_T, device subthreshold slope is also unaffected and off current or subthreshold leakage current is also unaffected due to the reason that I_{off} increases in proportional to intrinsic carrier concentration n_i^2 and n_i is modified by $(m_n mp)^{0.5}$ and since strain reduces both m_n and mp, a very minute reduction of I_{off} can be expected with negligible impact as the I_{off} is already in the 10–13 A range for nanoscale MOSFETs. Strain's enhancement of electron and hole mobility can best be understood by first noting that when a tensile strain is applied, it splits the sixfold degeneracy of silicon conduction band and lowers the twofold degeneracy perpendicular Δ valleys with respect to the fourfold in-plane Δ valleys in energy space where the Δ_2 valley effective mass is lower than Δ_4 valley effective mass (in-plane transverse effective mass $m_t = 0.19\ m_o$ and out-of-plane longitudinal mass $m_l = 0.98\ m_o$. Such energy splitting suppresses inter-valley carrier scattering between the Δ_2 valleys and Δ_4 valleys. Reduced intervalley scattering extends the time between collision and simultaneous reduction of effective mass in Δ_2 valley population enhance electron mobility by strain, considering $\mu_n = \frac{q\tau}{m_n^*}$. Similarly, strain splits the valence band degeneracy at the Γ-point and shifts the spin orbit band improving the in-plane hole mobility. In the study of strain's effect on device performance, mostly mobility impact through effective mass alterations are stressed although strain can also effect band gap of a material due to changes in conduction band maxima point and valence band minima point although the change is very minute compared to overall band gap value. Strain's impact on V_T mainly comes from bulk potential which is inversely proportional to n_i enclosed in a logarithmic term. So, even though n_i will decrease due to reduced effective mass of both electron

and hole under strain where the change in band gap is assumed minimal, bulk potential will increase very minutely and overall V_T will have a negligible positive shift. The gain in drive current therefore can lead to larger I_{on}/I_{off} ratio. But it has been observed that high value of strain decreases the band gap. As a result of which n_i will increase due to exponential factor involving band gap. Increase of n_i for highly strained devices will result in decrease of bulk potential and lowering of V_T. Increase of n_i will also increase subthreshold leakage current I_{off}. So with the improvement of drive current with higher strain applied to the device, the off-state current can also increase substantially. Observing the advantage of strain to boost device current, the author now discusses some relevant articles that discuss the strain, its origin and overall device physics for n channel MOSFETs.

In their classic paper entitled "Strain: A Solution for Higher Carrier Mobility in Nanoscale MOSFETs" [154] by Scott E. Thompson et al., the authors start with scaling impact on conventional MOSFETs by noting that as device dimensions shrink into deep submicron regime, many physical phenomena such as short channel effects (SCE), velocity saturation, high leakage current and dielectric breakdown limit the benefits of conventional scaling. To continuously improve device performance, new device architectures, new channel and substrate materials and strain engineering in the source and channel of a MOSFET have been proposed and investigated. Among all these new technologies, strain engineering during the past decade has been the dominant technology to enhance device performance while providing a low cost and low risk technique by maintaining the traditional MOSFET fabrication processes. The authors provide a very important fact in this paper that it has been confirmed both experimentally and theoretically that strain has the potential to enable drive current improvement of ~ 4.5 × in Si p-MOSFETs and ~ 2 × in Si n-MOSFETs without a significant increase in leakage current. Figure 3.2 is an illustrative picture from this reference article.

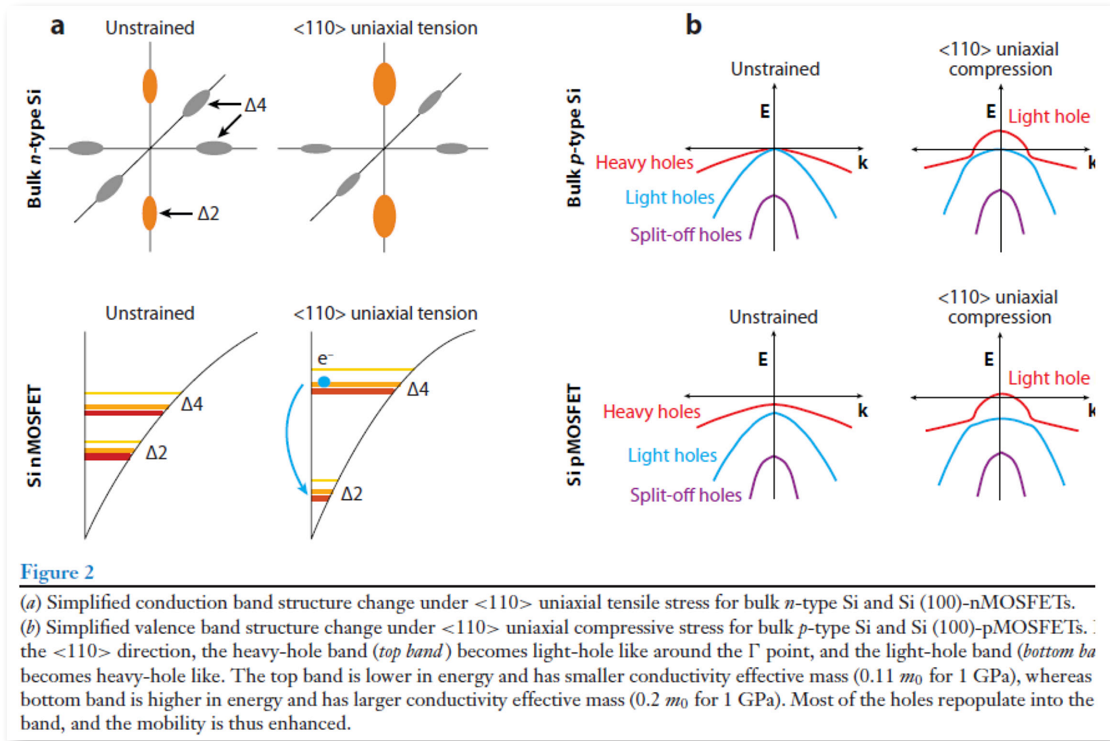


Figure 2

(*a*) Simplified conduction band structure change under <110> uniaxial tensile stress for bulk *n*-type Si and Si (100)-nMOSFETs. (*b*) Simplified valence band structure change under <110> uniaxial compressive stress for bulk *p*-type Si and Si (100)-pMOSFETs. the <110> direction, the heavy-hole band (*top band*) becomes light-hole like around the Γ point, and the light-hole band (*bottom ba* becomes heavy-hole like. The top band is lower in energy and has smaller conductivity effective mass (0.11 m_0 for 1 GPa), whereas bottom band is higher in energy and has larger conductivity effective mass (0.2 m_0 for 1 GPa). Most of the holes repopulate into the band, and the mobility is thus enhanced.

Figure 4.1: This Figure is taken from the article “Strain: A Solution for Higher Carrier Mobility in Nanoscale MOSFETs” [154] by Scott E. Thompson et al. Copyright permission is required for this Figure.

With the fourth generation of strained-Si technology now in commercial production, strain enhanced performance and power saving are present in nearly in all VLSI logic chips manufactured today. The authors introduced the concept of deformation potential theory which models the coupling between acoustic waves and electrons in solids. In deformation potential theory, the strain-induced band edge shift is proportional to the strain tensor. The electron mobility change can be ascribed to “electron transfer” and altered intervalley scattering rate caused by valley energy shift. The picture of strain-induced hole mobility change is more complicated, owing to the strong valence band warping and presence of light hole mass, heavy hole mass and split-off band in the valence band structure for silicon. Thus, the hole transport under strain cannot be simply explained by band edge shift. Band calculation method such as k.p method give more accurate valence band structure by constructing a strain Hamiltonian in terms of the angular momentum derived by symmetry consideration. k.p. method with the above inherent advantage has been used by researchers to systematically study the valence band effective masses and deformation potentials in strained silicon. This study revealed the key factors that affect the hole mobility in semiconductors—band splitting

and warping, mass change and consequently the change in density of states (DOS) which alter band occupation and phonon scattering. The authors continued that the most effective empirical method to predict device behavior under strain is by measuring piezoresistance coefficients (π coefficients). Researchers have used this technique to investigate the drive current improvement under strain. In 1968, Colman et al. measured by the π coefficients in p-type inversion layer for the firs time. Two decades later, the first silicon n- and p-MOSFETs with biaxial stress induced by a $Si_{1-x}Ge_x$ buffer layer were demonstrated by Wesler et al. in 1992 and Nayak et al. in 1993 respectively. A 2.2 times improvement of electron mobility and 1.5 times improvement in holy mobility were reported. In 2005 Lee et al. published the review of the history and progress in high mobility biaxially strained Si, SIGe and Ge channel MOSFETs. Even though the predominant focus of the industries in the 1980s and 1990s was on biaxially stresses devices, the current focus has been shifted to uniaxial stress. Uniaxial stress has several advantages over biaxial stress, such as larger mobility enhancements and a smaller shift in threshold voltage. Starting at the 90 nm technology node, uniaxial stress has been successfully integrated into the mainstream MOSFET process flow to improve device performance. Encouraged by the strain enhanced planar MOSFETs, researchers recently applied uniaxial stress to multi-gate devices with metal gate and high k dielectric as performance booster. The authors provide important concepts that to obtain insights into the underlying physics of strain-enhanced devices, it is necessary to first understand strain, of which the effect on current drivability is quantified as a π-coefficient. Strain is defined as the percentage change of material's lattice constant. Strain can result from phonon induced lattice vibrations, lattice-mismatched film growth and applied external mechanical stress. Advantageous strain reduces crystal symmetry, thus lifting band degeneracy and causing band warping. Any strain can be decomposed into a hydrostatic strain and two types of shear strain. One type of shear strain is related to the change of lengths along the three axes, the other is related to the rotation of the axes of an infinitesimal cube. For cubic crystals such as Si and Ge, hydrostatic strain does not break the crystal symmetry and hence, only shifts energy levels without lifting band degeneracy. Thus this component of strain is not important for carrier mobility improvement. Large hydrostatic strain is undesirable owing to band gap narrowing, strain relaxation and MOSFET threshold voltage shifts. It is the shear component of stress that causes sub-band splitting and affects semiconductor transport property. Strain is introduced into the device channel preferably by applying uniaxial stress. The uniaxial stress is longitudinal when is applied parallel to the channel and transverse when is applied perpendicular to the channel. Large magnitudes of uniaxial stress ~ 1 Gpa are being incorporated into the p-channel devices of 65 nm node and an even higher stress is applied in the 32 nm technology node as is evident from the significantly large saturation current 1.55 mA/μm for N-MOSFETs and 1.21 mA/μm for p-MOSFETs. However, because of many process parameters undergo modifications when fabricating strained MOSFETs, there is some uncertainty about whether strain alone is responsible for device performance improvement, for example, reduced external resistance

may also be the cause. The authors in this paper provide important information that the π-coefficient gives straightforward experimental information about strain-enhanced carrier mobility in semiconductors. This coefficient is defined as the normalized change of resistivity with stress $\pi = \frac{\Delta\rho}{\rho\sigma}$ where σ is the applied stress and ρ is the resistivity, which can be calculated as $\rho = \frac{1}{qn\mu_n + qp\mu_p}$. For MOSFETs under steady state, the electron and hole densities are approximately constant, this the π-coefficients are mostly determined by changes in carrier mobility with stress. The π-coefficients give us a straightforward idea about how much drive current enhancement b achieved under a particular stress and it has therefore widely been used in industry to predict strained device performance. The authors then discuss in detail a very needful section on insights into the physics of strained classical Si MOSFETs. The mobiliy enhancement effect in strained-Si devices can be divided into two parts: (1) the reduction in average conductivity effective mass caused by carrier repopulation and band warping; and (2) suppression of intervalley scattering rate by sub-band splitting and change in density of states (DOS). For a bulk silicon conduction band, there are six degenerate valleys with the minimum located near the X point. The applied external stress shifts and splits these sub-bands. For example, the <110> longitudinal tensile stress causes the energy of the Δ2 sub band to shift down and Δ4 subband to shift up, resulting electrons repopulating from Δ4 valley to Δ2 valley. Because the conductivity effective mass m^*_o in Δ2 valley is 0.19 m^*_o is smaller than 0.315 m^*_o in Δ4 valley, the repopulation initiated in this way causes the effective mass to decrease and carrier mobility to increase. Additionally, band splitting causes the scattering rate to change. The dominant scattering mechanisms in strained Si devices are intervalley phonon scattering and surface roughness scattering at moderate to high vertical and lateral fields. As the sixfold conduction band splits, the intervalley scattering rate becomes lower owing to the smaller DOS thus resulting in higher carrier mobility. Strained n-MOSFETs have different mobility enhancement factors than those of bulk silicon. Because of the electric field confinement, Δ2 valley and Δ4 valley are nondegenerated even for unstrained Si. The energy splitting between these valleys depends on the magnitude of electric field and the difference in their out-of-plane confinement effective mass. For example, for the case of (100) surface unstrained silicon n-MOSFET under a high electric field, electrons predominantly occupy the lower energy Δ2 valley. Thus the effective mass change in strained silicon nMOSFET is smaller than bulk silicon. Furthermore, surface roughness scattering dominates under typical high vertical fields through gate with ultrathin gate oxide and this contributes to the 2D carrier transport which makes the current transport in n-MOSFETs more difficult to predictably model. As a result, the electron mobility enhancement in strained silicon n-MOSFETs is often significantly different than that of the bulk Si. Also the bulk Si mobility values are derived from density of states (DOS) mass whereas for strained silicon n-MOSFET conductivity effective mass with lower value is used. P-channel MOSFETs are important part of CMOS logic devices and enhancing mobility and drive current of p-MOSFETs should be also the

focus of device engineers but due to the inherent nature of valence band in silicon, effective mass enhancement of hole has been stalled for most device engineering methods over the years that have been successfully applied for n-channel MOSFET mobility improvements. Therefore, in this paper the authors address this issue and discuss the strain,s band structural effects for p-MOSFETs. For unstrained bulk Si, the valence band minima is located at the Γ point where the heavy hole and light hole mass energy band introduce degeneracy. The spin-orbit split-off band is located 44 meV below from valence band maxima point and is thus not important for hole transport. In unstressed silicon, 80% of the hole carriers occupy the heavy hole band which has a conductivity effective mass of 0.59 m_0 along the <110> direction. Under <110> uniaxial compression, which can be theoretically shown to provide the largest enhancement, the degeneracy is lifted and band warping occurs. At room temperature, band warping induced effective mass reduction is the dominant factor for mobility enhancement in p-MOSFETs under uniaxial stress (< 1 G Pa). At this stress, the splitting between top band and bottom band is small compared with Si optical phonon energy 61.3 meV. Thus the phonon scattering effect is negligible. For unstrained p-MOSFETs, however, the degeneracy of the heavy hole and light hole is lifted by the surface electric field confinement. The splitting between the heavy hole mass and light hole mass energies increase with an increasing gate field. With typical device operation where the surface electric field reaches 1 MV/cm, the summation of confinement induced band splitting and strain induced band splitting can grow and be larger than optical phonon energy. Therefore, for high vertical and lateral fields, the optical phonon scattering reduction under strain becomes significant and must be considered so that the mobility values for strained p-MOSFETs are not overestimated. The band edges also shift under applied external stress. The strain induced band edge shift can be either additive or subtractive to the confinement induced splitting and therefore band gap of the material is also altered. For conduction bands, both biaxial tensile stress and <110> uniaxial tensile stress cause splitting that is additive to the (100) confinement induced splitting. For valence bands, the splitting due to <110> uniaxial compression is additive to the (100) confinement effect and biaxial strain is subtractive. If the stress induced splitting is subtractive to confinement induced splitting, the mobility enhancement for strained n-MOSFETs is lower than bulk silicon. Thus, for MOSFETs operating under a high electric field to benefit from the strain, it is critical that the chosen strain produce splitting that is additive to the confinement effect. For need of space of this book, the author excludes some of the other discussion in this paper but mentions the section summary for speeded up realization by the readers. Strain induced carrier mobility improvement for n-MOSFETs and p-MOSFETs with different surface orientations and channel directions show that (100) surface <110> uniaxial stress is the most effective for performance benefits of both n and p-channel MOSFETs. In the following section summary, the authors list for strain effects on mobility for non classical MOSFETs first for UTB SOI devices. Carrier mobility in UTB SOI devices has a similar enhancement under low stress when compared to bulk silicon devices and a larger enhancement in carrier mobility under high stress for

UTB SOI devices under high stress results larger phonon scattering rate reduction owing to the extra subband splitting caused by geometrical confinement. Wide FinFETs have a similar mobility enhancement as bulk silicon devices with same surface orientation whereas narrow FinFETs have a larger enhancement. This is because of the strong subband modulation caused by geometry confinement increasing the number of carriers that can be affected by strain. Ge and Si_{1-x} Ge_x p-MOSFETs show the same strained behavior as the silicon MOSFETs but the enhancement saturates under higher stress. n-type GaAs MOSFETs benefit from biaxial stress owing to the monotonic increase in splitting between the Γ and L valleys under biaxial stress. As with the k.p. method, p-type GaAs MOSFET has similar enhancement factor under <100> uniaxial compression to that of silicon channel MOSFETs. Finally, the authors provide important projections and analysis on future generations of strain technology. The industry has adopted uniaxial strain over biaxial strain because of the former's larger performance improvement. Uniaxial stress introduced via process by capping layers and epitaxial SiGe source drain regions. Lately, tensile strained nMOSFETs with [illegible]e source and drain regions have attracted interests. The lattice mismatch between SiC [illegible]s in longitudinal tensile strain and vertical compressive strain inside the channel. Both types of strain enhance electron mobility. Also, the offset between the conduction band of SiC source and the strained channel results in higher carrier injection velocity. However, as device dimensions successively reduce, the volume of Si/Ge stressor in the source and drain also reduces, making it difficult to maintain high level strain in short channel devices. To date, strain has been an effective performance booster even for the 22 nm technology. However, the effectiveness of strain as a performance booster beyond the 22 nm node remains to be seen. Theoretical proves that strained Si technology provides performance improvements even for ballistic channel MOSFETs. Strain induced reduction in carrier mass results in increasing source carrier injection velocity and thus leads to drive current enhancements in ballistic transistors. This ensures the scalability of strain technology, but the real technology challenge will be to effectively incorporate high level stress in these nano devices.

The author next presents critical analysis of another classic paper titled "High Hole and Electron Mobilities Using Strained Si/Strained Ge Heterostructures" [162] by E. A. Fitzgerald et al. As we have seen during the evolution of scaling to sustain device performance, it was relatively easier and more feasible to enhance the electron mobilities of n-type MOSFETs but similar engineering approach employed for p-type MOSFETs rarely have increased the hole mobility and this is detrimental from the point of observation of requirement of symmetrical transfer curve for a CMOS and reducing its noise margins. As the authors of this paper state in the introductory section of the paper that strained layers of Si and mole-fraction-engineered SiGe offer dramatic mobility improvement over bulk silicon. Relaxed SiGe buffers on Si wafers with low defect densities are ideal platform for including these layers. Additionally, the technologies for processing these structures are similar to standard silicon CMOS processing. N-channel MOSFETs using strained Si grown on

(100) relaxed Si_{1-x} Ge_x have been fabricated and studied by many researchers and electron mobility enhancement of 1.7–2 times over bulk silicon has been observed. Important variables in determining the enhancements are the strained silicon thickness and the degree of strain in the Si layer. The strain levels in the above mentioned heterostructures are from 0.8% to 1.4% corresponding to a Ge fraction of 20% to 35% in the Si_{1-x} Ge_x virtual substrate. Bi-axial tensile strain breaks the six fold degeneracy of Si's conduction band resulting in reduced intervalley scattering in strained Si n-MOSFETs and enhanced electron mobility. Additionally, for in plane transport, electrons show low transverse electron mass $m_t = 0.19\ m_o$ resulting in further enhancement. A minimum strain of 0.8% corresponding to Si grown on relaxed $Si_{0.8}Ge_{0.2}$ is required for sufficient conduction band splitting to completely suppress intervalley scattering and from an enhancement peak, little gain in electron mobility can be deduced by further increasing the strain. The enhancements in p-MOSFET devices are more striking and vary with gate overdrive and strain. Biaxial tensile strain also splits the heavy hole and light hole band degeneracy though the rate of subband splitting in valence band is seen to be lower than in conduction band. The striking factor is unlike n-MOSFETs or strain on electrons, both the out-of-plane and in-plane conductivity effective mass for hole are reduced by strain. Similar to n-MOSFETs, the enhancement in hole mobility can degrade for higher value of strain. With a strain lower than 0.8%, the strained silicon p-MOSFET lose its mobility enhancement at strong gate overdrive and in some cases mobility drops down than bulk Si case. At low vertical fields, the hole function tends to be weighted below the surface due to the band offset which acts as a barrier to the holes between the relaxed Si_{1-x} Ge_x and the strained Si. The magnitude of valence band offset ΔEV increases with increasing mole fraction of Ge. As gate overdrive is increased, the wavefunction can overcome the barrier of the offset and shift towards the surface. The gradual influence of strained Si on valence band results in increasing hole mobility with E_{eff} up to 0.35 MV/cm to 0.5 MV/cm, where a mobility peak is typically observed. It also shows that for low gate voltage or gate voltage overdrive, i.e., for subthreshold or weak inversion condition hole mobility improvement will be very low irrespective of the magnitude of strain applied. With higher Ge content in the buffer layer, the barrier to hole occupation in the strained Si cap increases and a stronger vertical field is needed to shift the hole wavefunction upto the surface. Beyond 0.5 MV/cm, the mobility in strained silicon p-MOSFETs slowly decreases. A biaxial tensile strain greater than 1% in Si can cause the out-of-plane mobilities higher than in-plane mobilities This leads to a hole wavefunction that is substantially spread in the vertical direction. Experimental results show the extent of hole wavefunction shift is more than 5 nm which implies the hole wave function samples the Si_{1-x} Ge_x buffered p-MOSFET structure below the strained silicon. Under this situation, even at high vertical fields, the properties of holes in the relaxed SiGe below the strained silicon can still make a significant contribution to the hole transport. To improve the hole transport properties and mobilities, the authors in this paper proposed the dual channel heterostructure PMOSFET where a compressively strained $Si_{1-y}Ge_y$ layer is grown upon a relaxed Si_{1-x} Ge_x buffer layer $(y > x)$

with capped by tensile strained silicon. The introduction of compressive strain in the $Si_{1-y}Ge_y$ layer reduces the intervalley scattering due to breaking of valence band degeneracy and also reduces the out-of-plane and in-plane effective masses analogous to the tensile strain in the buffer SiGe layer. The resulting band structure creates a quantum well for the holes in the strained $Si_{1-y}Ge_y$ structure. The valence band offset which confines the holes in the compressively strained is also advantageous since these holes have an intrinsically higher hole mobility than Ge rich $Si_{1-y}Ge_y$ compressively strained layer. The higher hole mobility in strained $Si_{1-y}Ge_y$ occurs because the band structure crosses from Si like X valley to Ge like L valley for a Ge mole fraction of 0.8 and thus strained $Si_{1-y}Ge_y$ exhibits near bulk Ge like hole mobilities which is 1,900 cm^2/V-s. Increasing strained Ge thickness gives rise a 80% hole mobility enhancement grown on $Si_{0.5}Ge_{0.5}$. Thus experimental evidence shows that a large enough strained Ge thickness is required for obtaining highest hole mobilities in these dual channel heterostructures. The thickness should be large enough to accommodate the vertical extent of hole wavefunction which is determined by the amount of strain in the Ge channel. The authors provide more elaboration on this strained Ge thickness effect on hole mobility that the hole wavefunction penetrates the underlying buffer layer as well as top capped Si layer. The penetration of hole wavefunction in the relaxed $Si_{1-x}Ge_x$ layer below is not desirable because it provides a very low hole mobility in the channel and reduces the overall hole mobility enhancement. Thus, it is expected that higher hole mobility can be achieved if the hole wavefunction penetration in the underlying buffer layer can be reduced so that holes are confined in the strained Ge layer. It has been determined experimentally that a thin Si cap of 3 nm would provide the highest hole mobility since the hole wavefunction will be thwarted by the Si top capping layer and instead will be confined to remain in the strained Ge layer. The authors in this paper next puts forth the idea of dual channel heterostructure NMOS where for CMOS applications similar advantages for electron transport properties and mobilities can be deduced for this structure. The band structure of the dual channel heterostructure provides a quantum well for electrons in the surface. A thick layer of top Si capping layer should cause all the electron wave function to remain in top strained silicon layer and not penetrate the buried channel. Unlike p-MOSFETs where Ge rich compressive layer reduced intervalley scattering for holes, electrons residing near the strained Ge-strained silicon interface may undergo strong X valley to L valley scattering owing to the fact that Ge conduction band minima is toward <111> direction and Si conduction band minima is toward <100> direction. Electron mobility degradation in buried strained Ge layer is much higher compared to buried $Si_{0.2}Ge_{0.8}$ layer. The only way to obtain higher electron mobility in dual channel heterostructure NMOS is to completely isolate the electron wave function from the buried compressive layer. This is made possible by increasing the Si cap thickness to 6–7 nm.

4.1 IMPORTANT OBSERVATIONS ON THE IMPRECISE PRACTICE OF QUOTING EFFECTIVE MASSES FOR ELECTRONS AND HOLES IN SI AT T = 300 K

The author considers it highly appropriate about the inaccurate values used for density-of-states effective masses for electrons and holes for silicon material consideration that go into total device modeling based transport analysis such as mobility calculation and saturated drift velocity calculation. Most authors have used the $m_l = 0.98\ m_o$ and $m_t = 0.19\ m_o$ with 6 ellipsoidal surfaces for Si and refer these m_l and m_t values for T = 300 K which gives $m_n = 1.08\ m_o$. Now Robert F. Pierret in his book *Advanced Semiconductor Fundamentals* ([2] 2nd edition) has pinpointed this error in accurate effective mass calculation for electrons at T = 300 K. Measurements for Si list the m_l value as 0.916 and m_t value as 0.19 only at T = 4 K. So for T = 300 K or for any other substrate or lattice temperature, proper modeling equations must be developed for m_l and m_t for silicon as a function of temperature and should be validated by measurements. Robert F. Pierret using the m_l and m_t value at T = 4 K with six ellipsoidal surfaces in Si, computed the mn or effective mass of electron at T = 4 K to be $1.062\ m_o$. In another book by Pierret titled *Semiconductor Device Fundamentals* using the effective mass m_n dependence on substrate or lattice temperature by the equation $\frac{mn}{mo} = 1.028 + (6.11 \times 10\text{-}^{4})T - (3.09 \times 10^{-7})T^2$, the silicon effective electron mass at T = 300 K is 1.18 mo and this is also can be verified at T = 4 K to be $1.03\ m_o$ although this equation is not valid at such low temperature. Also from mn = $1.18\ m_o$ at T = 300 K for Si as per the above equation cannot generate the ml and mt values for Si at T = 300 K as there are two unknown parameters m_l and m_t linked with one equation $m_n = 6^{2/3}\ (m_l\ mt_2)^{1/3}$ where m_l and m_t values must be known for all temperatures in silicon and the current practice of using Δ_2 valley in plane effective mass 0.19 m_0 and out of plane Δ_4 valley effective mass 0.98 m_o for tensile strain applied to n-MOSFET actually are not precise values at T = 300 K but have the resemblance from T = 4 K values, so these values will alter as proper modeling equation for m_l and m_t as a function of substrate or lattice temperature is established and this will make the calculations discussed in this chapter from various referenced articles for electron and consequently hole mobilities due to strain more accurately represented. Therefore, there is a need to modify the above equation for mn as a function of temperature down to 4 K so that correct $1.062\ m_o$ value can be acquired and for Si and Ge materials, the exact dependence of ml and mt values as a function of temperature and specially for T = 300 K needs to be developed at least for silicon where applying strain, the mobility values are impacted by mt and ml and so far, the reference articles report erroneous mobility values by application of strain in Si for both electrons and holes at T = 300 K since m_l and m_t values are incorrectly used from T = 4 K reference with some difference in m_l value of $0.916\ m_o$ reported by Pierret and $0.98\ m_o$ reported by most referenced articles on strain's impact on n- and p- MOSFETs. This difference is also making the reported values for electron mobility under application of strain imprecise and need

AU/ COMP: Which version of 'mo' is right?

recalibration. Unfortunately, almost all the authors in referenced articles have overlooked these key requirements of accurate effective mass calculations for electrons and holes as a function of temperatures and separately for ml and mt dependence on temperature, specially, for very low temperatures where high enough mobility is generally found for electrons. One relatively time-feasible way to extract the effective mobility of electron in silicon as a function of temperature is to follow the following procedure. First, the band diagram of conduction band for silicon has to be constructed at some five or six temperature values ranging from 80–500 K by using empirical pseudopotential method or more accurate k.p method. Then near the conduction ban profile at the conduction band minima E_c, a parabolic fit needs to be superimposed on this profile in the small vicinity around E_c. Then this approximated parabola can be represented as $E\text{-}Ec = C_1 k^2$ and knowing the E value at a particular k value on the parabolic curve in the vicinity of E_c, C_1 coefficient can be extracted. Now the effective mass of electron at that particular temperature or silicon can be readily extracted by this equation, $\frac{2\,C_1}{\hbar^2} = \frac{1}{m_n}$. Following this procedure some sample effective mass for electron data at some sample substrate or lattice temperature can be extracted. Then a polynomial fitting curve with fitting coefficients can be deduced so that the polynomial curve is within 5% of the actual m_n versus T curve derived from band structure. This procedure can be reliably followed for other materials such as Ge, GaAs and other III-V materials and a trend of effective mass for electrons and holes as a function of substrate or lattice temperature can be extracted for these materials and now can be more precisely used for carrier transport based analysis in these device materials. The lattice/substrate temperature dependence of conductivity effective mass for electrons and holes are also need to be developed in analytical equation form to address the current transport in thin channels near the surface of Si/SiO_2 of n- and p-MOSFETs. Conductivity effective mass is used in computing conductivity of thin inversion channels in surface and additionally electron and hole mobility from Drude's mobility model and drift velocity in low and high lateral field. The conductivity effective mass m_{nc} of electron reported by most reference articles for Si is 0.26 m_o at T = 300 K using m_l = 0.98 m_o and mt = 0.19 mo and the equation $\frac{3}{m_{cn}} = \frac{2}{m_t} + \frac{1}{m_l}$. But the m_{cn} value calculated at T = 300 K in this way is erroneous as the m_l and m_t values used for Si for electron actually are from T = 4 K reference from Pierret's textbook mentioned earlier where the author stress, the correct ml value at T = 4 K 0.916 m_o must be practiced while calculating either density of states effective mass or conductivity effective mass of electron of Si material. Using Pierret's referred m_l = 0.916 m_o, the conductivity effective mass of electron at T = 4 K is m_{cn} = 0.258 m_o which is not that different than 0.26 m_o value used in reference articles. But this m_{cn} value is at T = 4 K and for T = 300 K, m_l and m_t for electrons in Si need to be computed and then used in the conductivity of effective mass calculation equation mentioned above. Besides, m_{cn} dependence of T should be established from 4 K to around 700 K. For instance, when ballistic n-MOSFET nanowire is used, electron's conductivity effective mass m_{cn} will decrease substantially and since there is almost no collision with

dopants, the scattering time is also very extended making a very high mobility value achievable with high on current I_{on}. So, we can readily see the importance of developing analytical equations for m_l and m_t versus T for electrons in Si as well as developing two other analytical equations of density of states effective mass of electrons mn versus T for Si and conductivity effective mass of electrons in Si m_{cn} versus T. The above analysis is equivalent for calculations of hole mobility dependence on substrate temperature for both density of states effective mass and conductivity effective mass following information given in Pierret's first referenced textbook in this sub-section. According to Drude's mobility model the other parameter that considerably influences channel mobility of n- and p-MOSFETs is scattering time or mean time between collisions. At low lateral fields, Coulomb or ionized impurity scattering and phonon vibrations related scattering are important. At high lateral fields, high vertical fields and high substrate doping, the phonon scattering is still important but more important is interface roughness scattering which reduces scattering time at a faster rate when ultrathin gate dielectric is used for n- and p-MOSFETs. So, along with effective mass for electrons and holes dependence on substrate temperature, we have to also determine some form of analytical modeling equations for temperature dependence of various scattering events, first as an individual component and then applying Matthiesen's rule to get the effective scattering time and then replace this value in Drude's mobility equation at that particular temperature. Authors in the literature have identified these scattering processes for strain's impact particularly at T = 300 K but ranging all T such that 4–700 K, these principal scattering events and corresponding scattering time are not precisely calculated although intensive Monte Carlo simulations have been performed. This reveals another shortcoming from modeling perspective of establishing temperature dependent analytical equations of various scattering processes accurately. The author following the polynomial fitting of derivation of $\frac{m_n}{m_o}$ as a function of temperature of R. F. Pierret as given in earlier pages, developed another suitable polynomial fitting equation for temperature dependence of $\frac{m_n}{m_o}$ that extends down to 4 K and reveals $m_n = 1.062\ m_o$. Pierret's both the equations for electron and hole effective masses as a function of temperature are not ideally precise at temperatures below T = 200 K. The equation derived by author of this book for electron effective mass in Si as a function of temperature are well within 3% deviation of Pierret referred equation for $\frac{m_n}{m_o}$ for lower ranges of T values below 200 K. For low temperature values below 100 K and down to 4 K, conduction band E-k diagram configuration for electrons in Si and consequently effective mass extraction in the vicinity of conduction band minima requires more simulation processing time, hence the mn values extracted at these extremely low substrate temperature will show more variation in analytically computed equation developed and experimental measurement procedure needs to be developed to extract mn at these low T values which have not been done by researchers and not listed in textbooks. The modified and developed equation by the author is given here, $\frac{m_n}{m_o} = 1.061 + (4.552 \times 10^{-4})T - (1.527 \times 10^{-7})T^2$. This equation can be used to determine values of mn of electron in silicon at low values of T close to 4 K as well as with precision comparable to

Pierret's equation up to higher temperatures up to 900 K. The fact that this newly developed equation computes m_n at very low T down to 4 K, can be suitably used to compute mobility in this range which will be very high for ballistic MOSFET with no scattering at all. We should carefully note that despite the requirements of proper modeling of m_n, m_l, and m_t of electron effective mass, longitudinal mass and transverse mass, respectively, with respect to substrate/lattice temperature from very low T such as 4 K to very high T such as 700 K, over this range m_n varies only 1.304 m_o–1.062 m_o = 0.242 m_o but the relatively large gain in transport mobility comes from extremely low scattering time in the vicinity of 4 K when ballistic transport is possible. So in the same temperature range 4–700 K, scattering time may vary from few 10^{-11} s to a fraction of 10^{-13} s. Therefore, identifying key scattering events are important for conventional MOSFET, wide band gap III-V MOSFET, ultrathin body FDSOI MOSFET, strained channel n-MOSFET and multiple and surrounding gate MOSFET. Researchers when computing scattering time stressed upon ionized impurity scattering, phonon scattering, intervalley scattering or optical phonon scattering and surface roughness scattering. One of the key scattering event neutral impurity scattering has not been conclusively modeled by researchers while determining overall scattering time and mobility of n-channel MOSFET as a function of temperature. For degenerately doped n-channel MOSFET, between $10^{17}/cm^3$ to few $10^{18}/cm^3$, there is incomplete ionization of dopants even at T = 300 K and some number of donor atoms are then neutral and they will also impact carrier transit and scattering which need to be properly assessed. At lower substrate temperatures down to 4 K, neutral impurity scattering inclusion becomes must as there are more incomplete ionization of donors as temperature reduces and a significant number of donors remain neutral considering 4 K operation. Neutral impurities not being charged, do not create scattering cross section over which electrons have to transport with their thermal energy, so electrons retain their velocity and momentum as there is no scattering through neutral impurities through attractive or repulsive force-field as the impurities are charge neutral. So, neutral impurity scattering may be beneficial at lower substrate temperature as the number of collisions with charged donors are reduced and mobility may be further enhanced although after careful observation of scattering time due to neutral impurity scattering event has been accurately modeled. Let us picture the scattering from ionized impurity and neutral impurity of acceptors. When the acceptors are ionized, they contain extra electron per impurity which repel the electron transport over it, so here repulsive scattering is dominant. When these acceptor dopants are neutral, they contain a vacant hole per acceptor and now carriers see an attractive force by the hole where scattering may result in carrier trapping or carrier loss by these vacant holes of neutral acceptors in the depletion region. Ideally, the reverse electric field of the depletion region should oppose the electron trapping by vacant hole of a neutral carrier, So neutral impurities mostly provide direct attractive scattering through their vacant holes and carrier loss through trapping is a rare event paving the path to extending scattering free transport as they are not charged dopants. Both processes will operate here but due to reverse bias electric field in the depletion region, elec-

tron trapping is insignificant and mostly forward scattering by the vacant state of neutral impurities impact carrier mobility at very low substrate temperatures at subthreshold and deep subthreshold operation of an n-channel MOSFET. Some researchers at T = 300 K have incorrectly used the concept of incomplete ionization for substrate doping density $10^{19}/cm^3$ or higher and hence applied neutral impurity scattering to determine mobility in these highly degenerate substrates. But as has been mentioned in previous discussion that after $10^{19}/cm^3$ substrate doping, the ionization percentage rapidly reaches to 100% as doping becomes more degenerate and therefore, using neutral impurity scattering for these degenerate doping values will generate inaccurate value for either surface mobility or bulk mobility. Also, the nature of distribution of incompletely ionized dopants with dopant activation energies needs to be looked upon as temperature goes to 4 K. If a larger sequence of donor states remains unionized at 4 K, it will reduce collisions significantly during carrier transport extending scattering time and spatial evolution of mobility being higher but at the same time carriers may be trapped in the holes associated with neutral impurities leading to attractive scattering and carrier loss but the reverse bias electric field present in the depletion region should resist electron occupation by these vacant states of neutral acceptors but due to attractive force induced forward scattering at a neutral acceptor site, overall extension of scattering time may still happen at these low temperatures when substantial neutral impurities are present in the depletion region of an n-channel MOSFET. But as the temperature is reduced from 300 K but not so low, the placement of these unionized dopants can be random amongst the donor states making the gain in scattering time minimal and hence minimal increment in spatially evolution of transport mobility due to randomly happening placement of neutral dopants.

CHAPTER 5

Device Physics-Based Scaling Insights on Multiple Gate and Surrounding Gate nm Dimension n-Type MOSFETs

Multiple gate and surrounding gate MOSFETs increase the gate coupling to channel by increasing the channel potential close to the surface and ensure a smoother distribution of inversion channel charge and its thickness from source to drain. With the multiple gate architecture, need for boosting the channel potential by increasing the depletion capacitance in relation to oxide capacitance by increasing the substrate doping concentration is very much relaxed or obviated as with near intrinsic doping value of the substrate, when multiple or surrounding gate architecture is selected, the channel potential is boosted and gate integrity is enhanced as the gates are now only at the top or bottom of the substrate but also are fully surrounding it, so the overall depletion capacitance C_d is enhanced and higher level of inversion channel charge at lower gate voltage is now possible. Intrinsic doping capability with multiple gate architecture also frees the device from various short channel effects that come out of higher doping in the substrate. For all multiple gate architectures, the inversion channel profile is rather roundabout and volume inversion of the substrate away from the oxide-semiconductor interface is possible. This has several beneficial impacts. The roundabout nature volume inversion of the substrate in the middle screens the underlying ionized impurity based scattering, so Coulomb scattering is decreased and also since this inversion charge channel lies away from the surface, surface roughness scattering and boundary layer scattering effects are also reduced. As a result of these scattering reduction, the carrier mobility is expected to be enhanced and total drive current is also enhanced. Therefore, the author has selected this Chapter to analyze some of the classic reference papers on multiple and surrounding gate architectures of n-channel MOSFET and probe with device physics based attributes and implications.

The first reference paper being analyzed in this context focused on multiple gate architecture deployment of n-channel MOSFET towards the end of Moore's Law-based device architectural evolution, the authors are Isabelle Ferain, Cynthia A. Colinge and Jean-Pierre Colinge and the paper is titled "Multigate transistors as the Future of Classical Metal-Oxide-Semiconductor Field-Effect Transistors" [175]. As the authors opine in this reference paper that in a classical bulk MOSFET, the gate electrode is positioned on the top of an insulator or oxide that covers the chan-

nel region of the device between the source and drain. In such a configuration, the device is planar and essentially two dimensional. Electrostatic control of the channel by the gate is achieved through the capacitive coupling between the gate and channel through the ultrathin gate oxide. The scaling law requires a reduction of source and drain junction depth profiling reduction by the same scaling factor with which the channel length is continuously reduced. This reduces short channel effect by rendering the subdued control of the channel by the source and drain field encroachment toward the channel. Decreasing the thickness of the gate insulator oxide, thus ultrathin gate oxide yields a similar result, by improving the capacitive coupling between gate and channel. In addition, replacing silicon dioxide SiO_2 with other silicate based high-k gate dielectric possessing higher dielectric constant are capable of increasing the gate coupling to the channel and enhance the channel charge density. The gate's electrostatic control of the channel can also be improved by modifying the shape of the MOSFET. The electrostatics of a long channel MOSFET are essentially one dimensional. Early textbooks on the physics of semiconductor devices used "gradual channel approximation" which solves the one dimensional Poisson's equation-the equation that governs the connectivity between surface electric field and surface charges vertically from the gate of the MOSFET through the channel down to the substrate. Short channel effects in which electric fields emanate from source and drain junctions and encroach towards the channel laterally or horizontally create a second dimension to the problem. Multigate MOSFETs introducing gate height, i.e., a third dimension, to counteract short channel effects, mainly the loss of gate to channel integrity by the two-dimensional source and drain field penetration into the channel effects. The notion multigate is more synonymous to surrounding gate which means the gate is now wrapped around several sides of the channel region. Over the years the industry manufacturing of multigate MOSFETs have resulted in a number of device architectures such as, fin field effect transistors (Fin-FET MOSFETs), triple-gate (Tri-gate MOSFETs), gate all around MOSFETs with cylindrical architecture in which the gate surrounds all edges of the substrate and Π-gate and Ω-gate structures and these two are so named because of the shape of their gate electrode structure. The authors also noted that the best practice of using multiple gate devices is on silicon-on-insulator (SOI) substrate. Multigate FETs can also be made with bulk silicon wafers instead of an SOI substrate. To fabricate such device, silicon fins are etched on a bulk silicon wafer and field oxide is deposited to to avoid the formation of an inversion channel between the fins. Ion implantation is then used to introduce the desired doping profile to the channel and the gate stack is deposited. In this architecture, the gate is wrapped around the top and edges of the silicon fin, creating a tri-gate structure and the source and drain are formed by ion implantation. Bulk tri-gate devices with a fin width as short as 10 nm have been shown to demonstrate good short channel immunity down to sub 20 nm gate length regime. SOI multigate FETs are easier to fabricate than their bulk counterparts because of the inherent device isolation provided by the buried-oxide layer. Bulk silicon substrates, by contrast, are more readily available in large quantities making the large volume manufacturing much simpler. The

authors then extended the conformability of multigate n-channel FETs by opining that apart from their aspect ratio based gate structural modification, multigate MOSFETs in function are similar to conventional MOSFETs. In particular, "technology boosters" techniques that are used in standard CMOS based device architectures, can also be applied to multigate FET architectures in a relatively straightforward manner. The most common technology boosters are the use of high-k gate oxides with metal gates, the use of mechanically induced strain in the channel to boost the transport mobility and the use of silicon epitaxy and metal silicides which reduce the resistances of source, drain and gate electrodes. Now the author focuses discussion on the main theme of this paper and that is reduction of short channel effects narration. Subthreshold swing and DIBL are caused by lateral encroachment of source and drain field lines into the channel. We are more familiar with drain field encroachment towards the channel for nm scale source and drain separation which further reduces gate to source barrier and injection of electrons from source end towards the channel. On the other hand, when high substrate doping is used, the channel potential close to the source end is also lifted or high and electrons therefore are attracted to these high channel potentials also denoting forward source to channel field enhancement, a side effect of source to channel electric field encroachment. Both the effects compete for available inversion charge density underneath the physical gate and reduce threshold voltage. The distribution of electrical potential in the channel region of a MOSFET can be derived directly from Maxwell's equation: $\nabla \cdot D = \rho$ where $D = \varepsilon E$ is the electric displacement field, ε is the permittivity under consideration, E is the electric field and ρ is the electron charge. The three-dimensional Poisson's equation shows how the gate competes with the source and drain to uniformly control the charge in the channel, shown by $\frac{dE_x}{dx} + \frac{dE_y}{dy} + \frac{dE_z}{dz} = -\frac{\rho}{\varepsilon}$ = constant value. For the multigate structures, the gate control is exerted in the y and z directions and competes with the variation of electric field in the x direction, which arises from the source and drain. Since the sum of all the terms of Poisson's equation is constant, any increase of in the control of top and bottom gates, i.e., $\frac{dE_z}{dz}$ or left and right side of the gates $\frac{dE_y}{dy}$ will decrease the penetration of source and drain regions into the channel, $\frac{dE_x}{dx}$. The authors carefully suggested that based on the above three dimensional Poisson's equation, it is possible by using few simplifying assumptions to calculate a most important 2D scaling length parameter or natural length λ which gauges the extent of the acuteness of penetration of source and drain field into the channel and for better short channel immunity, λ must be minimized. A device will be free of short channel effects, such as DIBL and threshold voltage roll-off, if the gate length is at least six times greater than computed λ in the channel as opined by the authors of this paper. If the transistor has a square cross section that is $T_{si} = W_{si}$ where T_{si} is the thickness of silicon substrate and W_{si} is width of the silicon substrate, then the value of the natural length is given by $\lambda_1 = \sqrt{\frac{\varepsilon_{si}}{\varepsilon_{ox}}} t_{ox}\, t_{si}$ in a single gate MOSFET, $\lambda_2 = \sqrt{\frac{\varepsilon_{si}}{2\,\varepsilon_{ox}}} t_{ox}\, t_{si}$ in a double gate MOSFET, $\lambda_3 = \sqrt{\frac{\varepsilon_{si}}{4\,\varepsilon_{ox}}} t_{ox}\, t_{si}$ in a gate all

around MOSFET structure. As can be seen, λ_3 has the lowest value of the three structures and therefore mostly recommended for sub 20 nm gate length MOSFET device. These equations indicate that short channel effects can be minimized by decreasing the oxide thickness t_{ox}, decreasing the silicon film thickness where channel is formed t_{si}, using high permittivity gate insulator with higher ε_{ox}. It has been shown using numerous extensive simulations that the natural length of a tri-gate device is given by $\sqrt{\frac{1}{3}}\,\lambda_1$. The concept of an effective number of gates N can then be devised for which the natural length is $\lambda_N = \sqrt{\frac{\varepsilon_{si}}{N\,\varepsilon_{ox}}\,t_{ox}\,t_{si}}$, which clearly shows the benefits of multigate structures in reducing short channel effects. As both channel length L and λ reduce, the ratio $\frac{L}{\lambda_N}$ shows that for lower value of this parameter DIBL is maximized and rises to 300 mV/V and for a DIBL around 50 mV/V to lower, $\frac{L}{} > 4$. The subthreshold slope shows the similar DIBL dependence and hence it is advised when designing multigate structures of n-channel FET, $\frac{L}{\lambda_N}$ should be sufficiently high value. Below is an illustrative figure taken from this article that highlights the DIBL dependence on $\frac{L}{\lambda_N}$.

Multigate gate all around structures is also the industry adopted structure for n-channel silicon nanowire architecture with nanometer diameter 10 nm to 5 nm for boosting the drive current, boosting channel mobility from uniform volume inversion effect and reduced surface roughness scattering, I_{on}/I_{off} ratio greater than 106, low off state leakage current, lower DIBL and lower subthreshold slope in the range 60 mV/decade limit for T = 300 K. With regard to FinFETs which is still existing for logic devices in the 20 nm regime as when device dimensions are in the 10 nm range or below, gate all around nanowire FET has increasing become the norm for device architecture selection because of their superior drive current I_{on}, very low I_{off}, high I_{on}/I_{off} ratio and scalability of subthreshold slope near 60 mV/decade limit at T = 300 K. The 3D nature of FinFET, the author believes warrants close device physics based probe and therefore, another classic review paper entitled “FinFETs: From Devices to Architectures” [182] by D. Bhattacharya and Niraj K. Jha are analyzed here. As the authors wisely point out in the introduction section of this review paper, among all multiple gate FETs (MGFETs), FinFETs (a type of DGFETs) and Tri-gate FETs have emerged as the most desirable alternatives to conventional MOSFETs in the 20 nm channel length regime due to their simple structure and ease of fabrication. Taking notes from the authors’ presentation from the structural configuration of FinFETs, while the planar MOSFET channel is horizontal, the FinFET channel (also known as the fin) is vertical. Hence the height of the channel (H_{Fin}) determines the width of the FinFET. This leads to a special property of FinFETs known as width quantization. This property says that the FinFET width should be a multiple of H_{Fin}, that is, the width can be increased by using multiple fins. Thus, arbitrary FinFET widths are not possible. Although smaller fin height offers more flexibility, they lead to multiple fins which in turn leads to more silicon area. On the other hand, taller fins lead to less silicon footprint but also

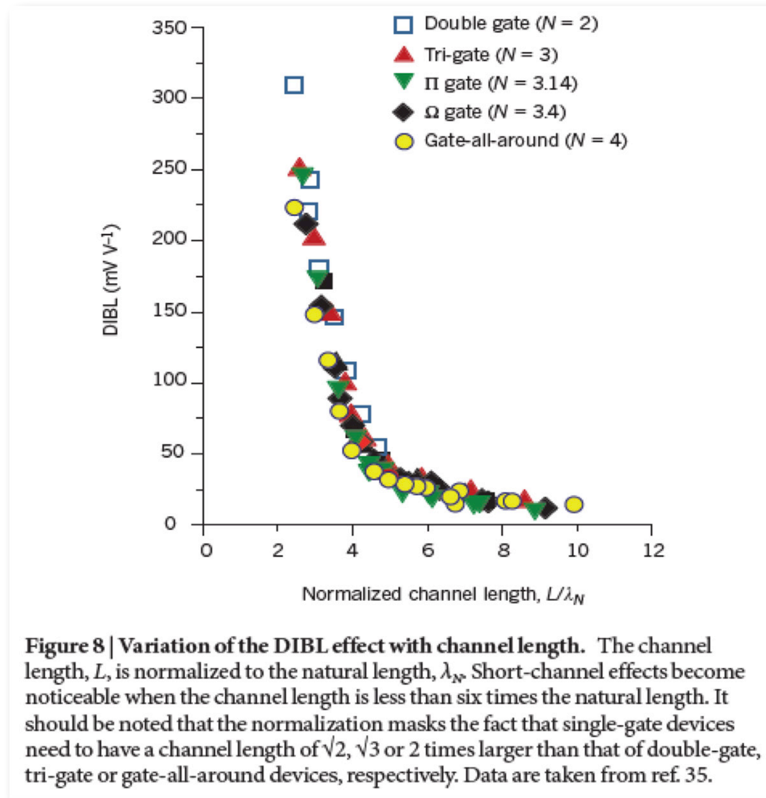


Figure 8 | Variation of the DIBL effect with channel length. The channel length, L, is normalized to the natural length, λ_N. Short-channel effects become noticeable when the channel length is less than six times the natural length. It should be noted that the normalization masks the fact that single-gate devices need to have a channel length of $\sqrt{2}$, $\sqrt{3}$ or 2 times larger than that of double-gate, tri-gate or gate-all-around devices, respectively. Data are taken from ref. 35.

Figure 5.1: This figure is taken from article "Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transistors" [175] by Jean-Pierre Colinge et al.. Copyright permission is required for this figure.

may lead to structural instability. Typically, the fin height is determined by the process engineers and kept below four times the fin thickness. Although FinFETs implemented in SOI wafers are very prevalent, FinFETs are also implemented on conventional bulk silicon wafers extensively. Unlike bulk FinFETs where all fins share a common silicon substrate, fins in SOI FinFETs are isolated by the buried oxide from the substrate. When comparing Tri-gate FET with FinFET, the authors referred to the important research finding of Yang and Fossum et al., where out of detailed performance simulation the authors Fossum et al. concluded that FinFETs are superior to Tri-gate FET when considering scaled performance. Yang and Fossum et al. showed in their analysis that although undoped Tri-gate FETs may enjoy more relaxed body thickness, they are not competitive with FinFETs with regard to SCE metrics. When trying to achieve comparable SCE metrics, Trigate FET lose the scaling advantage and suffer from significant layout area disadvantage. Nevertheless, Trigate FETs are designed, fabricated and analyzed by companies such as Intel for inclusion in their microprocessor unit and conclusive remarks about selection of a single device out of FinFET and Trigate FET have been declared as premature by the Niraj K. Jha et al. in their review article being discussed here. FinFETs can be fabricated with their channel in different directions in a single die. Fabrication of planar MOSFET channels on any plane other than <100> is difficult due to process variation and interface traps. However, FinFETs can be fabricated along the <110> plane as well. This results in enhanced hole mobility. <110>-oriented FinFETs can be fabricated by simply rotating the transistor layout by 45° in the plane of a <100> wafer. Thus, n-FinFETs implemented along <100> and p-FinFETs implemented along <110> planes lead to faster logic gates since this gives designers the opportunity to combat the inherent mobility disparity between electrons and holes. Next the authors of this FinFET review paper goes to discuss two types of FinFETs, shorted gate (SG) FinFETs and independent gate (IG) FinFETs. SG FinFETs are also known as three terminal (3T) FinFETs and IG FinFETs as four terminal (4T) FinFETs. In SG FinFETs, both the front and back gate terminals are physically shorted whereas, in IG FinFETs the gates are physically isolated. Thus in SG FinFETs, both gates are jointly used to control the electrostatics of the channel. Hence SG FinFETs show higher on current Ion as well as higher Ioff compared to IG FinFET. IG FinFETs offer the flexibility of assigning different gate voltages on front and back gate, such as using the back gate bias to modulate the threshold voltage V_{th} defined by the front gate voltage to speed up the device and also control subthreshold leakage current. However, IG FinFETs incur a high area penalty due to the need of placing two separate gate contacts. SG FinFETs can be further categorized based on asymmetries in their device parameters. Normally the work functions of the metal gates of both front and back gate are the same. However, while the gates are tied, their work functions can be suitably chosen by selecting different gate material with same gate potential voltage. This leads to an asymmetric gate (ASG) SG FinFET. ASG SG FinFETs can also be fabricated with selective doping of the two gate stacks. They have very promising short channel characteristics and provide two orders of magnitudes of lower Ioff compared to SG

FinFET at the price of lower I_{on}. The authors took the reference from Goel et al. that asymmetric-drain-spacer-extended (ADSE) FinFETs can lead to improved short channel characteristics because of an indirect increase in channel length. However, this asymmetry comes out of increased layout area. This asymmetry also destroys the conventional interchangeable source drain concept in CMOS. An asymmetry is created in the source to drain current I_{ds} and drain to source current I_{sd} because of extra underlap in the drain side. This asymmetry affects the FinFET pass transistor performance. A good alternative the authors mention in their review study is the asymmetric oxide thickness (AT_{ox}) SG FinFETs with good subthreshold slope and higher gate to channel coupling capability. Use of IG FinFETs or 4-T FinFETs offer an extra tuning capability on V_{th} which can be linearly increased or decreased. This asymmetry can be achieved using an ion bombardment enhanced etching process. Finally, the authors refer to the asymmetric fin heights in fabrication of FinFETs. Since the channel width or extent of inversion charge density is proportional to fin height, p-FinFETs with taller fins can compensate for the lower mobility of p-FETs in an inverter design with its corresponding n-FinFET.

Next, the authors of this FinFET review paper shine light on important process variations associated with FinFET manufacturing. As it is well known that reduced feature size and limited photolithographic resolution cause statistical fluctuations in the nanoscale device parameters. These fluctuations cause variations in performance benchmark metrics such as V_{th}, subthreshold slope, drain induced barrier lowering (DIBL), I_{off} and most importantly I_{on}. From random dopant fluctuations (RDF) variation perspective where $\pm$ 3 σ variation of V_{th} is considered, the inherent advantage of FinFETs having dual gate surrounding the fin comes in the form of lower body doping capability and adjusting the V_{th} through the work functions of front and back gates. Reduced body doping reduces RDF effect on all the benchmark parameters mentioned above. Hence designers can keep the thin channel (fin) close to intrinsic level doping $10^{15}/cm^3$. Low channel doping enhances carrier mobility and also reduces ionized impurity scattering and gate driven vertical surface field leading to reduced surface roughness scattering as for 3D device architectures with aspect ratio mismatch, interface and edge roughness scattering are dominant. FinFETs do suffer from other process variations, such as variations in gate length L_{FG} and L_{BG}, fin thickness T_{si}, gate oxide profiling fluctuations (T_{Fox} and T_{Box}) and gate underlap (L_{GUN}). For example, when fabricated, the gate oxides are on the etched sidewalls of the fin and may suffer from nonuniformity bringing potential line edge roughness (LER) of the fin, an extra source of process variation being researched extensively. LER also causes variation in fin thickness. The effect of temperature variation for SOI FinFETs can be significant on device and circuit performance degradation due to poor heat removal capability of buried oxide in SOI FinFETs. Author Niraj K. Jha with another researcher and paper have evaluated SG, IG and ASG FinFETs under temperature variation and found that even though I_{off} degraded for all the variants of FinFETs, ASG FinFETs still performed superior under temperature variation in fin thickness. The review paper by D. Bhattacharya and N. K. Jha et al.

provide further analysis and discussion on FinFET device characterization by process simulation and device simulation with compact models, FinFET standard six transistors SRAM cells with other variants, inverter gates, NAND gates, latches and flip-flops and then circuit level analysis with focus on optimization, novel interconnect structure and logic synthesis and finally FinFETs being used in architectural level analysis of computer chip.

CHAPTER 6

Device Physics-Based Scaling Insights on Tunneling Based n-Type MOSFET Device Architectures

From the requirements for faster turn on and off device switch, device engineers at the end of Moore's Law started focusing on steep slope devices with still considerable drive current I_{on} and lower I_{off}. So far the devices that have been analyzed and documented in Chapter 1–5, use electron injection into the channel over a gate to source barrier governed by thermionic emission and maximum possible initiation velocity at the source end where the channel L is assumed to just about to start. The device engineering focus on these devices is to bring down the subthreshold slope of these devices to the room temperature limit of 60 mV/decade and is limited by minimum gate to source barrier thermionic emission. In order to achieve subthreshold slope in the vicinity of 20–30 mV/decade alternative carrier transport mechanism needed to be evolved and hence, the concept of tunneling field effect device has been devised providing steeper slope amenable to faster logic switching and transitions and at the same time consuming lower switching current during these transitions. As the name reveals, tunnel field effect devices utilize tunneling of electrons from a reverse biased P^+-N^+ junctions where due to reverse bias, the junction profile is very steep and its width is very narrow. When the top gate overlaps this P^+-N^+ junction area, minority electrons in P^+ material can tunnel through the very narrow width of the reverse biased junctions formed there and these electrons subject to similar electron density in an n type drift region, can be collected by the heavily doped n+ drain side. Tunneling is more sensitive to width of the reverse biased energy profile of the P^+-N^+ junction rather t[illegible]vel of the two sides of the P^+-N^+ junction. It ha[illegible]ook regarding overview of this Chapter 6 that for tunneling field effect transistor, the subthreshold slope factor SS has a different expression as per Alan Seabaugh et al. paper and is effectively tailorable with its defining parameters to reduce subthreshold slope less than 60 mV/decade. Since the mechanism is through tunneling, tunneling induced electron drift velocity tend to two to three orders of magnitude lower than thermionic emission or over-the-barrier induced drift velocity. Therefore, different I_{on} enhancement engineering techniques are employed in the source sided junction like strained Si-Ge layer or extra pocket implant and recently vertical staggered heterojunction tunnel FET has shown promise to close the gap between I_{on} provided by conventional MOSFET based devices and tunneling field effect devices. Due to steep slope characteristics of tunnel FET, I_{off} is very low

and these types of devices are most ideal for subthreshold or near threshold ultra low power logic switches where conventional MOSFETs and other derivatives discussed from Chapter 2–5 of these devices are inefficient. In the introductory section of this book documenting the references that will be analyzed in this book will make the content of unmanageable as the review is meant to be within certain page limits. So as the a[illegible]evious paragraphs to the readers that some of these classic papers as mentioned in the Introduction section for every Chapter must be identified by th[illegible] and consulted on their own due to the page[illegible]this book. Therefore, in this Chapter 6, tunnel FET overview, its structural design and performance enhancement method will be discussed and frequently happening notorious ambipolar transport that reduces the I_{on} and enhances I_{off} will be discussed with methods developed in literature will be also outlined.

Since Alan Seabaugh et al.'s classic paper on Tunnel FET has been analyzed to the extent that the readers can get the required idea and knowledge from the analysis presented in the Introduction section of this book, the author turns to another classic review on tunnel FET "Tunneling Field Effect Transistors: Prospects and Challenges" by Ian A. Young et al. As the authors discuss in this paper, silicon's large indirect band gap and large carrier mass mean drive current for silicon TFET is very low. But due to the availability of high quality materials together with years of know-how, Si and Si:Ge TFETs have been studied the most. III-V materials for TFETs attracted attention such as InAs with some of them offering lower band direct band gap and lower tunneling effective mass. Eventually broken band gap vertical III-V heterojunction TFETS have shown the highest drive current attainable. The authors in this paper provide a comprehensive analysis and review of GaSb/InAs TFET for gate length of 13 nm that has been successfully manufactured by Intel where these researchers are professionally engaged. GaSb/InAs heterojunction tunnel FET's I_{on} is larger than silicon MOSFET for $L_G = 13$ nm up to a gate bias of 0. 6 V. When measuring I_{off} of GaSb/InAs TFET, the authors observed that TFET off current shows strong V_{ds} bias susceptibility. Even with 10 nm drain underlap and low drain doping, heterojunction TFET's I_{off} increases sharply as the drain voltage increased beyond 0. 3 V. Whereas heterojunction TFET can support $I_{off} < 10$ pA/μm at $V_{ds} = 0.3$ V, the lowest I_{off} achievable at $V_{ds} = 0.5$ V is 1 nA/μm. Another advantage emerging from GaSb/InAs heterojunction tunnel FET is its lower gate capacitance than Si MOSFET due to the low density of states (DOS) of InAs conduction band. This is advantageous for lowering dynamic power and circuit delay although from inversion channel carrier density view point, lower gate capacitance means lower gate to channel coupling and loss of inversion charge density. When device parameter variations are considered, TFETs are affected more from random dopant fluctuations (RDF) at its source to channel junction since it modifies the tunneling transmission area shape or its width and height. It should be mentioned here that only for a limited range of gate to source junction overlap voltage, energy states at the P^+ side and N^+ side are aligned so that electrons tunnel from valence band state of P^+ region to the conduction band state of N^+ region. Next, the authors discuss important subthreshold slope (SS) variability and its degradation by

defects and traps positioned near the source or adjacent to source to channel region. III-V materials are more prone to nonuniformities in their crystal structure such as vacancies or interstitials, interface states and charged ionized or unionized impurities coming out of dopants implanted as part of the desired structure. The main consequence of these nonuniformities is to add extra electron/hole states inside the ideal band gap weakening the energy filtering which the TFET uses to achieve steep SS. Defects can be modeled by introducing trap states into the existing band structure and allowing transport through these states having trap lateral size and energy spectrum varying with the defect type. The TFET is more susceptible to trap induced degradation than the MOSFET, if the location and energy of the traps are assumed closer to N-TFET's worst case point such as inside the channel with the energy levels 0.1–0.2 eV below the conduction band. However, this issue is not limited to TFET, the same trap-induced tunneling (TAT) process would cause the leakage issues in the MOSFET through the GIDL (Gate induced drain leakage) process when the traps are extended along the channel drain junction of the transistor. GIDL and band to band tunneling (BTBT) current are important components of reverse bias leakage current for TFETs when $V_{gs} < 0$. If the body doping is very high in the substrate, severe band bending happens near source to substrate p-n junction near the surface and generate BTBT current which may be an issue for MOSFET but surely an issue for TFET where both p and n region doping at the source junction is very high. Thus the specific energy and location of distribution of the traps inside the energy gap for fabricated experimental devices will impact the level of SS degradation in the TFET and MOSFET I-V characteristics. The authors infer that the main reason for SS degradation is the trap assisted tunneling which is more prominent in materials like III-V other than silicon possessing high bulk and interface defects. The experimental realization of TFET's significant advantage of SS over MOSFET is highly dependent on progress in the quality of these novel materials and their minimum defect density enabled crystal growth. Another non ideality concern is the density of states (DOS) extending into the band gap (band tail) due to high doping density. This effect is due to the nonhomogeneous distribution of the doping atoms creating different local potentials when compared with homogeneous doping condition. When statistically modeled over large areas, this effect can be described as an exponential decay of DOS into the band gap. Since the TFETs considered for future technology nodes have very small cross sections (5 nm nanowire), the averaging effect is not valid. A more appropriate approach is to consider this issue as a part of source doping variation and study its effect on Ioff variation. Through a 3 σ variation of random dopant induced source doping variation, it is found its effect is not as critical as 3 σ induced V_T variation on I_{off}, implying high doping induced band tails are not critical from off state leakage current performance of TFET. Again, the authors stresses the point that to subdue the effect of band tails due to high doping of source p-n junctions, high quality epitaxial layer where the dopants are at interstitial positions should be ensured during fabrication and thermal annealing treatment but defect degeneration due to heavy dopant implant or low quality epi-growth is expected to bring the same SS degradation problems that resulted out of traps in these devices. Other

sources of recombination/generation mechanisms (SRH) mostly happening with in direct band gap materials are also source of leakage and are more compounded at elevated temperatures and strongly depend on material quality. With regard to scaling requirements of these n-TFET design, the authors noted that the TFETs with narrow band gap and longer channel length, ambipolar transmission is the main source of SS degradation and high $I_{off.}$ However, in short gate length TFETs, the shorter tunneling path between source and drain aggravates direct source to drain tunneling leakage while the increased bandgap due to confinement reduces the ambipolar leakage problem. Unlike in the MOSFET, gate oxide scaling does not improve short channel effects in TFETs significantly instead body thickness scaling is the most important parameter. Compared to MOSFET built with double gate structure and nanowire (NW) structure which almost have similar drive current for low gate to source voltage V_{gs} up to 0. 5 V, significant upshift in heterojuction TFET built with nanowire is observed compared to double gate heterojunction TFET implying nanowire radius or body thickness scaling is important metric here for heterojunction TFET I_{on}. Figure 6.1 is taken from this reference article that points to this discussion.

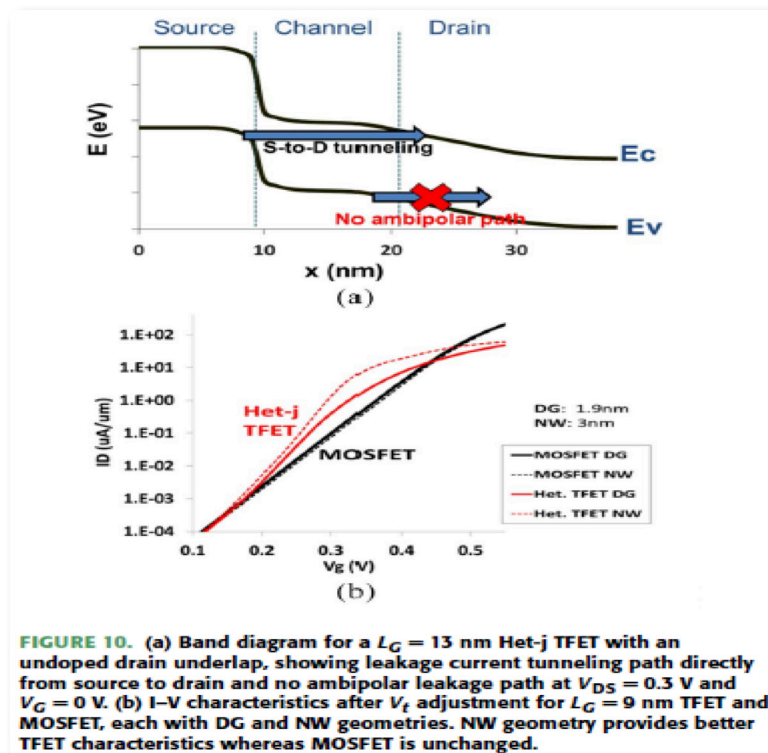


Figure 6.1: This Figure is taken from article " Tunnel Field-Effect Transistors: Prospects and Challenges" by Ian A. Young et al., *Journal of the Electron Device Society*, Vol. 3, No. 3, May 2015, page no. 93. Copyright permission is required for this figure.

The authors project that the body thickness requirement is much tighter for TFET than MOSFET with 3 nm NW recommended for L_g = 9 nm. The authors also observe importantly that as LG scales, two dimensional semiconductors with intrinsically superior electrostatics may benefit the performance of the TFETs.

For enhancing the I_{on} of TFET, the author discusses another reference article "Drive Current Boosting of N-Type Tunnel FET with Strained Sige Layer at Source" [196] by Santanu Mahapatra et al. In their paper, the authors propose that the enhancement in I_{on} is achieved by introducing a thin strained SiGe layer on top of the silicon source and the drive current I_{on} increases exponentially with the Ge mole fraction. For a tunnel FET, the I_{on} is proportional to the electron/hole transmission probability T(E) in the BTBT mechanism which is given by

$$T(E) = \exp\left(-\frac{4\sqrt{2\,m^*}\,E_G^{\frac{3}{2}}}{3\,|e\,|\,\hbar(E_G+\Delta\varphi)} \sqrt{\frac{\varepsilon_{si}}{e_{ox}} t_{ox}\, t_{si}} \right) \Delta\varphi \ ,$$

where m^* is the tunneling effective mass, e the electron charge, E_g is the band gap of semiconductor material, $\Delta\varphi$ is the energy range over which tunneling takes place, ε_{si} and ε_{ox} are the permittivity of silicon and oxide respectively, t_{ox} is the gate dielectric thickness and t_{si} is the TFET body thickness. Profiling the source to channel p-n junction energy profile to increase the $\Delta\varphi$ and decreasing the E_g are the two most viable option to increase T(E). When the body thickness t_{si} is decreased considering lower nm NW radius, the sub bands in the conduction band have a wider separation, so an electron from an energy state in the valence band on P^+ region has a misaligned state on the N^+ side reducing $\Delta\varphi$ and T(E) is decreased. Also when ε_{ox} is increased and t_{ox} is decreased, C_{ox} or oxide capacitance is increased creating more inversion charge in the n^- drift region and decreasing the reverse bias through decreased channel potential in the depletion region of P+-N+ junction at the source. As a result, the energy band becomes less steeper, tunneling width widens and tunneling rate of electrons reduces decreasing T(E). Finally, when ε_{si} is increased, the reverse bias width is increased analogous to reverse bias controlled more steeper energy profile in the P^+-N^+ junction at the source contributing to more tunneling rate and increased T(E). A note on tunneling effective mass m^*. From the equation it infers that the tunneling mass contribute to faster tunneling rate by increasing mobility of tunneling electrons. But generally, when m^* is low, density of states (DOS) in the conduction band is low and sub band energy separation in conduction band is wider. Both of these factors should reduce tunneling rate but it looks like it more contributes to tunneling mobility. The authors introduced a thin SiGe layer on top of Si source material and varied E_G to tune to maximizing T(E) by varying Ge mole fraction. The device performance is more sensitive to degenerate doping at the source and the abruptness of its junction profile. In the device simulation model, degeneracy induced band gap narrowing model and exact Fermi-Dirac statistics are used by the authors. Also nonlocal tunneling criterion is used which causes the electrons to be generated at the end of the tunneling path as implied by the tunneling physics. The band gap at the source

end reduces as the Ge mole fraction is increased. This results in a reduction of tunneling band gap E_G and increase in T(E). The authors also found through their simulation that there is negligible change in tunneling width and height with increase of V_{ds}. Additionally, this increase in I_{on} leads to a reduction of average subthreshold swing since the device threshold voltage now falls in the steeper region of the curve. The authors through their simulation recorded an I_{on} = 580 µA/µm, I_{off} = 0.52 fA/µm and average SS of 13 mV/decade is obtained for a Ge mole fraction of 0.7 with V_{dd} = 1.2 V. The authors also found through simulation that tunneling of carriers mostly happens at the surface level and BTBT generation is found to be within maximum to about 20 nm thickness of SiGe layer. Any additional increase of SiGe layer beyond 20 nm contribute to the I_{off} of the device. Increasing Ge mole fraction within this 20 nm thickness of SiGe layer increases the T(E) and I_{on}. The authors further explained the reasons behind placing the strained SiGe layer on top of Si material. First, the active area is located only at the source channel junction and the drain voltage V_{ds} does not have any impact on BTBT tunneling at the source junction. Adding strained SiGe layer on drain would add strain to the channel from both sides and may be a cause of either increase of mobility or decrease of it as a combined effect. The effect of strain in the silicon channel away from the source junction is not considered. After closely probing this article where the authors provided the transmittance T(E) , the tunneling mass m* should be variable with increasing Ge mole fraction of the SiGe layer on top of Si source and with strained SiGe layer and both these engineering modifications of the TFET's source configuration should decrease m* considerably (m* decreases with increasing Ge mole fraction in SiGe layer) and further increase T(E). The authors of this paper seemed to overlook this modeling requirement of m* of top SiGe layer which is strained and whose Ge mole fraction is varied up to 0.7. For an n-TFET structure, the carrier density that is available for tunneling is $\frac{n_i^2 eff}{N_A}$ where N_A is degenerately doped acceptors in P^+ material of source junction. N_A doping even though degenerate must be such that $n_i^2 eff$ increases considerably in 10th power owing to more intense band gap narrowing effect on Ge mole fraction reduced Si band gap E_G so that the minority electron density available for tunneling is increased up to almost $10^{11}/cm^3$ or more for a N_A doping of $10^{20}/cm^3$. It should be also noted that the top Si:Ge layer's strain on Si P^+ material increases with Ge mole fraction and additionally reduces band gap E_G of silicon P^+ material and this should further lift ni,eff value and the minority electron concentration from the above stated value beneficial for overall tunneling current. The drift region adjacent to N^+ material should be quite intrinsically doped to match the overall tunneling electron density that emerges out of N^+ material of the source junction. Intrinsically doped n- drift region will also invert with low enough gate voltage and moderate C_{ox} as high C_{ox} is detrimental to T(E) shown before in the analysis. The drain which is N^+ should be biased with high V_R so that the energy band profile of the source junction becomes sufficiently steep and its width narrow paving the path for interband tunneling

in some heterojunction III-V TFET structure. Intrinsic drift region will ensure negligible voltage drop in the n- drift region and most reverse bias voltage will appear at the P^+-N^+ source junction.

CHAPTER 7

Conclusions

The author wishes to make few essential comments and observations that came out of this extensive review work documented in Chapter 1–6. First, towards the end of Moore's Law with ITRS projections and roadmap, a suitable device architecture has to be chosen for maximum I_{on} so that it closely attains the value listed by ITRS. For maximum I_{on} at room temperature T = 300 K when the channel length is 10 nm or below, gate-all-around nanowire with cylindrical gate must be chosen with near intrinsic doping in the range of $10^{15}/cm^3$. The readers by this time should be aware about the pros and cons of this nanowire architecture that delivers the best I_{on}. For more current drive requirements, these nanowires can be driven in parallel or stacked on top of each other in the form of nanosheets and researchers have shown that nanosheet architecture with wrapped around gate is compatible with high I_{on} requirement. Now what about minimum value of I_{off} needed for better I_{on}/I_{off} or maximum value of I_{off} the device can tolerate before power dissipation becomes an issue? For minimum I_{off}, gate-all-around nanowire tunnel FET should be the obvious choice as it provides the lowest subthreshold current at a very low gate voltage. But then there are other components of I_{off} in TFET structure such as trap assisted tunneling (TAT) and band to band tunneling (BTBT) current at reverse gate bias or small forward gate bias for TAT that can elevate the value of I_{off} and decrease I_{on}/I_{off} ratio. Also for gate-all-around nanowire when the gate voltage is small, close to threshold voltage, the carriers forming the channel can leak through the ultrathin gate oxide to the positive gate contact forming gate leakage current. It has been shown that as the device dimensions shrink, gate leakage current increases at a faster rate than subthreshold leakage current and therefore, I_{off} controlling to keep its value below maximum tolerable level could be challenging when we are entering the 5 nm and 3 nm nanowire gate length or radius node. For ultralow power dissipation core processors in digital central processing unit (CPU) tunnel FET architecture should still be chosen as despite it has elevated I_{off}, its I_{on} is about 100th fraction of gate-all-around nanowire FET architecture for good level of gate voltage up to 0.5 V and this is due to the tunneling transmission probability based carrier injection happens at a much lower rate than over the top thermionic emission based carrier injection over gate to source barrier in FETs other than TFETs. Lower I_{on} at a low gate voltage below threshold for the TFET structure will enable ultra low power functioning of the chip wherever required. Now we need to embark on more interesting projection of ballistic or quasi-ballistic transport when the device is in the range of 10 nm or below, particularly in nanowire configuration. Although the paper by Lundstrom et al. has shown that the maximum drift velocity at the source point or where L = 0 defines the condition for maximum attainable I_{on}, from L=0 to few 0.2 L in the channel when the channel potential

is increasing downward from top of the gate potential to the drain potential at V_{ds} at the end of channel, electrons may backscatter significantly to the source side and compensate the drive current from source to drain. How we can reduce this backscattering factor? Increasing V_{ds} value has been already suggested in the paper by Lundstrom so that channel potential drop from drain to source are a little abrupt and these back scattering electrons positioned within 0.2 L witness a higher barrier on V_{GD} potential drop (taking V_G =0 and then increasing downward to V_{ds} = 0.5 V, so that V_{GD} = -0.5 V) because closer to source V_c or channel potential value is raised when V_{ds} is increased and that makes V_{gx} to steeper decline where V_x is the potential of V_c at some point where L is within 0.2 of channel length. When the back scattering electrons on the V_{GD} side of the channel potential close to source are restricted within a higher potential drop (V_{gx} slightly more negative as V_x is raised as discussed), these electrons will not be able to back scatter to the source side of V_{gs} potential and then the drift velocity at source point L = 0 should be higher. We can also increase the forward drift field by slight negative voltage applied to the source like -0.1 V or -0.15 V. This will reduce the V_{gs} barrier by the forward biasing of the source to substrate junction and total V_{ds} is also now greater as Vs< 0 V. Then the channel potential from drain to source will be elevated as they gradually decline and spatial positions closer to the source will witness higher channel potential or more V_{gx} drop at that point where channel potential is V_x. This will enhance the barrier more for electrons within 0.2 L to back scatter to the source and hence the maximum source point drift velocity can be ensured. There is another device engineering method available for increasing the abruptness of V_{GD} decline. If we use graded doping with slightly higher doping closer to the source than the substrate and gradually adjust or align the doping to actual substrate doping near middle of the channel to the drain end, then again the channel potential closer to the source will be raised due to higher substrate doping but fixed V_{ds} value. This will induce sharp decline at V_{GD} profile from top of V_G and then smoothening the abruptness and eventually aligning with V_{GD} profile near drain side. Higher substrate doping near the source end to 0.2 L in the channel will alter the V_T and number of inversion charge density and they also witness higher barrier from the V_{gx} channel profile where V_x is the channel potential with the slightly higher substrate doping and again the number of back scattering electrons reaching to the source will be reduced since V_{gx} drop is now more abrupt within 0.2 L due to the use of graded doping or slightly higher substrate doping which gradually drops to the original doping of the substrate. Other process technology based accuracies must be ensured so that V_{gs} [illegible]pt and this barrier is not too high and in line with scaling associated with supply voltage and also V_{GD} potential profile should be also abrupt with top of the barrier should be a single point in terms of V_G potential assignment. Apart from the device engineering adaptations mentioned heretofore, self aligned gate to source and drain must be established with no underlap of gate with source or drain junction. Ultra shallow source and drain junction should be formed so that the curved portion of the junction is minimal both at the source inception point and at the inception of drain junction. These technologies enable tight control on

V_T and its roll off prevention so that barriers at V_{gs} and V_{GD} both hold strong and backward scattered electron transport can be minimized.

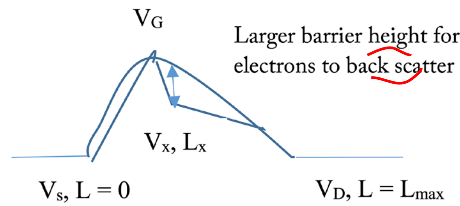


Figure 7.1: Using a higher step substrate doping within a small zone around source inception point, the barrier height of the potential profile of V_{GD} can be increased allowing less back scattering of electrons that are situated at top of V_{GD} (almost at L = 0) to approximately 0.2 L towards the drain side of the channel.

Also required is reduction of source and drain junction series resistance at the surface, for that the extent of source and drain junction can be terminated with shallow trench isolation or STI. Reduced series resistance at the source junction coming out of source contact resistance which has been discussed in detail in previous Section 1.1 and also from STI terminated junction length, width and area ensure that the drift velocity V_T at the source point L =0 remains high. Now Natori et al. in his paper has mentioned that for ballistic transport at room temperature or at temperatures much below than 300 K, low substrate doping or intrinsic doping should be used. But lower substrate doping at very low temperatures down to 4 K also increases carrier freeze out or incomplete ionization percentage so that the carrier density is very small number and could be even 10^5 to 10^7/cm^3 at 4 K if the original doping at T = 300 K is 10^{15}/cm^3. Then even if carrier mobility after ballistic transmission rises to 10^6 cm^2/V-s, the total current is still very low using nm dimension cross sectional area through which current flows. So we need to deal with this carrier freeze out at low temperatures where ballistic transport is possible. One way to maintain the degree of ballisticity high and ensure carrier freeze out is not significantly deteriorating the carrier density in transport is to degenerately dope the substrate either in 1D configuration nanowire or in 3D configuration nanowire such that it takes almost the form of a metallic nanowire. For degenerately doped 1-D nanowire, E_D or donor activation energy which is function of doping concentration shifts up in energy as E_c level is shifted up at lower substrate temperature close to 4 K, so E_C–E_D is still a very low value. Besides, when degenerate doping is used, individual donor sites are very much closely spaced and form band or become indistinguishable from one another and may overlap Ec which makes almost complete ionization of carriers at T = 300 K for degenerate doping levels in the range 10^{19}/cm^3 to 10^{21}/cm^3 and must be lower than the maximum number of atoms in a unit cell for silicon, i.e., 5 × 10^{22}/cm^3. Additionally, at temperatures below T = 300 K and at 4 K E_F or Fermi energy level is very much close to E_c but since degenerate doping has been used, E_D level shifts up

and at the same time E_F level may be inside the conduction band above than E_C level. The energy levels below E_F is partially filled if those energy levels are below E_F but not too much below. This together with E_c-E_D level reduction even at T = 4 K can make the ionization percentage 8–10% for highly degenerate doping, i.e., 1020/cm3. Then even if ballistic transmission related mobility drops to certain fraction of what can be maximally achieved by near intrinsic doping, the total drive current can still be some fraction of mA/μm. We have to ensure that the drift field E is not at E_{sat} and taking V_d = 0.1 V and L = 10 nm, the drift field E is 10^5 V/cm allowing the mobility value to peak by ballistic transmission and avoiding too much fraction of a carrier freeze out at T = 4 K or temperatures within 100 K from T = 300 K by choosing degenerate doping of the substrate. The author believes the observations enunciated in this conclusion section will be probed with a deeper eye by the readers and researchers and whenever possible, apply the correct device physics based reasonings to justify analytical modeling based equations and simulation outcomes that have been performed on each device architecture discussed in this book.

References

1. *Semiconductor Device Fundamentals*, Robert F. Pierret, Addison-Wesley Publishing Company Inc., 1996.
2. *Advanced Semiconductor Fundamentals*, Robert F. Pierret, Modular Series on Solid State Devices, Volume VI, second edition, Pearson Education Inc., 2003.
3. *Semiconductor Physics and Devices Basic Principles*, Donald A. Neamen, fourth edition, McGraw Hill Companies Inc., 2012.
4. *Fundamentals of Modern VLSI Devices*, Yuan Taur and Tak H. Ning, Cambridge University Press, 1998.
5. *Fundamentals of Modern VLSI Devices* (Expanded coverage), Yuan Taur and Tak H. Ning, Second edition, Cambridge University Press, 2009.
6. *Semiconductor Physics*, Karl W. Böer and Udo W. Pohl, Springer International Publishing AG, 2018.
7. *Fundamentals of Semiconductors-Physics and Materials Properties*, Peter Y. Yu and Manuel Cardona, fourth edition, Springer-Verlag Berlin Heidelberg, 2010.
8. *Fundamentals of Solid State Engineering*, Manijeh Razeghi, Kluwer Academic Publishers, 2002.
9. *Semiconductor Physical Electronics*, Sheng S. Li, second edition, Springer Science + Business Media LLC, 2006.
10. *Physics of Semiconductor Devices*, S. M. Sze and Kwok K. Ng, John Wiley and Sons Inc., 2007.
11. *Physics of Semiconductor Devices*, Jean-Pierre Colinge and Cynthia A. Colinge, Kluwer Academic Publishers, Springer US, 2002.
12. *Nanowire Transistors-Physics of Devices and Materials in One Dimension*, Jean-Pierre Colinge and James C. Greer, Cambridge University Press, 2016.
13. *Nanoscale CMOS-Innovative Materials, Modeling and Characterization*, Francis Balestra (editor), John Wiley & Sons Inc., 2010.
14. *Frontiers of Nanoscience–Materials and Processes for Next Generation Lithography*, Editor: Alex Robinson and Richard Lawson, Elsvier Ltd., 2016.
15. *Low Temperature Electronics-Physics, Devices, Circuits and Applications*, Edmundo A. Gutierrez-D, M. Jamal Deen, and Cor L. Claeys, Academic Press, 2001.

16. *Germanium Based Technologies—From Materials to Devices*, Cor Claeys and Eddy Simoen, Elsevier BV, 2007.

17. *Advanced Physical Models For Silicon Device Simulation*, Andreas Schenk, Springer-Verlag Wien, 1998.

18. *Fundamentals of Nanoscaled Field Effect Transistors*, Amit Chaudhry, Springer Science + Business Media New York 2013.

19. *Fundamentals of III-V Semiconductor MOSFETs*, Editors: Serge Oktyabrsky and Peide D. Ye, Springer Science + Business Media LLC, 2010.

20. *Strain Effect in Semiconductors-Theory and Device Applications*, Yongke Sun, Scott E. Thompson, and Toshikazu Nishida, Springer Science + Business Media LLC, 2010.

21. *Fundamentals of Tunnel Field-Effect Transistors*, Sneh Saurabh and Mamidala Jagadesh Kumar, Taylor, and Francis Group LLC, 2017.

22. *Tunneling Field Effect Transistor Technology*, Editors: Lining Zhang and Mansun Chan, Springer International Publishing Switzerland, 2016.

23. *Beyond CMOS Technologies for Next Generation Computer Design*, Editors: Rasit O. Topaloglu and H.-S. Philip Wong, Springer International Publishing AG (part of Springer Nature), 2019.

24. *Transport in Nanostructures*, David K. Ferry, Stephen M. Goodnick and Jonathan Bird, second edition, Cambridge University Press New York, 2009.

25. Miniaturization of CMOS, Henry H. Radamson, Xiaobin He, Qingzhu Zhang, Jinbiao Liu, Hushan Cui, Jinjuan Xiang, Zhenzhen Kong, Wenjuan Xiong, Junjie Li, Jianfeng Gao, Hong Yang, Shihai Gu, Xuewei Zhao, Yong Du, Jiahan Yu, and Guilei Wang, *Micromachines* 2019, 10, 293, pp. 1–52.

26. The end of CMOS scaling: toward the introduction of new materials and structural changes to improve MOSFET performance, T. Skotnicki, J. A. Hutchby, Tsu-Jae King, H-S. P. Wong, and F. Boeuf, *IEEE Circuits and Devices Magazine*, Volume 21, Issue 1, Jan-Feb. 2005, pp. 16–26.

27. State of the Art and Future Perspectives in Advanced Semiconductor Technology, Henry H. Radamson, Huilong Zhu, Zhenhua Wu, Xiaobin He, Hongxiao Lin, Jinbiao Liu, Jinjuan Xiang, Zhenzhen Kong, Wenjuan Xiong, , Junjie Li, Hushan Cui, Jianfeng Gao, Hong Yang, Yong Du, Buqing Xu, Ben Li, Xuewei Zhao, Jiahan Yu, Yan Dong, and Guilei Wang, *Nanomaterials* 2020, 10, 1555, pp. 1–86.

28. Mosfet scaling trends, challenges, and key associated metrology issues through the end of the roadmap, Peter M. Zeitzoff and Howard R. Huff, *AIP Conference Proceedings* 788, 203 (2005), pp. 203–213.

29. MOSFET performance scaling-part I: Historical Trends, Ali Khakifirooz and Dimitri A. Antoniadis, *IEEE Transactions on Electron Devices*, Vol. 55, No. 6, June 2008, pp. 1391–1400.

30. MOSFET performance scaling-part II: Future directions, Ali Khakifirooz and Dimitri A. Antoniadis, *IEEE Transactions on Electron Devices*, Vol. 55, No. 6, June 2008, pp. 1401–1408.

31. Perspectives on scaling theory and CMOS technology-understanding the past, present and future, D. Foty, *Proceedings of the 2004 11th International Conference on Electronics, Circuits and Systems*, 2004, Dec 2004, pp. 631–637.

32. Scaling MOSFETs to the limit-A physicist's perspective, M. V. Fischetti, *Journal of Computational Electronics*, Vol.2, 2003, pp. 73–79.

33. Shrinking limits of silicon MOSFETs: numerical study of 10 nm scale devices, Y. Naveh and K. K. Likharev, *Superlattices and Microstructures*, Volume 27, Issues 2–3, February 2000, pp. 111–123.

34. Performance limitations of Si bulk CMOS and alternatives for future ULSI, Krishna C. Saraswat, Donghyun Kim, Tejas Krishnamohan, and Abhijit Pethe, *ECS Transactions*, Volume 8, No. 1, pp. 9–15.

35. Device physics at the scaling limit: what matters? [MOSFETs], M. Lundstrom, *IEEE International Electron Devices Meeting* 2003, pages 33.1.1–33.1.4.

36. In search of "Forever," continued transistor scaling one new material at a time, Scott E. Thompson, Robert S. Chau, Tahir Ghani, Kaizad Mistry, Sunit Tyagi, and Mark T. Bohr, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 18, No. 1, Feb. 2005, pp. 26–36.

37. Transistor scaling with novel materials, Meikei Ieong, Vijay Narayanan, Dinker Singh, Anna Topol, Victor Chan, and Zhibin Ren, *Materials Today*, Volume 9, Number 6, June 2006, pp. 26–31.

38. Nanoscale silicon MOSFETs: A theoretical study, Victor A. Sverdlov, Thomas J. Walls and Konstantin K. Likharev, *IEEE Transactions on Electron Devices*, Vol. 50, No. 9, September 2003, pp. 1926–1933.

39. Physics and technology of ultra short channel MOSFET devices, D. A. Antoniadis, *IEEE International Electron Devices Meeting* 1991, pages 2.1.1–2.1.4.

40. A nanotechnology enhancement to Moore's Law, Jerry Wu, Yin-Lin Shen, Kitt Reinhardt, Harold Szu, and Boqun Dong, *Applied Computational Intelligence and Soft Computing*, Hindawi Publishing Corporation, Volume 2013, Article ID 426962, pp. 1–13.

41. A perspective on the theory of MOSFET scaling and its impact, Tak H. Ning, *IEEE SSCS Newsletter*, Winter 2007, pp. 27–30.

42. Generalized scale length for two dimensional effects in MOSFET's, David J. Frank, Yuan Taur, and Hon-Sum P. Wong, *IEEE Electron Device Letters*, Vol. 19, No. 10, October 1998, pp. 385–387.

43. MOSFET scaling limits determined by subthreshold conduction, Joseph M. Pimbley and James D. Meindl, *IEEE Transactions on Electron Devices*, Vol. 36, No. 9, September 1989, pp. 1711–1721.

44. Temperature scaling theory for low-temperature-operated MOSFET with deep submicron channel, You-Wen Yi, Kazuya Masu, Kazuo Tsubouchi, and Nobuo Mikoshiba, *Japanese Journal of Applied Physics*, Vol. 27, Number 10A, 1988, pp. L1958–L1962.

45. Nanoscale MOSFETs scaling, Y. Naveh, V. Sverdlov, and K. Likharev, *Device Research Conference, Conference Digest*, June 2001, pp. 61–62.

46. MOSFET scaling-the driver of VLSI technology, Dale L. Critchlow, *Proceedings of the IEEE*, Vol. 87, No. 4, April 1999, pp. 659–667.

47. Scaling equations for the accurate prediction of CMOS device performance from 180 nm to 7 nm, Aaron Stillmaker and Bevan Baas, Integration, *The VLSI Journal* (58) 2017, pp. 74–81.

48. Extremely scaled silicon nano-CMOS devices, Leland Chang, Yang-Kyu Choi, Daewon Ha, Pushkar Ranade, Shiying Xiong, Jeffrey Bokor, Chenming Hu, and Tsu-Jae King, *Proceedings of the IEEE*, Vol. 91, No. 11, November 2003, pp. 1860–1873.

49. Characterization and compact modeling of nanometer CMOS transistors at deep-cryogenic temperatures, Rosario M. Incandela, Lin Song, Harald Homulle, Edoardo Charbon, Andrei Vladimirescu, and Fabio Sebastiano, *IEEE Journal of Electron Devices Society*, Vol. 6, April 2018, pp. 996–1006.

50. Cryogenic MOS transistor model, Arnout Beckers, Farzan Jazaeri and Christian Enz, *IEEE Transactions on Electron Devices*, Vol. 65, No. 9, September 2018, pp. 3617–3625.

51. MOSFET characterization and modeling at cryogenic temperatures, Chao Luo, Zhen Li, Teng-Teng Lu, Jun Xu, and Guo-Ping Guo, *Cryogenics*, Vol. 98, March 2019, pp. 12–17.

52. CMOS scaling into the nanometer regime, Yuan Taur, Douglas A. Buchanan, Wei Chen, David J. Frank, Khalid E. Ismail, Shih-Hsien Lo, George A. Sai-Halasz, Raman G. Viswanathan, Hsing-Jen C. Wann, Shalom J. Wind, and Hon-Sum Wong, *Proceedings of the IEEE*, Vol. 85, No. 4, April 1997, pp. 486–504.

53. Silicon CMOS devices beyond scaling, W. Haensch, E. J. Nowak, R. H. Dennard, P. M. Solomon, A. Bryant, O. H. Dokumaci, A. Kumar, X. Wang, J. B. Johnson, and M. V. Fischetti, *IBM J, Research and Development*, Vol. 50, No. 4/5, July/September 2006, pp. 339–361A.

54. Fundamental limits of silicon technology, Robert W. Keyes, *Proceedings of the IEEE*, Vol. 89, No. 3, March 2001, pp. 227–239.

55. Limits of integrated circuits manufacturing, Robert Doering and Yoshio Nishi, *Proceedings of the IEEE*, Vol. 89, No. 3, March 2001, pp. 375–393.

56. Materials And Process Limits In Silicon VLSI technology, James D. Plummer and Peter B. Griffin, *Proceedings of the IEEE*, Vol. 89, No. 3, March 2001, pp. 240–258.

57. New Paradigm of Silicon Technology, Tadahiro Ohmi, Shigetoshi Sugawa, Koji Kotani, Masaki Hirayama, and Akihiro Morimoto, *Proceedings of the IEEE*, Vol. 89, No. 3, March 2001, pp. 394–412.

58. Challenges of 22 nm and beyond CMOS technology, Huang Ru, Wu HanMing, Kang JinFeng, Xiao DeYuan, Shi XueLong, AN Xia, Tian Yu, Wang RunSheng, Zhang LiangLiang, Zhang Xing, and Wang YangYuan, *Science in China Series F: Information Sciences*, 52 (9), 2009, pp. 1491–1533.

59. High dielectric constant materials for nanoscale devices and beyond, Hiroshi Iwai, Akira Toriumi, and Durga Misra, *The Electrochemical Society Interface*, Winter 2017, pp. 77–81.

60. Silicon device scaling to the sub-10 nm regime, Meikei Ieong, Bruce Doris, Jakub Kedzierski, Ken Rim, and Min Yang, *Science*, Vol. 36, December 2004, pp. 2057–2060.

61. The end of Moore's Law: A new beginning for information technology, Thomas N. Theis and H.-S. Philip Wong, *Computing in Science & Engineering* (published by IEEE), Vol. 19, Issue 2, Mar-Apr. 2017, pp. 41–50.

62. Towards the limits of conventional MOSFETs: case of sub 30 nm NMOS devices, G. Bertrand, S. Deleonibus, B. Previtali, G. Guegan, X. Jehl, M. Sanquer, and F. Balestra, *Solid State Electronics* (48), 2004, pp. 505–509.

63. Nanoscale CMOS, Hon-Sum Philip Wong, David J. Frank, Paul M. Solomon, Clement H. J. Wann, and Jeffrey J. Welser, *Proceedings of the IEEE*, Vol. 87, No. 4, April 1999, pp. 537–570.

64. Beyond the conventional transistor, H.-S. P. Wong, *IBM Journal Research and Development*, Vol. 46, No. 2/3, March/May 2002, pp. 133–168.

65. High k/metal gate science and technology, Supratik Guha and Vijay Narayanan, *Annual Review of Materials Research*, 39, 2009, pp. 181–202.

66. Recent advances in high performance CMOS transistors: From planar to non-planar, Suman Datta, *The Electrochemical Society Interface*, Spring 2013, pp. 41–46.

67. Roadmap for 22 nm and beyond, H. Iwai, *Microelectronic Engineering*, 86, 2009, pp. 1520–1528.

68. Leakage current: Moore's Law meets static power, N. S. Kim, T. Austin, D. Baauw, T. Mudge, K. Flautner, J. S. Hu, M. J. Irwin, M. Kandemir, and V. Narayanan, *Computer* (published by IEEE), Vol. 36, Issue 12, Dec 2003, pp. 68–75.

69. Cool chips: Opportunities and implications for power and thermal management, Sheng-Chih Lin and Kaustav Banerjee, *IEEE Transactions on Electron Devices*, Vol. 55, No. 1, January 2008, pp. 245–255.

70. Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits, Kaushik Roy, Saibal Mukhopadhyay, and Hamid Mahmoodi-Meimand, *Proceedings of the IEEE*, Vol. 91, No. 2, February 2003, pp. 305–327.

71. MOSFET scaling into the 10 nm regime, Leland Chang and Chenming Hu, *Superlattices and Microstructures*, Vol. 28, No. 5/6, 2000, pp. 351–355.

72. Emerging applications for high k materials in VLSI technology, Robert D. Clark, *Materials*, 7, 2014, pp. 2913–2944.

73. Device scaling limits of silicon MOSFETs and their application dependencies, David J. Frank, Robert H. Dennard, Edward Nowak, Paul M. Solomon, Yuan Taur, and Hon-Sum Philip Wong, *Proceedings of the IEEE*, Vol. 89, No. 3, March 2001, pp. 259–288.

74. Challenges and design choices in nanoscale CMOS, Siva G. Narendra, *ACM Journal on Emerging Technologies in Computing Systems*, Vol. 1, No. 1, April 2005, pp. 7–49.

75. Semiconductor logic technology innovation to achieve sub-10 nm manufacturing, Klaus Schuegraf, Mathew C. Abraham, Adam Brand, Mehul Naik, and Randhir Thakur, *IEEE Journal of Electron Devices Society*, Vol. 1, No. 3, March 2013, pp. 66–75.

76. On-chip solid-state cooling for integrated circuits using thin-film microrefrigerators, Ali Shakouri and Yan Zhang, *IEEE Transactions on Components and Packaging Technologies*, Vol. 28, No. 1, March 2005, pp. 65–69.

77. High-end server low temperature cooling, R. R. Schmidt and B. D. Notohardjono, *IBM Journal of Research and Development*, Vol. 46, No. 6, November 2002, pp. 739–751.

78. Future of nano CMOS technology, Hiroshi Iwai, *Solid State Electronics*, 112, 2015, pp. 56-67.

79. On the scaling of subnanometer EOT gate dielectrics for ultimate nano CMOS technology, Hei Wong and Hiroshi Iwai, *Microelectronic Engineering*, 138, 2015, pp. 57–76.

80. MOSFET gate leakage modeling and selection guide for alternative gate dielectrics based on leakage considerations, Yee-Chia Yeo, Tsu-Jae King, and Chenming Hu, *IEEE Transactions on Electron Devices*, Vol. 50, No. 4, April 2003, pp. 1027–1035.

81. CMOS design near the limit of scaling, Y. Taur, *IBM Journal Research and Development*, Vol. 46, No. 2/3, March/May 2002, pp. 213–222.

82. Limits of lithography, Lloyd R. Harriot, *Proceedings of the IEEE*, Vol. 89, No. 3, March 2001, pp. 366–374.

83. Promising lithography techniques for next-generation logic devices, Rashed Md. Murad Hasan and Xichun Luo, *Nanomanufacturing and Metrology*, 2018, 167–181.

84. Advanced scanning probe lithography, Ricardo Garcia, Armin W. Knoll, and Elisa Riedo, *Nature Nanotechnology*, Vol. 9, August 2014, pp. 577–587.

85. Sub-10 nanometer feature size in silicon using thermal scanning probe lithography, Yu Kyoung Ryu Cho, Colin D. Rawlings, Heiko Wolf, Martin Spieser, Samuel Bisig, Steffen Reidt, Marilyne Sousa, Subarna R. Khanal, Tevis D. B. Jacobs, and Armin W. Knoll, *ACS Nano*, 2017, 11, pp. 11890–11897.

86. Computational lithography and computational metrology for nanomanufacturing, Shiyuan Liu, *Proceedings of the 2011 6th IEEE International Conference on Nano/Micro Engineered and Molecular Systems*, February 2011, pp. 1093–1099.

87. Review of computational lithography modeling: focusing on extending optical lithography and design-technology co-optimization, Kafai Lai, *Advances in Optical Technologies*, 1(4), 2012, pp. 249–267.

88. Lithography for enabling advances in integrated circuits and devices, C. Michael Garner, *Philosophical Transactions of the Royal Society A*, 370, 2012, pp. 4015–4041.

89. Lithography, metrology and nanomanufacturing, J. Alexander Liddle and Gregg M. Gallatin, *Nanoscale*, 3, 2011, 2679–2688.

90. MAPPER: High throughput maskless lithography, E. Slot, M. J. Wieland, G. D. Boer, P. Kruit, G. F. ten Berge, A. M. C. Houkes, R. Jager, T. van de Peut, J. J. M. Peijster, S. W. H. K. Steenbrink, T. F. Teepen, A. H. V. van Veen, and B. J. Kampherbeek, *Emerging Lithographic Technologies XII, Proc. of SPIE*, Vol. 6921, 2008, pp. 6921P1–6921P9.

91. Current challenges and opportunities for EUV lithography, Harry J. Levinson and Timothy A. Brunner, International Conference on Extreme Ultraviolet Lithography, *Proceedings of SPIE*, Vol 1089, 2018, pp. 1080903-1 to 1080903–7.

92. Nanoimprint lithography, Stephen Y. Chou, Peter R. Krauss, and Preston J. Renstrom, *Journal of Vacuum Science and Technology B: Microelectronics and Nanometer Structures Processing, Measurement and Phenomena*, 14, 4129, 1996, pp. 4129–4133.

93. Quantum interferometric optical lithography: Exploiting Entanglement to Beat the Diffraction Limit, Agedi N. Boto, Pieter Kok, Daniel S. Abrams, Samuel L. Braunstein, Collin P. Williams, and Jonathan P. Dowling, *Physical Review Letters*, Vol. 85, No. 13, September 2000, pp. 2733–2736.

94. 2 nm quantum optical lithography, E. Pavel, S. Jinga, E. Andronescu, B. S. Vasile, G. Kada, A. Sasahara, N. Tosa, A. Matei, M. Dinescu, A. Dinescu, and O. R. Vasile, *Optics Communications*, 291, 2013, pp. 259–263.

95. The physics and chemistry of the Schottky barrier height, Raymond T. Tung, *Applied Physics Review*, 1, 011304, 2014, pp. 011304-1–011304-54.

96. Technology options for reducing contact resistances in nanoscale metal-oxide semiconductor field-effect transistors, Yee-Chia Yeo, *IEEE 5th International Nanoelectronics Conference (INEC)*, 2013, pp. 128–131.

97. Schottky barrier height tuning via the dopant segregation technique through low-temperature microwave annealing, Chaochao Fu, Xiangbiao Zhou, Yan Wang, Peng Xu, Ming Xu, Dongping Wu, Jun Luo, Chao Zhao and Shi-Li-Zhang, *Materials*, 9, 315, 2016, pp. 1–10.

98. Study of the intrinsic limitations of the contact resistance of metal/semiconductor interfaces through atomistic simulations, A. Dabral, G. Pourtois, K. Sankaran, W. Magnus, H. Yu, A. de. Jamblinne de Meux, A. K. A. Lu, S. Clima, K. Stockbro, M. Schaekers, N. Collaert, N. Horiguchi and M. Houssa, *ECS Journal of Solid State Science and Technology*, 7(6), 2018, pp. N73–N80.

99. Full band calculations of the intrinsic lower limit of contact resistivity, J. Maassen, C. Jeong, A. Baraskar, M. Rodwell and M. Lundstrom, *Applied Physics Letters*, 102, 111605, 2013, pp. 111605-1–111605-4.

100. Source/drain junctions and contacts for 45 nm CMOS and beyond, Mehmet C. Ozturk and Jing Liu, *AIP Conference Proceedings*, 788, 2005, pp. 222–231.

101. Contact engineering for nano-scale CMOS, Muhammad Hussain, Hossain Fahad and Ramy Qaisi, *Physica Status Solidi A*, 209, No. 10, 2012, pp. 1954–1959.

102. Theory of workfunction control of silicides by doping for future Si-nano-devices based on fundamental physics of why silicides exist in nature, T. Nakayama, K. Kakushima, O. Nakatsuka, Y. Machida, S. Sotome, T. Matsuki, K. Ohmori, H. Iwai, S. Zaima, T. Chikyow, K. Shiraishi and K. Yamada, *IEEE International Electron Devices Meeting*, 2010, Dec 2010, pp. 15.5.1–15.5.4.

103. Advanced source/drain technologies for parasitic resistance reduction, Yee-Chia Yeo, *IEEE International Workshop on Junction Technology Extended Abstracts*, June 2010, pp. 1–5.

104. Low Schottky barrier height for silicides on n-type Si (100) by interfacial selenium segregation during silicidation, Hoong-Shing Wong, Lap Chan, Ganesh Samudra and Yee-Chia Yeo, *Applied Physics Letters*, 93, 072103, 2008, pp. 072103-1–3.

105. Fermi-level depinning for low-barrier Schottky source/drain transistors, Daniel Connelly, Carl Faukner, P. A. Clifton and D. E. Grupp, *Applied Physics Letters*, 88, 012105, 2006, pp. 012105–1-3.

106. Essential physics of carrier transport in nanoscale MOSFETs, Mark Lundstrom and Zhibin Ren, *IEEE Transactions on Electron Devices*, Vol. 49, No. 1, Jan. 2002, pp. 133–142.

107. Elementary scattering theory of the Si MOSFETs, Mark Lundstrom, *IEEE Electron Device Letters*, Vol. 18, No. 7, July 1997, pp. 361–363.

108. Theory of ballistic nanotransistors, Anisur Rahman, Jing Guo, Supriyo Datta and Mark S. Lundstrom, *IEEE Transactions on Electron Devices*, Vol. 50, No. 9, Sep. 2003, pp. 1853–1864.

109. On the interpretation of ballistic injection velocity in deeply scaled MOSFETs, Yang Liu, Mathieu Luisier, Amlan Majumdar, Dimitri A. Antoniadis and Mark S. Lundstrom, *IEEE Transactions on Electron Devices*, Vol. 50, No. 4, April 2012, pp. 995–1001.

110. Ballistic MOSFET reproduces current-voltage characteristics of an experimental device, Kenji Natori, *IEEE Electron Device Letters*, Vol. 23, No. 11, Nov. 2002, pp. 655-–657.

111. Heat generation and transport in nanometer-scale transistors, Eric Pop, Sanjiv Sinha and Kenneth E. Goodson, *Proceedings of the IEEE*, Vol. 94, No. 8, August 2006, pp. 1587–1601.

112. Ballistic/quasi-ballistic transport in nanoscale transistor, Kenji Natori, *Applied Surface Science*, Vol. 254, Issue 39, July 2008, pp. 6194–6198.

113. On apparent electron mobility in Si nMOSFETs from diffusive to ballistic regime, Dimitri Antoniadis, *IEEE Transactions on Electron Devices*, Vol. 63, No. 7, July 2016, pp. 2651–2656.

114. Physics-based compact model of nanoscale MOSFETs – part I: Transition from drift-diffusion to ballistic transport, Giorgio Mugnaini and Giuseppe Iannaconne, *IEEE Transactions on Electron Devices*, Vol. 52, No. 8, August 2005, pp. 1795–1801.

115. Physics-based compact model of nanoscale MOSFETs – Part II: Effects of degeneracy on transport, Giorgio Mugnaini and Giuseppe Iannaconne, *IEEE Transactions on Electron Devices*, Vol. 52, No. 8, August 2005, pp. 1802–1806.

116. Current transport in nanoelectronic semiconductor devices, V. Sverdlov, H. Kosina, and S. Selberherr, *IEEE Conference on Emerging Technologies-Nanoelectronics*, Jan. 2006, pp. 490–495.

117. Wafer-scale, sub-5 nm junction formation by monolayer doping and conventional spike annealing, Johnny C. Ho, Roie Yerushalmi, Gregoy Smith, Prashanth Majhi, Joseph Bennett, Jeffri Halim, Vladimir N. Faifer, and Ali Javey, *Nano Letters*, Vol. 9, No. 2, 2009, pp. 725–730.

118. Controlled nanoscale doping of semiconductors via molecular monolayers, Johnny C. Ho, Roie Yerushalmi, Zachery A. Jacobson, Zhiyong Fan, Robert L. Alley, and Ali Javey, *Nature Materials*, Vol. 7, Jan. 2008, pp. 62–67.

119. Exploring the limits of n-type ultra shallow junction formation, Craig M. Polley, Warrick R. Clarke, Jill A. Miwa, Giordano Scappucci, Justin W. Wells, David L. Jaeger, Maia R. Bischof, Richard F. Reidy, Brian P. Gorman, and Michelle Simmons, *ACS Nano*, Vol. 7, No. 6, 2013, pp. 5499–5505.

120. Ultra-shallow junction formation-physics and advanced technology, B. Colombeau, S. H. Yeong, D. X. M. Tan, A. J. Smith, R. M. Gwilliam, C. M. Ng, K. R. C. Mok, F. Benistant, and L. Chan, *AIP Conference Proceedings*, 1066, 11, 2008, pp. 11–18.

121. Atomically precise placements of single dopants in Si, S. R. Schofield, N. J. Curson, M. Y. Simmons, F. J. Rueβ, T. Hallam, L. Oberbeck, and R. G. Clark, *Physical Review Letters*, Vol. 91, No. 13, September 2003, pp. 136104-1-4.

122. Realization of atomically controlled dopant devices in silicon, Frank J. Rueβ, Wilson Pok, Thilo C. G. Reusch, Matthew J. Butcher, Kuan Eng J. Goh, Lars Oberbeck, Giordano Scappucci, Alex R. Hamilton, and Michelle Y. Simmons, *Small* 2007, 3, No. 4, pp. 563–567.

123. Recent advance in single-ion implantation method for single-dopant devices, Takahiro Shinada, Masahiro Hori, Keigo Taira, Tetsuo Endoh, and Iwao Ohdomari, 2009 *IEEE International Workshop on Junction Technology*, June 2009, pp. 96–99.

124. Control of both number and position of dopant atoms in semiconductors by single ion implantation, Takahiro Shinada, Tomonori Kurosawa, Takahiro Kobayashi, Hideki Nakayama,

and Iwao Ohdomari, *IEEE 2006 International Workshop on Junction Technology*, May 2006, pp. 16–20.

125. Toward atomic-scale device fabrication in silicon using scanning probe microscopy, Frank J. Ruess, Lars Oberbeck, Michelle Y. Simmons, Kuan Eng J. Goh, Alex R. Hamilton, Toby Hallam, Steven R. Schofield, Neil J. Curson, and Robert G. Clark, *Nano Letters*, Vol. 4, No. 10, 2004, pp. 1969–1973.

126. High mass molecular ion implantation, Bill Chang and Michael Ameen, Chapter 5, *Crystalline Silicon-Properties and Uses*, July 2011, pp. 81–104.

127. Single-ion implantation for the development of Si-based MOSFET devices with quantum functionalities, Jeffrey C, McCullum, David N. Jamieson, Changyi Yang, Andrew J. Alves, Brett C, Johnson, Toby Hopf, Samuel C, Thompson, and Jessica A. van Donkelaar, *Advances in Materials Science and Engineering*, Volume 2012, Article ID 272694, pp. 1–10.

128. Nanometre-scale electronics with III-V compound semiconductors, Jesús A. del Alamo, *Nature*, Vol. 479, November 2011, pp. 317–321.

129. Nanometer-scale III-V MOSFETs, Jesús A. del Alamo, Dimitri A. Antoniadis, JianQiang Lin, Wenjie Lu, Alon Vardi, and Xin Zhao, *Journal of Electron Devices Society*, Volume 4, No. 5, September 2016, pp. 205–214.

130. Main determinants for III-V metal-oxide-semiconductor field effect transistors (invited), Peide D. Ye, *Journal of Vacuum Science and Technology A*, 26 (4), July/August 2008, pp. 697–704.

131. Enhancing hole mobility in III-V semiconductors, Aneesh Nainani, Brian R. Bennett, J. Brad Boos, Mario G. Ancona, and Krishna C. Saraswat, *Journal of Applied Physics*, 111, 103706, 2012, pp. (currently does not have access to view pdf).

132. Device physics and performance potential of iii-v field-effect transistors, Yang Liu, Himadri S. Pal, Mark S. Lundstrom, Dae-Hyun Kim, Jesús A. del Alamo, and Dimitri A. Antoniadis, Chapter 3, *Fundamentals of III-V Semiconductor MOSFETs* (book), edited by: S. Oktyabrsky and Peide D. Ye, Springer Science + Business Media LLC, 2010, pp. 31–49.

133. III-V ternary and quaternary compounds, Sadao Adachi, Part D:30, Springer *Handbook of Electronic and Photonic Materials* (Book), edited by: S. Kasap and P. Capper, Springer International Publishing AG 2017, pp. 725–741.

134. The importance of ballistic resistance in the modeling of nanoscale inGaAs MOSFETs, Jianqiang Lin, Xiaowei Cai, Dimitri A. Antoniadis, and Jesús A. del Alamo, *IEEE Electron Device Letters*, Vol. 38, No. 7, July 2017, pp. 851–854.

135. III-V nanowires-extending a narrower road, Lars-Erik Wernersson, Claes Thelander, Erik Lind, and Lars Samuelson, *Proceedings of the IEEE*, Vol. 98, No. 12, December 2010, pp. 2047–2060.

136. Band parameters for III-V compound semiconductors and their alloys, I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, *Applied Physics Review*, *Journal of Applied Physics*, Vol. 89, No. 11, June 2001, pp. 5815–5875.

137. Ultrathin body InGaAs MOSFETs on III-V-on-insulator integrated with silicon active substrate (III-V-OIAS), Jianqiang Lin, Lukas Czornomaz, Nicolas Daix, Dimitri A. Antoniadis, and Jesús A. del Alamo, *IEEE Transactions on Electron Devices*, Vol. 63, No. 8, August 2016, pp. 3088–3095.

138. Investigation of source starvation in high-transconductance III-V quantum-well MOSFETs, Martin Rau, Jianqiang Lin, Dimitri A. Antoniadis, Jesús A. del Alamo and Mathieu Luisier, *IEEE Transactions on Electron Devices*, Vol. 66, No. 11, November 2019, pp. 4698–4705.

139. Physical model of incomplete ionization for silicon device simulation, Andreas Schenk, Pietro P. Altermatt, and Bernhard Schmithüsen, *IEEE 2006 International Conference on Simulation of Semiconductor Processes and Devices* (SISPAD), September 2006, pp. 51–54.

140. Device modeling at cryogenic temperatures: Effects of incomplete ionization, Akin Akturk, Jeffrey Allnutt, Zeynep Dilli, Neil Goldsman, and Martin Peckerar, *IEEE Transactions on Electron Devices*, Vol. 54, No. 11, Nov. 2007, pp. 2984–2990.

141. Source/drain doping effects and performance analysis of ballistic III-V n-MOSFETs, Raseong Kim, Uygar E. Avci, and Ian A. Young, *Journal of the Electron Devices Society*, Vol. 3, No. 1, Jan 2015, pp. 37–43.

142. GaN nanowire n-MOSFET with 5 nm channel length for applications in digital electronics, Nadim Chowdhury, Giuseppe Iannaccone, Gianluca Fiori, Dimitri A. Antoniadis, and Tomas Palacios, *IEEE Electron Device Letters*, Vol. 38, No. 7, July 2017, pp. 859–862.

143. Route to high hole mobility in GaN via reversal of crystal-field splitting, Samuel Ponce, Debdeep Jena, and Feliciano Giustino, *Physical Review Letters*, 123, 096602, 2019, pp. 0966602-1 to 5.

144. Frontiers of silicon-on-insulator, G. K. Celler and Sorin Cristoloveanu, *Applied Physics Reviews*-Focused Review, *Journal of Applied Physics*, Volume 93, No. 9, May 2003, pp. 4955–4978.

145. Ultimately thin double-gate SOI MOSFETs, Thomas Ernst, Sorin Cristoloveanu, Gerard Ghibaudo, Thierry Ouisse, Seiji Horiguchi, Yukinori Ono, Yasuo Takahashi, and Katsumi Murase, *IEEE Transactions on Electron Devices*, Vol. 50, No. 3, March 2003, pp. 830–838.

146. The new generation of SOI MOSFETs, J.-P. Colinge, *Romanian Journal of Information Science and Technology*, Vol. 11, No. 1, 2008, pp. 3–15.

147. On the extension of ET-FDSOI roadmap for 22 nm node and beyond, C. Sampedro, F. Gamiz, and A. Godoy, *Solid-State Electronics*, 90, 2013, pp. 23–27.

148. Evidence of mobility enhancement due to back biasing in UTBOX FDSOI high-k metal gate technology, I. Ben Akkez, C. Fenouillet-Beranger, A. Cros, F. Balestra, and G. Ghibaudo, *IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference* (S3S), October 2013, pp. 1–2.

149. Planar fully-depleted-silicon-on-insulator technologies: Toward the 28 nm node and beyond, B. Doris, B. DeSalvo, K. Cheng, P. Morin, and M. Vinet, *Solid State Electronics* 117, 2016, pp. 37–59.

150. Gate length and performance scaling of undoped-body extremely thin SOI MOSFETs, Amlan Majumdar, Xinlin Wang, Arvind Kimar, Judson R. Holt, David Dobuzinsky, Raj Venigalla, Christine Ouyang, Steven J. Koester, and Wilfried Haensch, *IEEE Electron Device Letters*, Vol. 30, No. 4, April 2009, pp. 413–415.

151. Comprehensive analysis of short-channel effects in ultrathin SOI MOSFETs, Qian Xie, Chia-Jung Lee, Jun Xu, Clement Wann, Jack Y. C. Sun, and Yuan Taur, *IEEE Transactions on Electron Devices*, Vol. 60, No. 6, June 2013, pp. 1814–1819.

152. Ultra-thin body and buried oxide SOI substrate development and qualification for Fully Depleted SOI device with back bias capability, Walter Schwarzenbach, Bich-Yen Ngyuen, Frederic Allibert, Christophe Girard, and Christophe Maleville, *Solid State Electronics*, 117, 2016, pp. 2–9.

153. Analysis of self-heating effects in ultrathin-body SOI MOSFETs by device simulation, Claudio Fiegna, Yang Yang, Enrico Sangiorgi, and Anthony G. O'Neill, *IEEE Transactions on Electron Devices*, Vol. 55, No. 1, Jan. 2008, pp. 233–244.

154. Strain: A solution for higher carrier mobility in nanoscale MOSFETs, Min Chu, Yongke Sun, Ummaheswari Aghora, and Scott E. Thompson, *Annual Review Materials Research*, 2009, 39, pp. 203–229.

155. Modeling of modern MOSFETs with strain, V. Severdlov, O. Baumgartner, T. Windbacher, and S. Selberherr, *Journal of Computational Electronics*, 8, 2009, pp. 192–208.

156. Understanding starin-induced drive-current enhancement in strained-silicon n-MOSFET and p-MOSFET, Stefan Flachowsky, Andy Wei, Ralf Illgen, Tom Herrmann, Jan Hontschel, Manfred Horstmann, Wilfried Klix, and Roland Stenzel, *IEEE Transactions on Electron Devices*, Vol. 57, No. 6, June 2010, pp. 1343–1354.

157. Exploring the limits of stress-enhanced hole mobility, Lee Smith, Victor Moroz, Geert Eneman, Peter Verheyen, Faran Nouri, Lori Washington, Malgorzata Jurczak, Oleg Penzin, Dipu Pramanik, and Kristin De Meyer, *IEEE Electron Device Letters*, Vol. 26, No. 9, September 2005, pp. 652–654.

158. pMOSFET with 200% mobility enhancement induced by multiple stressors, Lori Washington, Faran Nouri, Sunderraj Thirupapuliyur, Gert Eneman, Peter Verheyen, Victor Moroz, Lee Smith, Xiaopeng Xu, Mark Kawaguchi, T. Huang, Khaled Ahmed, Miheala Balseanu, Li-Qun Xia, Meihua Shen, Yihwan Kim, Rita Rooyackers, Kristin De Meyer, and Robert Schreutelkamp, *IEEE Electron Device Letters*, Vol. 27, No. 6, June 2006, pp. 511–513.

159. Mobility enhancement the next vector to extend Moore's Law, Nidi Mohta and Scott E. Thompson, *IEEE Circuits and Devices Magazine*, Sep./Oct. 2005, pp. 18–23.

160. The origin of electron mobility enhancement in strained MOSFETs, G. Hadjisavaas, L. Tsetseris, and S. T. Pantelides, *IEEE Electron Device Letters*, Vol. 28, No. 11, No. 2007, pp. 1018–1020.

161. Strained Si/strained Ge dual-channel heterostructures on relaxed Si0.5 Ge0.5 for symmetric mobility p-type and n-type metal-oxide-semiconductor field-effect transistors, Minjoo L. Lee and Eugene A . Fitzerald, *Applied Physics Letters*, 83, 4202, 2003.

162. High hole and electron mobilities using strained Si/strained Ge heterostructures, Saurabh Gupta, Minjoo L. Lee, Chris W. Leitz, and E. A. Fitzrerald, *Series/Report No. Advanced Materials for Micro- and Nano-Systems* (AMMNS), pp. 1–7.

163. Future of strained si/semiconductors in nanoscale MOSFETs, S. E. Thompson, S. Suthram, Y. Sun, G. Sun, S. Parthasarathy, M. Chu, and T. Nishida, 2006 *IEEE International Electron Devices Meeting*s, Dec 2006, pp. 1–4.

164. Band structure, deformation potentials and carrier mobility in strained Si, Ge and SiGe alloys, M. V. Fischetti and S, E. Laux, *Journal of Applied Physics*, 80, 4, August 1996, pp. 2234–2252.

165. Physics of hole transport in strained silicon MOSFET inversion layers, Everett X. Wang, Philippe Matagne, Lucian Shifren, Borna Obradovic, Roza Kotlyar, Stephen Cea, Mark

Stettler, and Martin D. Giles, *IEEE Transactions on Electron Devices*, Vol. 53, No. 8, August 2996, pp. 1840–1851.

166. Physics of strain effects in semiconductors and metal-oxide-semiconductor field-effect transistors, Y. Sun, S. E. Thompson, and T. Nishida, *Journal of Applied Physics*, 101, 104503, 2007, pp. 104503-1–104503-22.

167. Strained Si, SiGe and Ge channels for high-mobility metal-oxide-semiconductor field-effect transistors, Minjoo L. Lee, Eugene A. Fitzerald, Mayank T. Balsara, Mathew T. Curie, and Anthony Locthtefeld, *Applied Physics Reviews-Focused Review*, *Journal of Applied Physics*, 97, 011101, 2005, pp. 011101-1–011101-27.

168. A 90-nm logic technology featuring strained-silicon, Scott E. Thompson, Mark Armstrong, Chris Auth, Mohsen Alavi, Mark Buchler, Robert Chau, Steve Cea, Tahir Ghani, Glenn Glass, Thomas Hoffman, Cha-Hong Jan, Chris Kenyon, Jason Klaus, Kelly Kuhn, Zhiyong Ma, Brian Mcintyre, Kaizad Mistry, Anand Murthy, Borna Obradovic, Ramune Nagisetty, Phi Nguuyen, Sam Sivakumar, Reaz Shaeed, Lucian Shifren, Bruce Tufts, Sunit Tyagi, Mark Bohr, and Youssef El-Mansy, *IEEE Transactions on Electron Devices*, Vol. 51, No. 11, November 2004, pp. 1790–1797.

169. Strain effects on three-dimensional, two-dimensional, and one-dimensional silicon logic devices: Predicting the future of strained silicon, Mehmet O. Baykan, Scott E. Thompson, and Toshikazu Nishida, *Journal of Applied Physics*, 108, 093716, 2010, pp. 093716-1 to 093716–24.

170. Enhancement of electron mobility in ultrathin-body silicon-om-insulator MOSFETs with uniaxial strain, *IEEE Electron Device Letters*, Vol. 26, No. 5, May 2005, pp. 314–316.

171. Strained-Si CMOS technology, S. Takagi, *Advanced Gate Stacks For High Mobility Semiconductors*, Chapter 1, pp. 1–19.

172. Conventional technology boosters for injection velocity in ultrathin-body MOSFETs, M. Ferrier, R. Clerc, L. Lucci, Q, Rafhay, G. Pananakakis, G. Ghibaudo, F. Boeuf, and T. Skotnicki, *IEEE Transactions on Nanotechnology*, Vol. 6, No. 6, November 2007, pp. 613–621.

173. Sub-band structure engineering for advanced CMOS channels, Shin-ichi Takagi, T. Mizuno, T. Tezuka, N. Sugiyama, S. Nakaharai, T. Numata, J. Koga, and K. Uchida, *Solid State Electronics*, 49, 2005, pp. 684–694.

174. Continuous MOSFET performance increase with device scaling: The role of strain and channel material innovations, D. A. Antoniadis, I. Aberg, C. Nı´ Chle´irigh, O. M. Nafayeh, A. Khakifirooz, and J. L. Hoyt, *IBM Journal of Research and Development*, Vol. 50, No. 4/5, July/September 2006, pp. 363–376.

175. Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transistors, Isabelle Ferain, Cynthia A. Colinge, and Jean-Pierre Colinge, *Nature*, Vol. 479, November 2011, pp. 310–316.

176. Scaling theory for cylindrical, fully-depleted, surrounding-gate MOSFET's, Christopher P. Auth and James D. Plummer, *IEEE Electron Device Letters*, Vol. 18, No. 2, Feb. 1997, pp. 74–76.

177. A novel scaling theory for fully-depleted multiple-gate MOSFET, including effective number of gates (ENGs), Te-Kuang Chiang, *IEEE Transactions on Electron Devices*, Vol. 61, No. 2, February 2014, pp. 631–633.

178. High performance and highly uniform gate-all-around silicon nanowire MOSFETs with wire size dependent scaling, S. Bangsaruntip, G. M. Cohen, A. Majumdar, Y. Zhang, S. U. Engelmann, N. C. M. Fuller, L. M. Gignac, S. Mittal, J. S. Newbury, M. Guillorn, T. Barwicz, L. Sekaric, M. M. Frank, and J. W. Sleight, 2009 *IEEE International Electron Devices Meeting* (IEDM), December 2009, pp. 12.3.1–12.3.4.

179. Generalized scale length for two-dimensional effects in MOSFET's, David J. Frank, Yuan Taur, and Hon-Sum P. Wong, *IEEE Electron Device Letters*, Vol. 19, No. 10, Oct. 1998, pp. 385–387.

180. Can 3-D devices extend Moore's Law b[illegible]2 nm technology node?, Marius Orlowski and Andreas Wild, *ECS Transactions*, 3 (6), 2006, pp. 3–17.

181. Compact modeling of cross-sectional scaling in gate-all-around FETs: 3-D to 1-D transition, Avirup Dasgupta, Priyank Rastogi, Amit Agarwal, Chenming Hu, and Yogesh Singh Chauhan, *IEEE Transactions on Electron Devices*, Vol. 65, No. 3, March 2018, pp. 1094–1100.

182. FinFETs: From devices to architectures, Debajit Bhattacharya and Niraj K. Jha, *Advances in Electronics*, Hindawai Publishing Corporation, Volume 2014, Article ID 365689, pp. 1–21.

183. Vertical GAA FETs for the ultimate CMOS scaling, Dmitry Yakimets, Geert Eneman, Pieter Schuddinck, Trong Huynh Bao, Marie Garcia Bardon, Praveen Raghavan, Anabela Veloso, Nadine Collaert, Abdelkarim Mercha, Diederik Verkest, Aaron Voon-Yew Thean, and Kristin De Meyer, *IEEE Transactions on Electron Devices*, Vol. 62, No. 5, May 2015, pp. 1433–1439.

184. Impact of self-heating effect on transistor characterization and reliability issues in sub-10 nm technology nodes, Yi Zhao and Yiming Qu, *Journal of the Electron Devices Society*, Volume 7, 2019, pp. 829–836.

185. CMOS beyond Si: Nanometer-scale III-V MOSFETs, Jesús A. del Alamo, Xiaowei Cai, Jianqiang Lin, Wenji Lu, Alon Vardi, and Xin Zhao, *2017 IEEE Bipolar/BiCMOS Circuits and Technology Meeting* (BCTM), October 2017, pp. 25–29.

186. Silicon nanowire FinFETs, C. Mukherjee and C. K. Maiti, Chapter 8, *Nanowires-Recent Advances*, Intech Open Science Publishing, pp. 151–176.

187. Enhanced coupling effects in vertical double-gate FinFETs, Sung-Jae Chang, Maylin Bawedin, Yufeng Guo, Fanyu Liu, Kerem Akarvardar, Jong-Hyun Lee, Jung-Hee Lee, Irina Ionica, and Sorin Cristoloveanu, *Solid State Electronics*, 97, 2014, pp. 88–98.

188. Impact of LER and random dopant fluctuations on FinFET matching performance, Emanuele Baravelli, Malgorzata Jurczak, Nicolo Speciale, Kristin De Meyer, and Abhishek Dixit, *IEEE Transactions on Nanotechnology*, Vol. 7, No. 3, May 2008, pp. 291–298.

189. Turning silicon on its edge-overcoming silicon scaling barriers with double-gate and FinFET technology, Edward J. Nowak, Ingo Aller, Thomas Ludwig, Keunwoo Kim, Rajiv V. Joshi, Ching-Te Chuang, Kerry Bernstein, and Ruchir Puri, *IEEE Circuits and Devices Magazine*, January/February 2014, pp. 20–31.

190. Experimental validation of self-heating simulations and projections for transistors in deeply scaled nodes, E. Bury, B. Kaczer, P. Roussel, R. Ritzenthaler, K. Raleva, D. Vasileska, and G. Groeseneken, *2014 IEEE International Reliability Physics Symposium*, June 2014, pp. XT. 8.1–XT.8.6.

191. Scaling theory for FinFETs based on 3-d effects investigation, Wenwei Yang, Zhiping Wu, and Lilin Tian, *IEEE Transactions on Electron Devices*, Vol. 54, No. 5, May 2007, pp. 1140–1147.

192. FinFET-A Self-aligned double-gate MOSFET scalable to 20 nm, Digh Hishamoto, Wen-Chin Lee, Jakub Kedzierski, Hideki Takeuchi, Kazuya Asano, Charles Kuo, Erik Anderson, Tsu-Jae King, Jeffrey Bokor, and Chenming Hu, *IEEE Transactions on Electron Devices*, Vol. 47, No. 12, December 2000, pp. 2320–2325.

193. FinFET scaling to 10 nm gate length, Bin Yu, Leland Chang, Shibly Ahmed, Hailong Wang, Scott Bell, Chih-Yuh Yang, Cyrus Tabery, Chan Ho, Qi Xiang, Tsu-Jae King, Jeffrey Bokor, Chenming Hu, Ming-Ren Lin, and David Kyser, Digest. *International Electron Devices Meeting*, Dec. 2002, pp. 1–4.

194. Tunnel field-effect transistors: Prospects and challenges, Uygar E. Avci, Daniel H. Morris, and Ian A. Young, *Journal of the Electron Devices Society*, Volume 3, No. 3, May 2015, pp. 88–95.

195. Low-Voltage Tunnel Transistors for Beyond CMOS Logic, Alan C. Seabaugh and Qin Zhang, *Proceedings of the IEEE*, Vol. 98, No. 12, December 2010, pp. 2095–2110.

196. Drive current boosting of n-type tunnel FET with strained SiGe layer at source, Nayan Patel, A. Ramesha and Santanu Mahapatra, Microelectronics Journal, 39, 2008, pp. 1671–1677.

197. Investigation on the Effects of Gate-Source Overlap/Underlap and Source Doping Gradient of n-Type Si Cylindrical Gate-All-Around Tunnel Field-Effect Transistors, Kong-Ming Liu and Ching-Ping Cheng, *IEEE Transactions on Nanotechnology*, Volume 19, 2020, pp. 382–389.

198. High On-Current Ge-Channel Heterojunction Tunnel Field-Effect Transistor Using Direct Band-to-Band Tunneling, Garam Kim, Jaehong Lee, Jang Hyun Kim and Sangwan Kim, Micromachines, 2019, 10, 77, pp. 1–8.

199. A Review of Sharp-Switching Devices for Ultra-Low Power Applications, Sorin Cristoloveanu, Jing Wan and Alexander Zaslavsky, *Journal of the Electron Devices Society*, Volume 4, No. 5, September 2016, pp. 215–226.

200. Direct and Indirect Band-to-Band Tunneling in Germanium-Based TFETs, Kuo-Hsing Kao, Anne S. Verhulst, William G. Vandenberghe, Bart Soree, Guido Groeseneken and Kristin De Meyer, *IEEE Transactions on Electron Devices*, Vol. 59, No. 2, Feb 2012, pp. 292–301.

201. Controlling Ambipolar Current in Tunneling FETs Using Overlapping Gate-on-Drain, Dawit B. Abdi and M. Jagadesh Kumar, *Journal of the Electron Devices Society*, Vol. 2, No. 6, November 2014, pp. 187–190.

202. A Comparative Study on Scaling Capabilities of Si and SiGe Nanoscale Double Gate Tunneling FETs, Toufik Bentrda, Faycal Djeffal, Hichem Ferhad and Zohir Dibi, Silicon 2020, 12, pp. 945–953.

203. Impact of Strain on Drain Current and Threshold Voltage of Nanoscale Double Gate Tunnel Field Effect Transistor: Theoretical Investigation and Analysis, Sneh Saurabh and M. Jagadesh Kumar, Japanese *Journal of Applied Physics*, 48, 2009, 064503, pp. 064503-1–7.

204. Design and Simulation of Improved Swing and Ambipolar Effect for Tunnel FET by Band Engineering Using Metal Silicide at Drain Side, Shih-Chieh Teng, Yen-So Su and Yung-Hsien Wu, *IEEE Transactions on Nanotechnology* (Letters), Volume 18, 2019, pp. 274–278.

205. Scaling Length Theory of Double-Gate Interband Tunnel Field-Effect Transistors, Lu Liu, Dheeraj Mohata and Suman Datta, *IEEE Transactions on Electron Devices*, Vol. 59, No. 4, April 2012, pp. 902–908.

206. Modeling and TCAD Assessment for Gate Material and Gate Dielectric Engineered TFET Architectures: Circuit-Level Investigation for Digital Applications, Upasana, Rakhi Narang, Manoj Saxena and Mridula Gupta, *IEEE Transactions on Electron Devices*, Vol. 62, No. 10, Oct. 2015, pp. 3348–3356.

207. Modeling the single-gate, double-gate, and gate-all-around tunnel field-effect transistor, Anne S. Verhulst, Bart Soree, Daniele Leonelli, William G. Vandenberghe and Guido Groeseneken, *Journal of Applied Physics*, 107, 024518, 2010, pp. 024518-1-8.

208. The Tunnel Field Effect Transistor, Devin Verreck, Anne Verhulst and Guido Groeseneken, J. Webster(Ed.), Wiley Encyclopedia of Electrical and Electronic Engineering, 2016, John Wiley and Sons, pp. 1–24.

209. High Performance Drain Engineered InGaN Heterostructure Tunnel Field Effect Transistor, Xioaling Duan, Jincheng Zhang, Jiabo Chen, Tao Zhang, Jiaduo Zhu, Zhiyu Lin and Yue Hao, Micromachines 2019, 10, 75, pp. 1–9.

210. Tunnel FET technology: A reliability perspective, Suman Datta, Huichu Liu and Vijaykrishnan Narayanan, Microelectronics Reliability, 54, 2014, pp. 861–874.

211. Auger generation as an intrinsic limit to tunneling field-effect transistor performance, James T. Teherani, Sapan Agarwal, Winston Chern, Paul M. Solomon, Eli Yablonovitch and Dimitri A. Antoniadis, *Journal of Applied Physics*, 120, 084507, 2016, pp. 084507-1–15.

212. Subthreshold-swing physics of tunnel field-effect transistors, Wei Cao, Deblina Sarkar, Yasin Khatami, Jiahao Kang and Kaustav Banerjee, AIP Advances, 4, 067141, 2014, pp. 067141-1–9.

Author's Biography:

Dr. Nabil Shovon Ashraf was born in Dhaka, Bangladesh in 1974. Currently Dr. Ashraf serves as an Associate Professor in the department of Electrical and Computer Engineering of North South University, Dhaka, Bangladesh. From September 2014 till April 2018 (correction from previous biography June 2018), he served as a faculty in the department of Electrical and Computer Engineering of North South University Dhaka, Bangladesh in the rank of Assistant Professor. He obtained a Bachelor of Technology degree in Electrical Engineering from Indian Institute of Technology Kanpur, India in 1997. He obtained a Master of Science degree in Electrical Engineering from University of Central Florida Orlando, U.S. in 1999. He obtained his Doctor of Philosophy degree in Electrical Engineering from Arizona State University Tempe, U.S. in 2011. From December 2011-May 2014 he was a Post Doctoral Researcher in the department of Electrical Engineering of Arizona State University Tempe. He was employed as design engineer in RF Monolithics Inc., a surface acoustic wave (SAW) based filter design company in Dallas, Texas, from August 1999-March 2001. From October 2003 till June 2006 he served on the faculty as Assistant Professor of the department of Electrical and Electronic Engineering of Islamic University of Technology, Gazipur, Bangladesh. To-date Dr. Ashraf has published six peer-reviewed journal articles (two IEEE EDS) and about 15 international conference proceedings (three IEEE EDS). He is the author of two books on semiconductor device scaling by Morgan & Claypool publishers (Released in 2018 and 2016) on substrate temperature engineering based device performance improvement and contributed to two book chapters on interface trap induced threshold voltage fluctuations in presence of random channel dopants of scaled n-MOSFET at the invitations of highly accomplished international book editors. He was cited in *Who's Who in America* for Marquis on-line biographies of distinguished and eminent researchers for two consecutive years of 2015 (69th edition) and 2016 (70th platinum edition). In 2017 Dr. Ashraf became the recipient of Albert Nelson Marquis Lifetime Achievement Award honored by Marquis Who's Who. Dr. Nabil Shovon Ashraf currently serves as a reviewer for a number of MDPI published journals and *Journal of Applied Physics*. He specializes in the area of device physics and modeling analysis of scaled devices for enabling improved device performance at the scaled node of current MOSFET device architectures.