APPLICATION OF MOSFET'S FLOATING BODY EFFECTS IN DYNAMIC MEMORY AND SPIKING NEURAL NETWORKS

Submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

by

Dibyendu Chatterjee

Roll No. 134076002

Supervisor:

Prof. Anil Kottantharayil

Department of Electrical Engineering

INDIAN INSTITUTE OF TECHNOLOGY BOMBAY

April 2021

Dedicated

to my family

for their

unfathomable love, inexhaustible care & invaluable support ...

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Dibyendu Chatterjee
Roll no. 134076002
Department of Electrical Engineering
IIT BOMBAY

14^{th} April 2021

Abstract

In this thesis, two major aspects of the modern computing system are studied and investigated. The first one is dynamic random access memory (DRAM) which is the primary memory of any computing system and the second one is the electronic neuron, a fundamental building block of an artificial neural network which drives the concept of neuromorphic computing.

In a conventional DRAM cell, the storage capacitor fabrication is a big challenge for sub-100 nm technology nodes. Novel floating body RAM (FB-RAM) or one transistor DRAM (1T-DRAM) or zero-capacitor RAM (Z-RAM) can be a promising solution to the scalability issue of the storage capacitor in a conventional DRAM cell.

In a Z-RAM cell, we take advantage of the floating body effects (FBE) to store excess charges at the body of an n-channel silicon on insulator (SOI) MOSFET. But low retention time due to over the barrier leakage from the body to source/drain is the main concern with all-Si Z-RAM cells. To increase the retention time, silicon is replaced by TiO_2 at source/drain of an n-channel SOI MOSFET. Since TiO_2 is an n-type high bandgap semiconductor with a high valence band offset with silicon ($\Delta E_V \approx 2\ eV$), the over the barrier leakage is significantly reduced. This leads to the improvement in both sense margin and retention characteristic as compared to an all-Si Z-RAM cell. Using well calibrated TCAD simulations, we demonstrate low bias programming for the proposed Z-RAM cell, which is a major advantage from an application perspective. At low drain bias, hole storage is initiated by band to band tunnelling which is subsequently taken over by impact ionization. We predict a retention time of 2 s and 70 ms at $T = 300$ K and 358 K respectively for a device gate length of 30 nm. We have optimized the device design to obtain a write '0' time of 6 μs. Multiple non-destructive reading operation for the proposed Z-RAM cell is also demonstrated.

On the other hand, the hardware implementation of an artificial spiking neural network (SNN) requires two fundamental building blocks namely, an artificial neuron and an artificial synapse. In this thesis, by utilizing the floating body effects, an artificial electronic neuron using an n-channel bulk FinFET with an n+ buried layer has been demonstrated using well calibrated TCAD simulations. The proposed neuron is seen to have a spiking frequency in the MHz range which is five orders of magnitude higher than that of a biological neuron and energy per spike of 6.3 fJ which is the lowest reported till date for the integrate block of the neuron. This can be a potential building block of a spiking neural network.

Keywords: One transistor one capacitor dynamic random access memory (1T-1C DRAM), zero capacitor random access memory (Z-RAM), silicon on insulator (SOI), impact ionization (II), floating-body effects (FBE), technology computer aided design (TCAD), bulk finFET, neuron, spiking neural networks (SNN).

Acknowledgement

I would like to express my hearty thanks to many people who have made this work possible. Firstly, I am fortunate to have Prof. Anil Kottantharayil as my supervisor whose constant guidance through various inputs have helped me transform ideas to concrete solutions. His constant push for solving problems which really make a difference to the research community, is deeply inspiring. As a presenter in his journal club group, I have got enough opportunities to constantly improve my presentation skills which helped me to boost my confidence. I will cherish my association with him in future in my life.

I would also like to thank my doctoral advisory committee members, Prof. Swaroop Ganguly and Prof. Pradeep Nair for their critical feedback and valuable suggestions which helped me improve the quality of the work.

It was a pleasant experience to have supportive and friendly association with my group mates Robin, Poonam, Premsai, Sreejith, Rajashekar who have stood by me whenever I needed them. I am thankful to my seniors Sanchar and Kalaivani for useful interactions (formal/informal) and making my early days in IITB comfortable. My best wishes to my juniors Jayashree, Saima, Durgaprasad and Tarun. Apart from my group, I would also like to thank Subrat, Aniket, Bhaskar, Dipankar, Harsh, Sachin, Bitan, Binit and Jaswant for the unforgettable memories with them in IITB to cherish. I wish them success in their life.

Most importantly, I would like to express my thanks to my loving parents, my younger brother (Vanu), my wife (Shruti) and my grandparents for their patience and invaluable support that they have shown over the years. I am lucky to have them as my family without whom it was not possible.

There could be people who have helped me in various stages of my Ph.D. journey at IIT Bombay, but I might have failed to recall them at this point of time, but I hope they would be remembered at appropriate time to come in my life.

Contents

List of Figures

List of Tables

List of Abbreviations

A2RAM	Advanced 2 Random Access Memory
BJT	Bipolar Junction Transistor
BOX	Buried Oxide
BL	Bit Line
BTBT	Band to Band Tunnelling
CMOS	Complimentary Metal Oxide Semiconductor
CMP	Chemical Mechanical Planarization
CPU	Central Processing Unit
CVD	Chemical Vapor Deposition
DIBL	Drain Induced Barrier Lowering
DRAM	Dynamic Random Access Memory
eV	electron Volt
F	Feature size
FD-SOI	Fully Depleted Silicon on Insulator
FET	Field Effect Transistor
GIDL	Gate Induced Drain Leakage
HDD	Hard Disk Drive
IC	Integrated Circuits

II	Impact Ionization
ITRS	International Technology Roadmap for Semiconductors
I-V	Current-Voltage
LIF	Leaky Integrate and Fire
MSDRAM	Meta-stable Dip Random Access Memory
PD-SOI	Partially Depleted Silicon on Insulator
PCM	Phase Change Memory
RDF	Random Dopant Fluctuations
RRAM	Resistive Random Access Memory
RT	Retention Time
SDD	Solid State Drive
SEM	Scanning Electron Microscope
SM	Sense Margin
SOI	Silicon on Insulator
SNN	Spiking Neural Networks
SRAM	Static Random Access Memory
SRH	Shockley-Read-Hall
1T1C	One Transistor One Capacitor
1T-DRAM	One Transistor Dynamic Random Access Memory
TCAD	Technology Computer Aided Design
WL	Word Line
Z-RAM	Zero Capacitor Random Access Memory
Z2FET	Zero Slope Zero Impact Ionization Field Effect Transistor

List of Symbols

β	Parasitic BJT gain
C_d	Depletion capacitance in $F.cm^{-2}$
C_{ox}	Gate oxide capacitance in $F.cm^{-2}$
D_{it}	Interface trap density in $cm^{-2}.eV^{-1}$
ΔE_C	Conduction band offset in eV
ΔE_V	Valence Band offset in eV
ϵ_0	Electrical permittivity of free space in $F.m^{-1}$
E_a	Activation energy in $J.mol^{-1}$
E_C	Bottom of the conduction band energy in eV
E_f	Fermi energy level in eV
E_g	Band-gap energy in eV
E_V	Top of the Valence band energy in eV
f_0	Output spiking frequency in Hz
GaP	Gallium phosphide
HfO_2	Hafnium oxide
I_0	Read state '0' current in A
I_1	Read state '1' current in A
I_b	Base current in A

I_{ch}	Channel current in A
I_{DS}	Drain to source current in A
I_{in}	Input current to a post synaptic neuron in A
I_{sub}	Substrate current in A
I_{th}	Threshold current in A
k_B	Boltzmann constant ($1.38064852 \times 10^{-23}\ m^2.kg.s^{-2}.K^{-1}$)
L_{ch}	Channel length in nm
L_G	Gate length in nm
L_n	Diffusion length of electrons in nm
L_p	Diffusion length of holes in nm
m_0	Electron mass ($9.109 \times 10^{-31}\ kg$)
M	Multiplication factor
μ_e	Electron mobility in $cm^2.V^{-1}.s^{-1}$
μ_h	Hole mobility in $cm^2.V^{-1}.s^{-1}$
n	Electron concentration in cm^{-3}
N_A	Acceptor type doping concentration in cm^{-3}
N_C	Effective density of states in the conduction band in cm^{-3}
n_i	Intrinsic doping concentration in cm^{-3}
N_D	Donor type doping concentration in cm^{-3}
N_V	Effective density of states in the Valence band in cm^{-3}
p	Hole concentration in cm^{-3}
P	Ferroelectric polarization in $C.m^{-2}$
q	Charge in C
Q_h	Excess hole density in the body in cm^{-2}

ρ	Charge density in cm^{-2}
S_0	Surface recombination velocity in $cm.s^{-1}$
Si	Silicon
SiC	Silicon carbide
SiGe	Silicon germanium
SiO_2	Silicon dioxide
Si_3N_4	Silicon nitride
T	Temperature in K
TiO_2	Titanium dioxide
τ	Carrier lifetime in μs
t_{body}	Body thickness in nm
t_{box}	Buried dioxide thickness in nm
t_{ox}	Silicon dioxide thickness in nm
t_{Si}	Silicon body thickness in nm
V_{DS}	Drain to source voltage in V
v_{eff}	Effective electron velocity in $cm.s^{-1}$
V_{GS}	Gate to source voltage in V
V_{BG}	Back gate voltage in V
V_{BLH}	Bit line voltage high in V
V_C	Capacitor voltage in V
V_{FB}	Flat-band voltage in V
V_{FG}	Front gate voltage in V
V_{offset}	Offset voltage in V
V_{WLL}	Word line voltage low in V

V_{WLH}	Word line voltage high in V
v_{sat0}	Saturation velocity in $cm.s^{-1}$
$V_{storage}$	Voltage across the storage capacitor in V
V_S	Source voltage in V
V_D	Drain voltage in V
V_G	Gate voltage in V
V_B	Body potential in V
V_{BB}	Body supply voltage in V
V_{DD}	Drain supply voltage in V
V_{TH}	Threshold voltage in V
V_{THF}	Front channel threshold voltage in V
Φ_B	Barrier height in eV
W_d	Maximum depletion layer width in nm
ZrO_2	Zirconium oxide

Chapter 1

Introduction

In the last century, one of the greatest inventions in the field of science and technology was the transistor. The invention of bipolar junction transistor (BJT) [1] by William Shockley, John Bardeen, and Walter Brattain in the year of 1947 opened a new era in the field of solid-state physics. They were jointly awarded the Nobel prize for this extraordinary breakthrough in the year of 1956. Looking back in time, Julius Edgar Lilienfeld first patented the concept of field effect transistor (FET) [2] in the year of 1926. However, it took almost 30 years for a working FET to be made until Dawon Kahng and Martin M. (John) Atalla at Bell Labs made it possible in 1959. During the same time, in the year of 1958, Jack Kilby from Texas Instruments invented integrated circuits (IC) [3] where several such transistors can be integrated on the same plane of a silicon wafer. These initial breakthroughs lead the way to manufacture high performance and small size computers and electronic devices. The hunt for smaller, faster, and low power systems began. In 1965, Gordon Moore, co-founder of Intel made a prediction that would set the pace for our modern digital revolution. From careful observation of an emerging trend, Moore extrapolated that computing would dramatically increase in power, and decrease in relative cost, at an exponential pace. He predicted that the number of transistors that can be packed into a given unit of space will double about every two years, though the cost per transistor is halved. In the year of 1966 [4], the first one transistor/one capacitor dynamic random access memory (DRAM) cell was invented by Robert Dennard at IBM and it was patented in 1968. Prior to the existence of the Moore's law, a computer architecture was described and designed by John von Neumann which is called von Neumann architecture [5]. Von Neumann architecture gave an opportunity to the designers in the Moore's law era to exploit the ever-increasing processing power of the microprocessors to build various

complex computational systems. Modern computers are mostly based on von Neumann architecture. But this architecture is not sufficient for today's era of big data where most of the task is data intensive. The data path between the central processing unit (CPU) and the memory unit becomes a bottleneck for data transfer, referred to as the von Neumann bottle-neck [6]. Even though since the last couple of decades both the logic and memory transistors are scaled down following the Mooré's law, in recent years, performance improvement due to scaling becomes increasingly challenging. To improve the performance of computing systems in the so-called "big data" era, we must think beyond von Neumann architecture. The conventional approach to computation will face a barrier in the next couple of years. There are two major factors behind this situation. (i) Scaling will reach its fundamental limit (atomic) beyond which the device cannot be miniaturized, (ii) power dissipation due to various leakage currents. For sub-100 nm technology nodes, leakage is one of the main concerns to deal with. Even though the scaling increases the integration density, it also increases the leakage current which in turn increases the power dissipation. Leakage also affects the performance of the memory devices. With the increase in leakage currents, memory performance degrades in terms of data retention. Also, the storage capacitor scaling in a dynamic random access memory (DRAM) cell for sub-100 nm technology nodes becomes increasingly difficult. To deal with these problems, we need novel approaches, need to search for proper materials that can be introduced besides silicon.

In today's data-centric world, an enormous amount of data is generated every day. For efficient management of this huge amount of unstructured data, we need i) a faster memory system with very high integration density so that whenever we need to process the data, we can fetch it from the memory system and do processing and ii) an intelligent computing system which is self-learning, energy-efficient and can do parallel processing like the human brain.

In this thesis, with the application of floating body effects, two major aspects of the modern computing system are conceptualized and demonstrated using TCAD simulations. The first one is TiO_2 source/drain zero-capacitor random access memory (Z-RAM) which can be an alternative to the conventional one transistor one capacitor dynamic random access memory (1T1C-DRAM) and the second one is the bulk FinFET based electronic neuron, a fundamental building block of an artificial neural network which drives the concept of highly energy-efficient neuromorphic computing. Chapter-2 of this thesis provides the literature survey where the conventional 1T1C-DRAM and

its limitations are discussed which is followed by the discussion of Z-RAM (or floating-body RAM or 1T-DRAM) as a possible replacement to the conventional 1T1C-DRAM. Then a brief literature on neuromorphic computing is also provided. In Chapter-3, we have implemented a TiO_2 source/drain partially depleted (PD) silicon on insulator (SOI) MOSFET based zero capacitor random access memory (Z-RAM). In this chapter, we have shown the superiority of the proposed TiO_2 source/drain Z-RAM cell as compared to the all-Si Z-RAM cell in terms of sense margin and retention characteristics. Since TiO_2 source/drain PD-SOI based Z-RAM uses impact ionization based programming, it required comparatively high drain bias which causes not only high power dissipation but also long term reliability issues. To solve this issue, we have implemented a TiO_2 source/drain fully depleted (FD) silicon on insulator (SOI) MOSFET based zero capacitor random access memory (Z-RAM) using commercial TCAD tool (Sentaurus^{TM}) in Chapter-4. Here, we have used parasitic BJT based programming method where comparatively low biases are applied for the memory operations. At the end of this chapter, a benchmark comparison is done with the other Z-RAM cells as well as with the state-of-the-art conventional 1T1C-DTAM cell. Chapter-5 describes the implementation of a highly scalable and CMOS compatible bulk-FinFET based ultra-low energy artificial neuron for spiking neural network (SNN) in a commercial TCAD tool (Sentaurus^{TM}). Chapter-6 concludes the thesis with a summary and a discussion of the scope for future work.

Chapter 2

Literature Review

In today's era of "Big Data", a high density faster memory system and an intelligent computing system are necessary for information processing of unstructured data. In this chapter, we review the literature on different types of memory in the memory hierarchy specially conventional Dynamic Random Access Memory (DRAM), where we discuss DRAM's operating principle and the challenges in modern DRAM cell. We also review the floating body RAM or zero capacitor RAM (Z-RAM) as an alternative to conventional DRAM cell. In the last section of this chapter, we review the literature on neuromorphic computing or artificial neural networks specially spiking neural networks where we discuss why we need neural network based computing system, historical background and the advancement in neural network based computing.

2.1 Semiconductor Memory Hierarchy

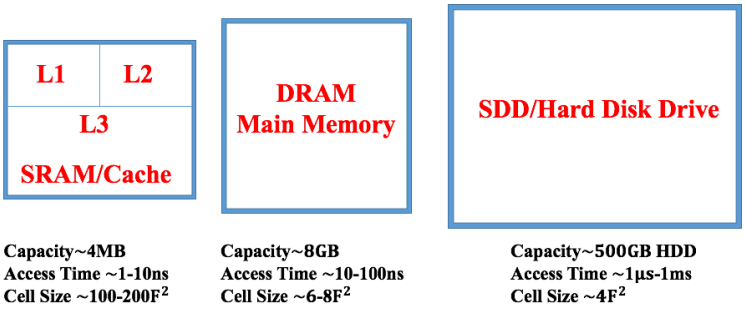

FIGURE 2.1: Semiconductor memory architecture.

If we divide a computer system into subsystems, then we will get two main parts. 1) computer memory where information or data is stored in the form of binary code and 2) Central Processing Unit (CPU) or logic unit which processes or operates on different data fetched from the memory whenever necessary. The basic unit of a memory which can store one bit is called memory cell. Based on proximity to the CPU, speed and volume, the memory is subdivided into three categories as shown in Fig. 2.1. First one is the Flash memory [7] or Solid State Drive (SSD) [8] or Hard Disk Drive (HDD) [9]. As the size of a unit cell of this kind of memory is $4F^2$, this is the densest memory inside a computer. The storage size of a HDD can be 512 GB - 1024 GB or even bigger. This memory is the furthest from the CPU and lowest in terms of speed. The access time of this memory is in the range of $1\ \mu s$ - $1\ ms$. A CPU cannot communicate with this memory directly as logic transistors inside a CPU are much faster. The access time of a logic transistor is in the range of $1ps - 1ns$. So, to bridge this gap between CPU and HDD, two more types of memory were introduced. i) Static Random Access Memory (SRAM) [10, 11] or cache memory which is the nearest to the CPU and fastest in terms of speed. Access time of this type of memory is in the range of $1\ ns$ - $10\ ns$. Since it is directly embedded to the CPU and one unit cell of it consists of six transistors with a cell size of $100F^2$, there is a limit of how big it can be. It's storage size is limited to 4 MB - 8 MB. Cache memory is subdivided into L1 cache, L2 cache, and L3 cache with increasing speed and proximity to the CPU. Cache is the costliest among all memories discussed. ii) Dynamic Random Access Memory (DRAM) [12] which is the most important type of memory in the whole memory architecture has a speed in the range of $10\ ns$ - $100\ ns$ and a cell size of $6F^2$. DRAM capacity can be as large as 32 GB. For certain applications, DRAM is embedded into the CPU and it is called embedded DRAM. The cell size of embedded DRAM is $30 - 50F^2$ which is still less as compared to the cell size of a SRAM cell [13] . So DRAM plays a crucial role in a computing system by bridging the gap between the CPU and the SSD or HDD.

As can be seen from Fig. 2.2(a), total memory IC market is expected to increase in the upcoming years after a dip in the year of 2019. DRAM and flash (NAND and NOR) have captured 98% of the total memory industry. DRAM alone has captured 53% [14] where as the market share of flash is 45% as shown in Fig. 2.2(b). DRAM market is expected to grow in the upcoming years.

Fig. 2.3 shows the sales and revenue growth of leading IC product segments in 2020 [14]. DRAM leads in terms of sales with $\$65,215$ million and flash leads in terms of the revenue growth percentage with 25%.

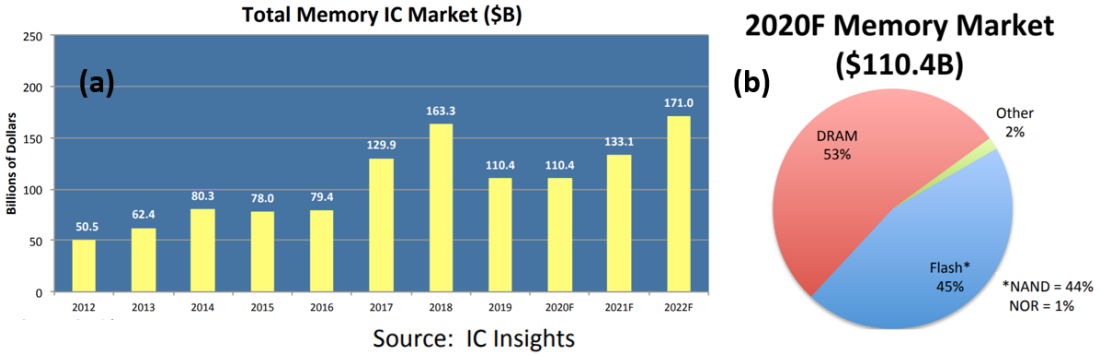

FIGURE 2.2: (a) Total memory IC market in billions. (b) Market share of DRAM. Figure taken from reference [14].

Leading IC Product Segments in 2020
Sales and Revenue Growth

Rank	Sales	$M	Revenue Growth	20/19 % Chg
1	DRAM	$65,215	NAND Flash	25%
2	NAND Flash	$55,154	Cellphone Application MPUs	24%
3	Computer CPU	$43,848	Wired Comm—Spcl Purp Analog	20%
4	Computer and Periph—Spcl Purp Logic	$31,340	Computer and Periph—Spcl Purp Logic	15%
5	Cellphone Application MPUs	$26,615	Wireless Comm—Spcl Purp Logic	12%

FIGURE 2.3: Sales and revenue growth of leading IC product segments in 2020 [14].

To continue the high market share, it becomes essential to keep on scaling the DRAM cell size until it reaches the fundamental limit after which it can not be scaled down further. To continue the scaling further below 100 nm node, new ideas are required and new materials to be introduced. In the following sections, the conventional DRAM cell is introduced and its limitations are discussed for sub-100 nm technology nodes. Floating-body (FB) RAM or 1T-DRAM or Z-RAM cell is introduced as a possible replacement of a conventional DRAM cell. Then the limitations of all-Si Z-RAM cell are discussed. To overcome the limitations of an all-Si Z-RAM cell, we proposed a TiO_2 source/drain based Z-RAM cell which is discussed in greater detail in Chapter 3 and Chapter 4.

2.2 Conventional 1T-1C DRAM

Conventional DRAM cell [12, 15] has a very simple structure as compared to a complex six transistor SRAM cell [10, 11]. It consists of an access transistor which acts as a

switch between the input and the output node and a storage capacitor as shown in Fig. 2.4.

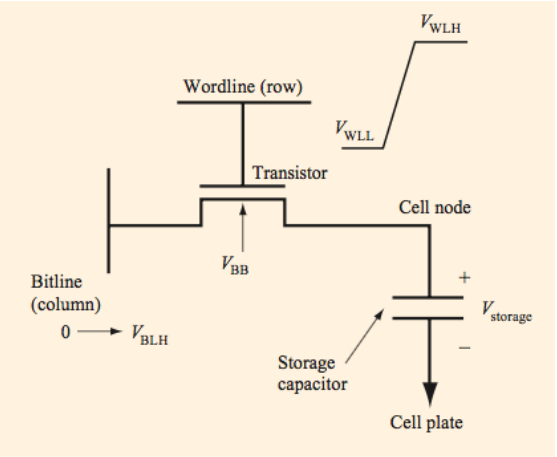

FIGURE 2.4: Conventional one transistor one capacitor DRAM Cell. Each cell consists of one transistor as a switch and a capacitor as storage node. V_{BLH}, V_{WLH}, V_{WLL}, V_{BB} are the bit-line high voltage, word-line high voltage, word-line low voltage, and body supply respectively. $V_{storage}$ is the voltage across the storage capacitor. This schematic is taken from reference [12].

The capacitor is used as a storage node. When the capacitor is charged, it is termed as logic 'state 1' and when there is no charge stored in the capacitor, it is termed as logic 'state 0'. As shown in Fig. 2.4, gate and drain are connected to the word line and bit line respectively. The operating principle of a conventional DRAM cell is discussed in the next section.

2.2.1 Operating Principle of a Conventional DRAM Cell

There are three operations which are performed by any type of memory. First, we erase/program some data in the memory and then we read it whenever required. For a conventional DRAM cell, these operations are performed in the following way. For programming or writing logic 'state 1', first, the bit line capacitor is fully charged to the supply voltage V_{DD}, and a positive gate voltage (V_{DD}) is applied to turn the access transistor on. Once the access transistor is turned on, a part of the bit line charge will pass through the transistor and charge the storage capacitor.

Now to read logic 'state 1', an intermediate voltage (in general $V_{DD}/2$) is applied at the bit line and the transistor is turned on by applying a positive gate voltage (V_{DD}). If the storage capacitor is charged, it transfers some part of it to the bit line. It increases

the bit line voltage slightly which is sensed by a sense amplifier and it gives an output of logic 'state 1'.

In the same way if the storage capacitor is not charged then some part of the bit line charge is transferred to the storage capacitor. This reduces the bit line voltage slightly and it is sensed by a sense amplifier to give an output of logic 'state 0'.

To erase or write logic 'state 0', data stored in the storage capacitor has to be removed. This can be done by first discharging the bit line capacitor then turning on the access transistor by applying V_{DD} to the gate terminal so that the stored charge in the storage capacitor gets drained by the bit line capacitor.

2.2.2 Challenges in Modern 1T-1C DRAM

With continuous scaling, the number of transistors per chip doubles in every technology generation. The major drawback of the scaling of a transistor is the leakage [12, 16]. With the shrinking of device dimensions, the gate oxide thickness reduces which increases the gate leakage, and at the same time the interaction between the source and drain increases with scaling which causes source to drain leakage due to drain induced barrier lowering (DIBL) effect. Further, the capacitor in a DRAM cell is not ideal. It leaks charges. That is why the capacitor must be refreshed after every read cycle. This leakage cannot be tolerated in a DRAM cell as it limits the charge storage capacity of the capacitor. There is an important parameter that must be specified and that is the retention time. The length of time during which a sufficient amount of charge can be retained in the storage capacitor to distinguish between logic state '1' and logic state '0' is termed as the retention time. As per International Technology Roadmap for Semiconductors (ITRS) [17], the retention time of a DRAM cell should be minimum $64\ ms$ at $358\ K$. This performance specification is constant for every technology generation. Now to achieve $64\ ms$ of retention time and to have sufficient signal to noise ratio, the capacitance of the storage capacitor in a DRAM cell should be at least $30\ fF$ [18]. But even though it is possible to scale down the access transistor of a DRAM cell below a certain technology node, the storage capacitor cannot be scaled down as we need $30\ fF$ of minimum capacitance to achieve $64\ ms$ of retention time and enough signal to noise ratio to distinguish between state '1' and state '0'. One way to get rid of the further scaling of the storage capacitor issue is to use of high-K dielectrics (HfO_2, ZrO_2, etc.) [19] instead of SiO_2 to increase the capacitance with the same device

dimensions. Another way is to increase the surface area of the capacitor by making a trench capacitor or stacked capacitor structure.

2.2.2.1 Trench Capacitor

As shown in Fig. 2.5 , to increase the surface area, a deep trench [20–22] is formed into Si by anisotropic etching process. Then a thin high-k dielectric is deposited followed by the deposition of metal electrode.

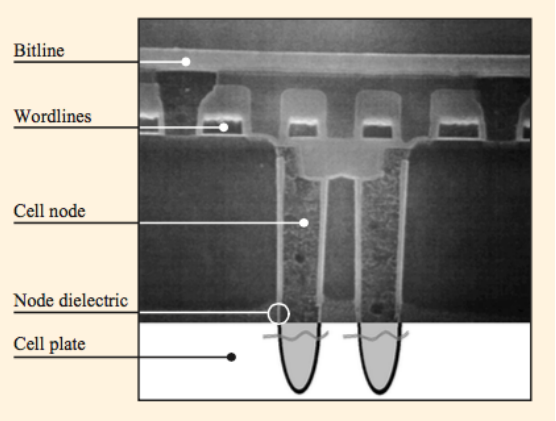

FIGURE 2.5: SEM photomicrograph of 0.25-μm trench DRAM cell suitable for scaling to 0.15 μm and below. Figure taken from reference [12].

One of the advantages of this type of structure is that even after the trench formation, the silicon surface is planer unlike the stack capacitor which is discussed in the next section.

The problem with this kind of structure is that, with the technology node, the trench has to be deeper which is a challenging task as we go further into the lower technology nodes. We cannot do high temperature processes once the capacitor dielectric is deposited as high-k dielectrics cannot sustain very high temperatures. Another problem is trap or defect generation [23] during the trench formation. These defects can act as a leakage path for the charge stored in the capacitor leading to lower retention time.

2.2.2.2 Stacked Capacitor

Another way to increase the surface area of the capacitor is by stacked capacitor structure [24, 25] as shown in Fig. 2.6.

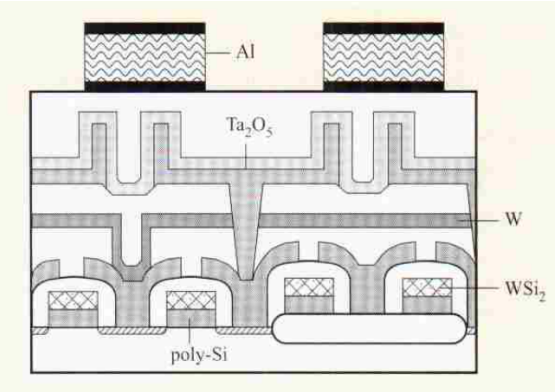

FIGURE 2.6: Schematic cross section of stacked capacitor cell suitable for $0.15\ \mu m$. Figure taken from reference [12].

There are several disadvantages with this kind of structure also. As the stacked capacitor is fabricated on top of the transistor, it is not planer and it can collapse during the subsequent process steps which may lead to the shortening of different parts. The aspect ratio also needs to be increased with every technology generation.

2.2.2.3 Sub-100 nm Technology Charge Leakage Issues:

As the device technology shrinks below 100 nm, several types of leakage appear (Fig. 2.7) in the device which can degrade the retention characteristic of the memory cell. These leakages are: 1) band to band tunnelling leakage (I_1) [16, 26], 2) sub-threshold leakage (I_2) [16, 27], 3) gate dielectric leakage (I_3) [16, 28], 4) gate induced drain leakage (GIDL) (I_5) [16, 29, 30], 5) leakage due to drain induced barrier lowering (I_6) (DIBL) [16], 6) leakage due to non-ideal storage capacitor.

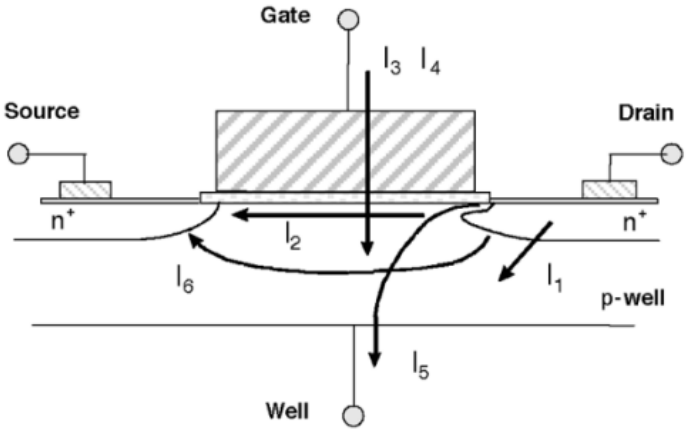

FIGURE 2.7: Summary of leakage current mechanisms of deep sub-micrometer transistors. Figure is taken from reference [16].

2.3 Floating-Body RAM Can be an Alternative

One transistor floating body RAM or zero capacitor RAM (1T-DRAM or Z-RAM) [35–42] cell can be a good alternative to the conventional one transistor one capacitor DRAM (1T-1C DRAM). Since 1T-DRAM and Z-RAM are the same, both the names are used interchangeably in this thesis. In a Z-RAM cell, the body of an n-channel SOI MOSFET is used as the storage node unlike the capacitor of a conventional DRAM cell. Since the body is floating, some floating body effects [43–48] arise at the input and output characteristics of the SOI MOSFET. These floating body effects are discussed in detail in the next section. There are two types of SOI MOSFETs, namely partially depleted (PD) SOI MOSFET and fully depleted (FD) SOI MOSFET. In a partially depleted (PD)-SOI MOSFET, as the depletion region does not cover the entire body thickness, there will be a quasi-neutral region at the body where the excess carriers (holes for an n-channel device) can be stored. The presence of excess holes at the body can be assigned the logic state '1' and the absence of excess holes can be assigned the logic state '0'. Excess holes can be generated either by impact ionization [49, 50] or by band to band tunnelling [51–53]. This can turn on the parasitic BJT [54, 55] leading to an increase in the drain current. The increase in the current can be used for reading logic state '1'. The difference between the logic state '1' (I_1) and logic state '0' (I_0) read currents is called the sense margin ($SM = \Delta I = I_1 - I_0$). The retention time specification is one of the most important figure of merit for a DRAM cell and it should be a minimum of $64\ ms$ at $T = 358\ K$ as per ITRS [17]. The details about 1T-DRAM or Z-RAM cell is discussed later in this chapter.

2.3.1 Floating Body Effects in SOI MOSFET

Although SOI technology has several advantages over bulk planer technology, it has some serious parasitic effects. Since there is a buried oxide layer in between the substrate and the transistor body, the body of an SOI MOSFET is floating i.e., it is not connected to ground, unlike the bulk planer MOSFET. There are several floating body effects [43–48] which are commonly seen in a PD-SOI MOSFET as it has a quasi-neutral region at the bottom of the silicon body layer. Floating body effects cause several problems in SOI MOSFET such as kink effect [43], hysteresis effect [44], parasitic BJT Effect [45].

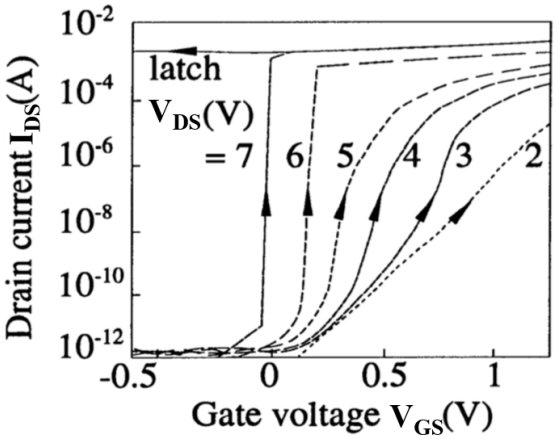

FIGURE 2.8: Experimental drain current (I_{DS}) versus gate voltage (V_{GS}) characteristics in SOI MOSFETs illustrating the sub-threshold slope steepening and the single transistor latch. Figure taken from reference [44].

2.3.1.1 Hysteresis Effect

Hysteresis or latch effect [44] is seen in the sub-threshold region of the transfer characteristics (I_{DS} vs V_{GS}) of an SOI MOSFET. Hysteresis phenomena can be explained by floating body effects. When the drain voltage is high enough, impact ionization occurs. Impact ionization causes electron-hole pair generation at the drain-body depletion region. The excess majority carrier holes are accumulated at the floating body and the excess minority carrier electrons go to the drain contact due to the presence of a large electric field at the drain-body junction. The accumulation of majority carriers at the body region causes the body potential to increase which leads to the reduction in threshold voltage and increase in the drain current. The increase in drain current causes more impact ionization at the drain junction and more holes are accumulated at the body region. This is a positive feedback process that occurs when the impact ionization current at the drain junction is larger than the drain to body leakage current, leads to a sharp increase in the sub-threshold drain current and the sub-threshold slope become almost $0\ mV/decade$. During the back sweeping of the gate bias, the high impact ionization current due to high drain voltage make the body potential high which in turn keeps the threshold voltage low to maintain the inversion layer and high drain current and a hysteresis appears until the positive feedback dies down. Fig. 2.8 shows the hysteresis effect in a PD-SOI MOSFET.

From Fig. 2.8, it is seen that the width of the hysteresis increases with the increase in drain bias. For a large drain bias, it is very difficult to turn the device off even if the gate to source voltage goes large in the negative direction. The reason for this is, for a

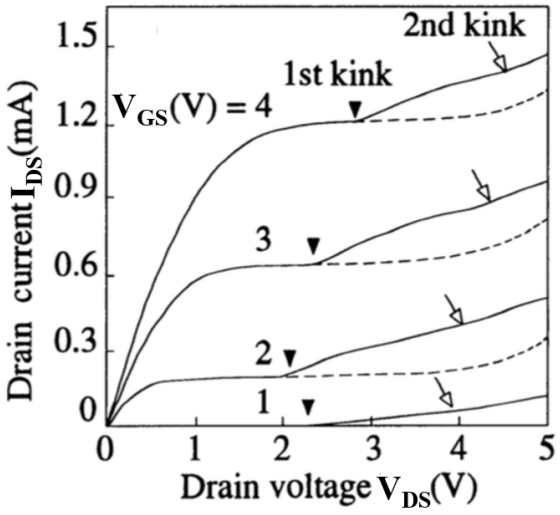

FIGURE 2.9: Experimental drain current (I_{DS}) versus drain voltage (V_{DS}) characteristics in SOI MOSFETs showing kink effect(solid line), which may be cancelled by grounding the body(dotted line). Figure taken from reference [43].

large drain bias the accumulation of majority carriers at the body is so large that it will sustain there for a longer amount of time and we get a very large hysteresis width. To get back to the zero state, the drain bias should be reduced to a very small value.

2.3.1.2 Kink Effect

Kink Effect [43] is seen in the saturation region of the output characteristics (I_{DS} vs V_{DS}) of an n-channel PD-SOI MOSFET. In a strong inversion region, when the drain voltage is large enough to cause impact ionization, the electron-hole pairs are generated at the drain to body junction and the generated holes enter into the floating body and stay there. As a result, the potential energy of the floating body increases and the threshold voltage decreases and at the same time the source to body barrier lowers which causes more and more minority carriers to enter into the channel and as a result of that, more impact ionization will take place at the drain junction which causes the drain current to rise and a kink (Fig. 2.9) appears in the output characteristics(I_{DS} vs V_{DS}). The extent of this kink is limited by the recombination of the majority carriers at the floating body. Eventually when the accumulated holes at the body are sufficient enough to make the base-emitter junction of the parasitic BJT forward biased a second kink (Fig. 2.9) appears which is seen in the short channel SOI MOSFET.

As shown in Fig. 2.9, for a larger gate bias the kink shifts towards right. This is due to the fact that, with the increase in gate bias, the resultant electric field at the drain

slows down the impact ionization process, which delays the onset of kink. Since in a fully depleted SOI MOSFET, there is no neutral region at the body for the accumulation of holes. So, the kink effect is not seen here.

2.3.1.3 Parasitic BJT Effect

The impact ionization current is much more important in SOI MOSFET as compared to bulk MOSFET. There is a parasitic BJT [45] in an SOI MOSFET in which the source acts as an emitter, floating body acts as the base and the drain as a collector of the parasitic BJT. When a high drain bias is applied, if the accumulated hole at the floating body is large enough to make the parasitic BJT turned on, extra minority carriers enter into the channel from source (emitter). The number of minority carriers enter into the channel depends on the gain (β) of the parasitic BJT, which can be given as follow:

$$\beta \approx 2(\frac{L_n}{L})^2 - 1 \tag{2.1}$$

Where, L_n is the diffusion length of minority carriers, and L is the channel length.

The collector current (βI_B) of the parasitic BJT contributes to the drain current and augments the impact ionization. The body (base) current I_B can be given as follow:

$$I_B = (I_{ch} + \beta I_B)(M-1) = \frac{I_{ch}(M-1)}{1-\beta(M-1)} \tag{2.2}$$

Where, M is the multiplication factor.

Therefore the total drain current is

$$I_D = M(I_{ch} + \beta I_B) = \frac{I_{ch}M}{1-\beta(M-1)} \tag{2.3}$$

The above equations show that the drain current increases both due to the impact ionization and parasitic BJT effect. The breakdown occurs when

$$1 - \beta(M-1) = 0 \tag{2.4}$$

So, from the above condition for breakdown, it is clear that premature breakdown depends on the parasitic BJT current gain.

2.3.2 Operating Principle of Floating Body RAM or Z-RAM

As there is no contact at the body of an SOI transistor (Fig. 2.10), the body is floating. Since there is a buried oxide (SiO_2) in between the transistor body and the substrate, there is a huge valence band offset between Si and SiO_2.

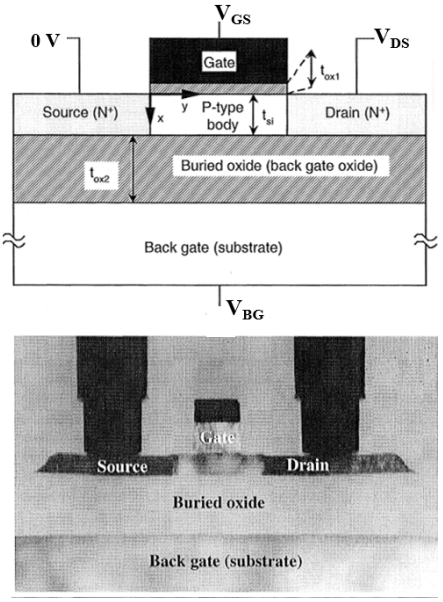

FIGURE 2.10: Schematic cross section of an n-channel SOI MOSFET. V_{GS}, V_{DS}, and V_{BG} represent the gate to source voltage and drain to source voltage, and back gate voltage. t_{si}, t_{ox1}, and t_{ox2} are the silicon body thickness, gate oxide thickness and buried oxide thickness respectively. Bottom is the SEM cross-section image of an actual SOI MOSFET [56].

Fig. 2.11 (a), (b), and (c) shows the energy band diagram in the vertical direction of a bulk, PD-SOI, and FD-SOI MOSFET respectively. Since the thickness of the body of an FD-SOI MOSFET is less than the depletion layer thickness, the entire body is depleted and there is no quasi-neutral region unlike in PD-SOI MOSFET as can be seen from Fig. 2.11.

Once the excess carriers are generated at the drain-body junction, excess holes come to the body region and excess electrons go to the drain contact due to drain to body depletion electric field. Excess holes at the body face a huge barrier due to large valence band offset between Si and SiO_2 and cannot escape through the substrate. So, these excess holes can be stored at the body for some time till they recombine in the body or at the body-dielectric/channel-dielectric interfaces or leak out from the body to source/drain. Depending on whether the excess holes are present at the body or not, the memory states are decided.

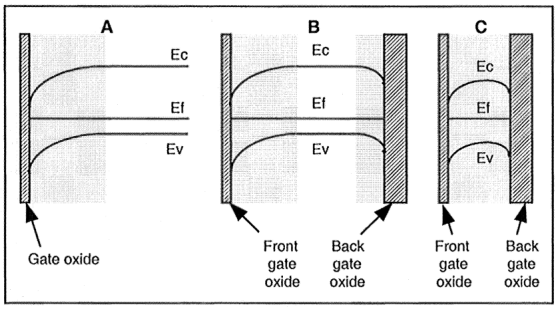

FIGURE 2.11: Energy band diagrams in (a) bulk, (b) partially depleted SOI and (c) fully depleted SOI. All devices are represented at threshold (front gate voltage = threshold voltage). The shaded areas represent the depleted zones. SOI devices are represented for a condition of weak inversion (below threshold) at the back interface [57]. E_C, E_V, and E_i are the conduction band energy, valence band energy and Fermi energy levels respectively.

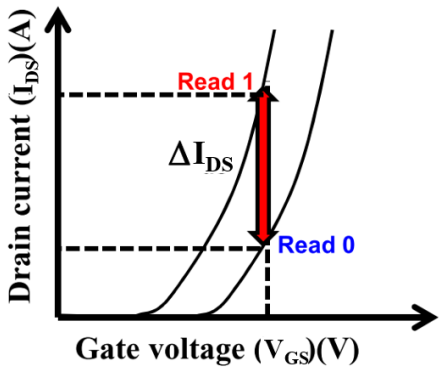

FIGURE 2.12: State '1' and state '0' are differentiated by two different threshold voltages in the transfer characteristics of the device [58]. V_{GS} and I_{DS} are the gate to source voltage and drain to source current respectively. ΔI_{DS} is the drain current difference between read state '1' and read state '0'.

The presence of excess holes at the body represents state '1' and the absence of excess holes represents states '0'. While in state '1', due to the presence of excess holes at the body, the body potential increases which reduces the threshold voltage of the device and we get a higher current. So, the two states are distinguished by the two different threshold voltages as shown in Fig. 2.12. The methods of excess carrier generation are discussed in the next section.

2.3.2.1 Different Programming Mechanisms

There are three different ways by which excess carriers are generated or the programming or writing state '1' can be done. These three programming methods are discussed below. In this thesis, we focus only on the n-channel device with a p-type body. Excess carriers

in this case refer to holes. The same concepts can be applied for the p-channel 1T-DRAM cell with an n-type body region.

Impact ionization based programming

One of the earliest method for state '1' programming is by impact ionization. First generation of Z-RAM cells was programmed using impact ionization mechanism [49, 50, 59]. In this method, excess holes are generated by impact ionization. In impact ionization based programming, the transistor operates in the inversion mode, i.e., a positive front gate bias (V_{GF}) which is greater than the front channel threshold voltage (V_{THF}) is applied and a highly positive drain bias (V_{DS}) is applied. Since the drain bias is sufficiently high to cause impact ionization at the drain-body junction, electron-hole-pairs are generated at the depletion region of the drain-body junction. Due to a large junction electric field, the excess electrons go to the drain contact and the excess holes come to the body region.

While programming, the front-gate bias V_{GF} remains unchanged. The hold and read operations also use the same front gate bias, while the drain bias V_{DS} is shifted from a low value (in the range of a mV) to a higher value (high enough to cause impact ionization). Since the body/drain junction is reverse biased, the excess holes which are generated by impact ionization flow to the body, resulting in an increase in the floating body potential above the level it had before the programming stage. V_{DS} is switched back to its low level (in the range of a mV) for state '1' reading (or hold). Since the excess generated holes stay inside the body, there will be a change in the body potential (ΔV_B) which is seen during writing state '1' or programming. A lowering of the "dynamic" V_{THF} is observed which increases the drain current. During the reading/holding, the body potential is comparatively higher due to higher number of stored holes than its steady-state value and the stored holes are gradually leaked out through the body-to-source/drain junctions. This leads to a drain current overshoot or returns to a steady state with time. Notice that the reading is non-destructive since it is performed at a low drain voltage (in the range of a mV).

It is clear that the high drain bias during write state '1' is a concern in the first generation of Z-RAM cells. Over time, it can degrade the gate oxide quality through hot carrier injection, which leads to the generation of interface states at the Si-SiO_2 interface, and eventually to premature oxide breakdown. A higher density of interface states at the Si-SiO_2 interface can degrade the retention time due to higher recombination.

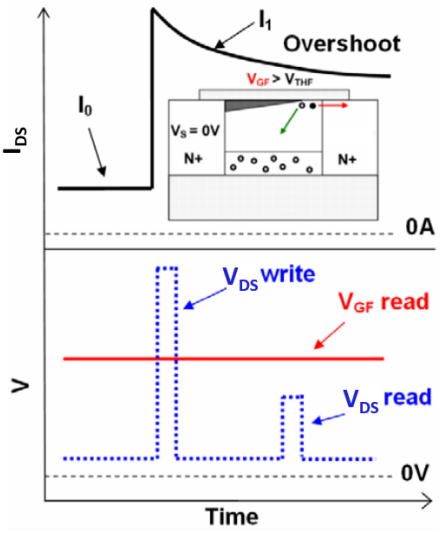

FIGURE 2.13: The front gate and drain biasing sequences and schematics for both the write/read state '1' programming by impact ionization. V_{DS} and V_{GF} are the drain to source voltage and front gate voltage respectively. V_{THF} is the front channel threshold voltage. I_1 and I_0 are the read state '1' and read state '0' currents respectively. The image is taken from reference [59].

Furthermore, to achieve faster programming, impact ionization rate should be increased which means higher drain voltage and higher power consumption.

Parasitic BJT based programming

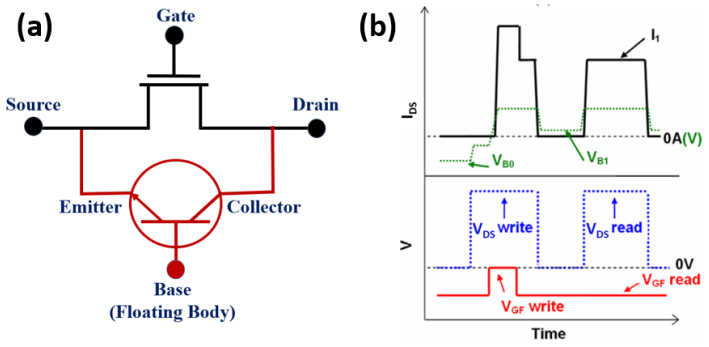

FIGURE 2.14: (a) Parasitic $n^+ - p - n^+$ bipolar junction transistor inside the n-channel SOI MOSFET. (b) The drain and front gate biasing sequences and schematics for both the write/read state '1' for parasitic BJT based programming. V_{DS} and V_{GF} are the drain to source voltage and front gate voltage respectively. V_{THF} is the front channel threshold voltage. I_1 and I_0 are the read state '1' and read state '0' currents respectively. V_{B0} and V_{B1} are the body potential during read state '1' and read state '0'. The image is taken from reference [59].

The problems of the first generation Z-RAM cell are solved up to a certain extent in the second generation of Z-RAM cells where the parasitic BJT (Fig. 2.14(a)) inside

an n-channel SOI MOSFET is used to write state '1' [54, 55, 59, 60]. In this case, source (N^+) acts as an emitter, body (P) as a base and drain (N^+) as a collector of the parasitic BJT inside the MOSFET. This parasitic BJT is turned on by increasing body potential through the accumulation of holes which is initiated by impact ionization at the drain-channel junction. Once the BJT is turned on, more electrons from the source (emitter of the parasitic BJT) will enter the channel (base of the parasitic BJT), and impact ionization rate increases at the drain-body junction and this is a regenerative process which eventually leads to latch-up of current. The increase in the current can be used for reading logic state '1'. The biasing mechanism for parasitic BJT based programming and corresponding drain current are shown in Fig. 2.14(b).

Band to band tunnelling based programming

Band to band tunnelling based programming [59, 61, 62] consumes the lowest power [62] among all the programming methods. In this method, the excess carriers are generated by gate induced drain leakage (GIDL) which is a band to band tunnelling mechanism at the gate-drain overlap region. The biasing scheme for this programming method is shown in Fig. 2.15(a). Here, the transistor operates in the accumulation mode i.e., a negative bias is applied at the gate and a positive bias is applied at the drain contact. The interface of the overlap region between gate and drain is inverted and a sheet of holes is created at the interface between gate and drain. The amount of holes at the interface is much higher which causes a band bending at the drain side, sufficient enough for band to band tunnelling of electrons from the valence band of the drain to the conduction band of the drain leaving behind the holes at the interface. Since a large electric field is present at the positively charged sheet and the drain, the holes come out from that region and populate the body.

At the onset of programming, since a large negative front gate bias is applied ($V_{GF} << V_{FB}$), the body potential decreases by dynamic gate coupling and becomes negative (Fig. 2.15(b)). The accumulation layer holes cannot be supplied within a quick time. Initially, the interface is depleted and it takes some time to go into accumulation from depletion. Excess holes generated by GIDL, start to fill the body gradually. As a result, the body potential keeps on increasing until it reaches a steady state when the transistor completely enters into accumulation mode. Programming time can be reduced by increasing the BTBT rate. BTBT rate can be enhanced by increasing the potential difference between the gate and drain terminal ($|V_G - V_D|$). Since the BTBT current is

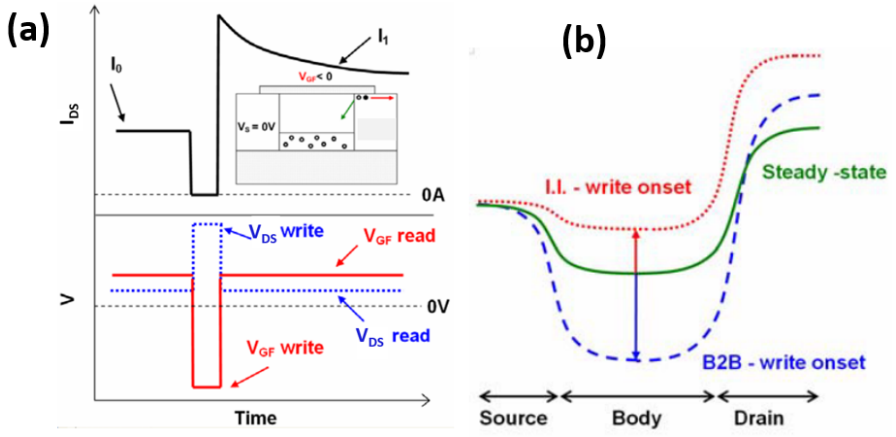

FIGURE 2.15: (a) The drain and front gate biasing sequences and schematics for both the write/read state '1' for band to band tunnelling based programming. V_{DS} and V_{GF} are the drain to source voltage and front gate voltage respectively. I_1 and I_0 are the read state '1' and read state '0' currents respectively. (b) Comparison between the impact ionization and the band to band tunnelling injection methods in terms of body potential variations during the programming. The image is taken from reference [59].

much lower as compared to the drain current generated by the impact ionization process, it consumes low power which is one of the advantages of BTBT based programming.

2.3.2.2 Erasing Mechanism

There are two different ways by which excess holes at the body region can be erased or writing state '0' is performed. These two methods are discussed below.

Erasing by forward biasing the drain-body junction

Excess carriers which are generated during the programming can be removed by forward biasing the drain-body junction [35]. In this method, a negative drain bias is applied as shown in Fig. 2.16. The drain-body depletion layer barrier reduces due to the applied forward bias at the drain terminal and the accumulated holes at the body which are generated during the programming stage cross the barrier and goes to the drain contact.

Since the time response of the forward body/drain junction to the bias switch is nearly instantaneous, erasing by forward bias is a reliable and fast. The resulting body potential variation is inefficient as the state '0' and the state '1' current nearly the same which results in the high power consumption during the read operation.

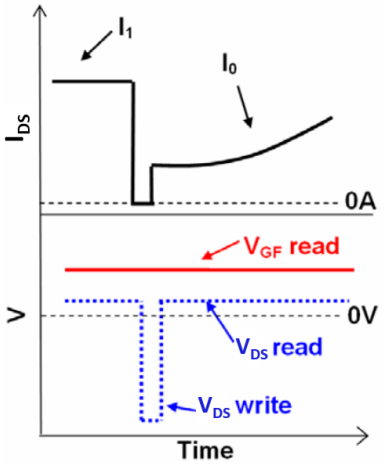

FIGURE 2.16: The drain and front gate biasing sequences and schematics for both the write/read state '0' for forward bias based erasing. V_{DS} and V_{GF} are the drain to source voltage and front gate voltage respectively. I_1 and I_0 are the read state '1' and read state '0' currents respectively. This image is taken from reference [59].

Erasing by capacitive coupling

A hole accumulation layer is created at the front interface when the front gate bias V_{GF} is lower than the front flat-band voltage V_{FB}. A fast removal can be performed by applying a front-gate pulse higher than V_{FB}. When V_{GF} increases, the body potential follows due to capacitive coupling. With the increase in body potential, the front interface becomes depleted and the body/drain and/or body/source junctions become "forward biased". Therefore, the excess holes are removed through the body/drain and/or body/source junctions [36, 63]. While reading, a negative bias is applied at the gate and as a result the body potential decreases to a negative value by capacitive coupling. The body potential is pinned to its negative value as the holes cannot be accumulated instantly. Depending on the hole charging via the junction leakage current, the steady-state during the reading is returned. Leakage is detrimental as it can change the retention time by increasing the state '0' drain current.

In Fig. 2.17, to prevent the parasitic BJT activation during read, body potential V_B is decreased by the capacitive coupling method. For a compatibility with the whole BJT method, the drain bias is kept at a high positive value during erase. In order to avoid the the parasitic BJT turn on, a positive source bias has to be applied . On the other hand, we cannot apply too high source bias in order to allow the excess holes to be removed from the body.

Difficulty arises in choosing proper combination of programming and erasing methods to achieve high performances, i.e. low programming voltage, high retention time as well

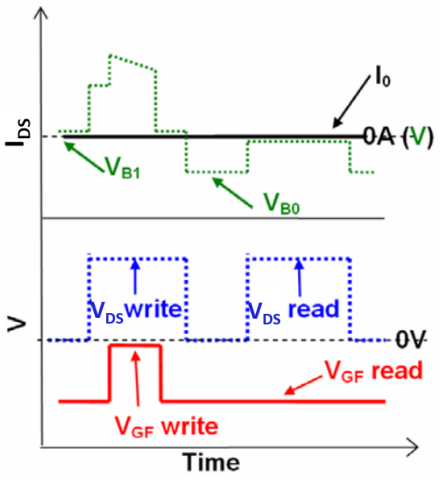

FIGURE 2.17: The drain and front gate biasing sequences and schematics for both the write/read state '0' for capacitive coupling based erasing. V_{DS} and V_{GF} are the drain to source voltage and front gate voltage respectively. I_1 and I_0 are the read state '1' and read state '0' currents respectively. V_{B0} and V_{B1} are the body potential during read state '1' and read state '0'. This image is taken from reference [59].

as sense margin and at the same time the best compatibility with a specific technology (PD-SOI, FD-SOI, double-gate, FinFET).

2.3.2.3 Reading Operation

After programming and erasing the Z-RAM cell, the states need to be read. Biasing schemes for read operation are shown for all the program (impact ionization based (Fig. 2.13), parasitic BJT based (Fig. 2.14) and band to band tunnelling based (Fig. 2.15) programming) and erase (forward biasing (Fig. 2.16) and capacitive coupling (Fig. 2.17)) methods. During the read operation, impact ionization is avoided by reducing the drain bias.

2.3.3 Limitations of all-Si Z-RAM cells

The retention time specification is one of the most important figure of merit for a DRAM cell and it should be a minimum of $64\ ms$ at $T = 358\ K$ as per ITRS [17]. For an all-Si Z-RAM cell, the retention time is less than $64\ ms$ [64, 65] which does not meet the ITRS specification. The reasons being the leakage of excess holes from the floating body to source/drain due to low barrier height for holes and Shockley-Read-Hall (SRH) recombination in the body. As the body is charged by excess injected holes, the barrier height reduces, leading to an unfavourable compromise between the

retention time and the sense margin. To overcome this problem, few hetero-structure based Z-RAM cells [66–69] have been proposed. In all these Z-RAM cells, the barrier height for holes at the body is increased either by using a high bandgap semiconducting material at the source/drain with a large valence band offset with Si [66, 67, 69] or by introducing a comparatively lower bandgap material like $Si_{1-x}Ge_x$ [68] at the body of an SOI MOSFET. Write '0' time is an important parameter for hetero-structure based Z-RAM cell as with the increase in valence band offset, write '0' time increases. Write '0' process and the time required for the same are not elaborated in the literature.

Other all-Si based Z-RAM cell architectures, namely meta stable dip DRAM (MSDRAM) [70], A2RAM [71] and zero-slope zero-impact ionization (Z2FET) DRAM [72], have been proposed for the possible replacement of the conventional DRAM cell. MSDRAM [70] makes use of the dynamic coupling between the front and back interfaces of a double gate FD-SOI MOSFET to give rise to hysteresis in the $I_{DS} - V_{GS}$ characteristics. MSDRAM suffer from high power dissipation due to a constant high positive voltage supply at the back gate and for ultra-thin SOI, electrons and holes cannot coexist at the same body due to super coupling effect. A2RAM [71] is similar to MSDRAM. But, instead of an electro-statically doped channel at the back, a physically n-doped channel is created below the p-type body. The fabrication process for A2RAM is complex. Variability for ultra-thin SOI A2RAM, and high power dissipation due to comparatively high biases used for programming are other disadvantages. Z2FET [72] is a lateral $p^+ - i - n^+$ diode with one top gate and either a bottom gate or fixed positive oxide charge created by CVD deposited SiO_2 at the top ungated portion to create a barrier for holes. For reliable operation of the device, precise control of the positive fixed oxide charge (Q_S) for low channel length device is desirable. For a back gated device, where there is no fixed oxide charge, a sufficiently large back gate voltage is required to create a barrier at the ungated region, leading to high power dissipation.

To overcome the limitations of different all-Si Z-RAM cells discussed above, we have utilized the floating body effects to demonstrate, an n-channel TiO_2 source/drain SOI MOSFET based dynamic memory cell. Chapter 3 and chapter 4 describe the applications of floating body effects in demonstrating the TiO_2 source/drain based Z-RAM cell.

2.4 Literature Review on Neuromorphic Computing

In almost every aspect of today's life, whether it is science and technology, entertainment and communications or process control, computer is essential. At present, around 10-15 % of the total global energy [73] is consumed in some form of information transmission, manipulation or processing. Since the data generation rate is increasing day by day, the energy consumption is also going to increase in the near future if the traditional von Neumann computer architecture [5, 73] is continued. We need a computer architecture which is more energy efficient. Neural network based architecture [74, 75] is currently one of the most energy efficient computer architecture. Since early 1990s, research in neuromorphic computing started to increase. Neuromorphic computing [76, 77] is brain inspired computing which is a million times more power and energy efficient than that time's best computing devices. Though, the traditional computing had achieved remarkable feats in those times, they failed to do some of the basic tasks like image and speech recognition which a biological system can do with ease. Hence, the ideas from the biological system were required for the implementation of a computing system which can do things like image and speech recognition with much less power consumption.

In recent times, we have seen unprecedented advancement in CMOS technology. Some of the semiconductor industry giants like TSMC and SAMSUNG have already announced 5 nm and even 3 nm CMOS technology [78] to revive the Moor's law. This progress in CMOS technology made a revolution in parallel processing. For example, parallel processing is inevitable in millions of cell phones nowadays, personal computers are having multiple processors, the top supercomputers in the world are having millions of CPU counts. But in today's data-centric world where every day over 2.5 quintillion [79] bytes of data are generated, it is very difficult to manage these data. It is only going to grow from there. By 2020, it is estimated that 1.7 MB of data [79] will be created every second for every person on earth. Using traditional computing systems, this huge amount of data can not be managed. Based on solid engineering and scientific data, it is predicted that traditional computing will face a huge challenge within the next ten years. There are two main reasons for this. First of all, present CMOS technology will reach its fundamental limit after which it can not be miniaturized further, and secondly, there will be huge power dissipation (20-30 megawatts [73] for exascale computing) by the traditional computing systems while properly managing these hugely generated data. Even though today's computing is very powerful but they fail to handle this huge amount of complex and unstructured data.

There are two approaches [73] to tackle this problem. One is the software approach where "machine learning" [80] algorithm is used to tackle these huge numbers of complex and noisy datasets which the traditional non-learning algorithms are not able to tackle. It is a rapidly growing field. At present day, this "machine learning" approach is so popular that almost every computing and internet company have opened a new branch on "machine learning". Even all the major universities and research institutes across the world have their R and D in this area. The second approach is the hardware implementation [81, 82] of neural networks. In this approach, behaviour of a biological neuron is mimicked through hardware. In the software approach, the conventional processor is optimized for "machine learning" algorithms to mimic the behaviour of a biological neuron. Even though this optimized processor is 120 times more power-efficient [73] than a general-purpose processor, they are not fundamentally different from an existing CPU. So the hardware approach is in line. In the early 1980s, researchers across the world have started working to model the behaviour of a biological neuron through analog CMOS based circuits. One of the greatest examples of the advancement in the hardware implementation of a neuromorphic computing system is IBM's "True North" chip [83] which is biologically inspired and it consists of one million of spiking neurons and 256 million synapses. There are 5 billion transistors in this chip. Even though this biologically inspired chip consumes only 65 milliwatts of power, it is still not as power-efficient as the biological system is. So, more research and innovation are required in this field.

2.4.1 Von Neumann versus Neuromorphic Computing

Traditional computer architecture follows the principle of von Neumann architecture [5] as shown in Fig. 2.18. The system with von Neumann architecture is divided into several components like central processing unit (CPU), arithmetic logic unit (ALU), memory unit (MU), and data paths. All these units are separate entities and not embedded which makes the whole system slow and less energy efficient.

These constraints limit the future development of this architecture. Even though the conventional von Neumann architecture based parallel computer consists of thousands and millions of traditional processors connected to each other increases the computational power several times, the basic processor is the same as used in a single serial computer. It limits the further development of von Neumann architecture based computing system.

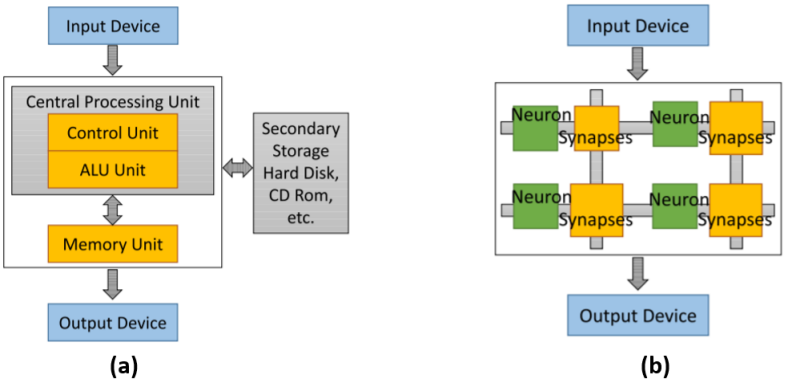

FIGURE 2.18: Comparison of high-level conventional von Neumann and neuromorphic computer architectures. (a) Schematic of a von Neumann architecture. The so-called "von Neumann bottleneck" [6] is the data path between the CPU (consists of Arithmetic Logic Unit (ALU) and Control Unit) and the memory unit. (b) Schematic of a basic concept of a neuromorphic architecture. A neural network-based architecture combines synapses and neurons into a fine grain distributed structure that scales both memory (synapse) and compute elements (soma) elements as the systems increase in scale and capability, thus avoiding the bottleneck between computing and memory. The schematics are taken from reference [84].

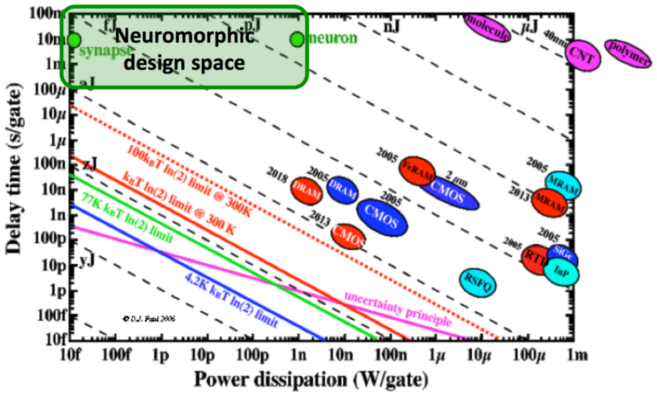

FIGURE 2.19: Delay time per transistor versus the power dissipation plot. The operating regime for neuromorphic devices is in the upper left corner indicating the extremely low power dissipation of biological synapses and the corresponding delay time. Systems built in this region would be more "brain-like" in their power and cycle times. The image is taken from reference [73].

On the other hand, the neuromorphic system functions in a completely different way which is very high power-efficient, self-learning and works in the principle of the human brain. One of the major advantages of a neural network [85] or neuromorphic computing over von Neumann or traditional computing is its ability to do highly power efficient [73] parallel processing. Delay versus power dissipation curve in Fig. 2.19 shows the difference between neuromorphic system and present-day technologies. It clearly shows the advantages of neuromorphic system or brain-inspired computing over the other technologies in terms of density and power efficiency.

2.4.2 Recent Advancement in Neuromorphic Systems

The concept of bio-inspired computing has been there for more than eighty years [86, 87]. Biologically inspired algorithms of neural computation are referred to as Artificial Neural Networks (ANNs). It can mimic the process by which the human brain acquires and processes sensory information. ANNs consist of two fundamental building blocks. 1) Artificial neuron, and 2) artificial synapse which is the connection between two neurons. Using these two building blocks information is processed according to a specified set of rules and equations which are called models of information processing in neural circuits. Neurons can be connected to each other through synapses by different ways which give rise to different models [88, 89]. Advances in consolidating the vast number of new findings and insights from neuroscience into such computational models in a biologically plausible way have been largely lacking in the mainstream ANN community.

While it has been known for long that neurons communicate with spikes (electrical pulses, action potentials), it was only in the early $1990's$ when studies found evidence for biological brains making use of the exact timing of single spikes to encode information [90, 91]. This observation gave rise to SNNs after their property of explicitly modeling individual spikes, rather than average firing rates like their predecessors. The utilization of spikes brings together the definitions of time varying Post-Synaptic Potential (PSP), firing threshold (ϑ), and spike latencies (Δ), as depicted in Fig. 2.20. They try to simulate the processes carried out between the neurons (synapses) in a network. More than two decades have passed since one of the most influential papers on the topic was published [92]. Nowadays, SNNs can still be considered a niche of artificial intelligence research.

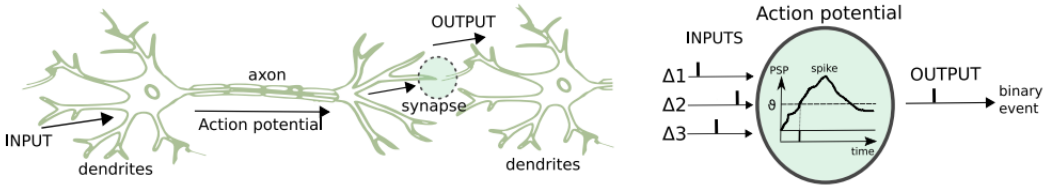

FIGURE 2.20: Biological neuron and its association with an artificial spiking neuron [93].

Spiking Neural Networks (SNNs) [94] is the new generation of ANNs [92] which are popular for its ability to capture the informational dynamics seen in biological neurons. In describing the realistic biological information processing, SNNs theory is

mostly accepted. SNNs are not only more realistic but also they are easy to implement on reliable hardware platforms because of their "integrate-and-fire" nature. For a better understanding, we explained both ANNs and SNNs in the below sections.

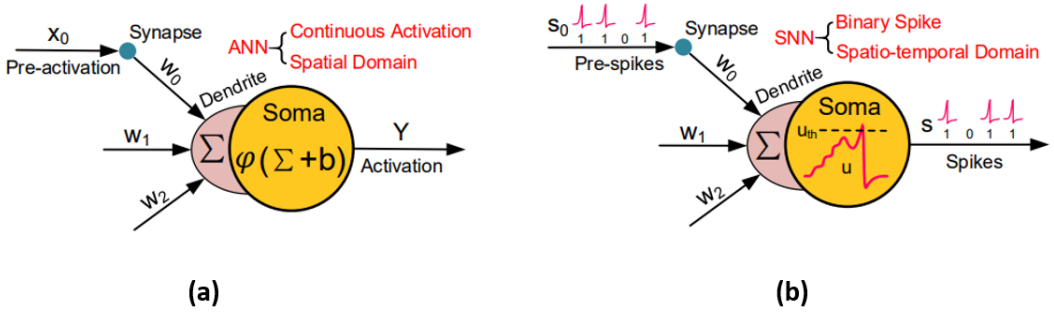

FIGURE 2.21: (a) Basic neuron model in ANNs. (b) Basic neuron model in SNNs. The figures are taken from reference [95].

2.4.2.1 Artificial Neural Networks (ANNs)

Artificial neurons act as node in artificial neural networks. An artificial neuron is a computational model inspired by the natural neurons. The complexity of real neurons is highly abstracted when modelling artificial neurons. Neurons accept weighted inputs through synapse and then computed by a mathematical function which determines the activation of the neuron. Another function computes the output of the artificial neuron depending on a certain threshold. ANNs combine artificial neurons in order to process information. Fig. 2.21(a) depicts a simple model of a typical artificial neuron. For an ANNs, the computational process is governed by the following equation.

$$y = \varphi(b + \sum_{j} x_j w_j) \tag{2.5}$$

Where,

x, and y are the input and output respectively. w and b are the synaptic weight and bias respectively and j is the index of the pre-synaptic input neurons. $\varphi(.)$ is a non-linear activation function, e.g., $\varphi(x) = ReLU(x) = max(x, 0)$. The communications between neurons in ANNs take place using the activations coded in high-precision and continuous values, and only propagate information in the spatial domain (i.e., layer by layer). It can be seen from the above equation, that the multiply-and-accumulate of inputs and weights is the major operation in ANNs. Even though ANNs are inspired by

the biological nervous system and are successfully used in various applications [96–98], their high abstraction compared to their biological counterpart [99] and their inability to capture the complex temporal dynamics of biological neurons have resulted in a new area of ANNs where the focus is placed on more biologically plausible neuronal models known as Spiking Neural Networks (SNNs).

2.4.2.2 Spiking Neural Networks (SNNs)

Spiking neural networks (SNNs) have the ability to capture the rich dynamics of biological neurons and to represent and integrate different information dimensions such as time, frequency, and phase. SNNs offer a promising computing paradigm and are potentially capable of modelling complex information processing in the brain [100–102]. SNNs are also potentially capable of dealing with large volumes of data and using trains of spikes for information representation [102]. Additionally, SNNs are suitable for implementation on low power hardware. A typical spiking neuronal model is shown in Fig. 2.21(b). Even though it has a similar structure, the behaviour is different as compared to the ANNs. In SNNs, the neurons communicate through electrical spikes coded in binary events rather than the continuous activations in ANNs. The input spikes from the pre-synaptic neurons comes to the soma (body of a neuron) through synapses and it is then integrated. Once the integrated value reaches a certain threshold, it produces an output spike. This behaviour is usually modeled by the popular Leaky Integrate and Fire (LIF) model [103]. The model consist of a capacitor (C) in parallel with a resistor (R) driven by a current $I(t)$ described by the following equation.

$$\tau \frac{\mathrm{d}V_c}{\mathrm{d}t} = -V_c(t) + RI(t) \tag{2.6}$$

Where, $\tau = RC$ is the membrane time constant. The working principle of LIF model is explained in detail in chapter 5. LIF model is the most widely used spiking neuron model. There also exist other neuron models in SNNs besides the LIF model, such as the model of Izhikevich [104] or Hodgkin & Huxley [105]. However, due to the higher complexity they are not widely used in practical SNNs.

2.4.2.3 Recent Advances in the Hardware Implementation of SNNs

The main goal of neuromorphic computing or artificial neural network is to mimic the functionalities of a human brain. Spiking Neural Network (SNN) is the new generation

of artificial neural networks which is more energy-efficient [92] than other neural networks as well as a von Neumann architecture based traditional computer. An artificial neuron and an artificial synapse are the two fundamental building blocks of an artificial SNN. Artificial neurons act as node of SNNs and the nodes are connected through artificial synapses. So, for the hardware implementation of SNNs, it is essential to realise both neuron and synapse using electronic devices and circuits. The functionalities of a neuron have been implemented in several CMOS based analog circuits [106–114], digital circuits [115, 116], and some non Si-based devices [117–119]. In all the implementations, there are two major challenges: (i) the implemented neuron must be energy and area efficient and (ii) it should be highly scalable so that it can match the high neuronal density of the human brain. Analog based implementation of a neuron is superior to digital implementation in terms of area and energy efficiency [108]. Joubert et. al. [108] compared LIF neuron using CMOS technology in both digital and analog implementation at $65\ nm$ node and claimed that the analog implementation consumes $5X$ less area and $20X$ less energy than digital implementation. Integrate and reset are the two circuit blocks of a LIF neuron. Efforts have been made to replace the area and energy inefficient CMOS circuit based integrate block with nano-scale devices [117, 119, 120]. Although, several CMOS based LIF neurons have been reported, it is believed that using nanoscale devices, neuromorphic systems can be improved in terms of both area and power consumption by almost a factor of 10, as compared to conventional CMOS neurons [153]. However, most of the novel nanoscale artificial neurons use non-Si materials [117–119, 157], which can be a disadvantage from a fabrication point of view. Recently S. Dutta et al. [120], demonstrated an analog based implementation of a low energy integrate block of a LIF neuron for SNN using a partially depleted silicon on insulator (PD-SOI) MOSFET with a spiking frequency in the order of MHz and energy consumption of 1.3×10^{-11} J/spike which is three orders of magnitude less energy efficient than a biological neuron ($\sim 10^{-14}$ J/spike [121]). So, the opportunities are there to design an electronic device which has an energy efficiency equivalent to that of a biological neuron.

In this thesis, we mainly focus on the realization of an electronic neuron for spiking neural networks. In Chapter 5, with the application of floating body effects, we demonstrate through well-calibrated TCAD [122] simulations, a highly scalable bulk FinFET based analog implementation of the integrate block of a LIF neuron. The charge integration mechanism is shown at the body of the proposed device through the addition of a buried n^+ layer. We also demonstrate that the proposed neuron is seen to have a spiking

frequency in the MHz range and the lowest energy consumption for the integrate block reported till date.

Chapter 3

TCAD Implementation of PD-SOI Based TiO_2 S/D Zero Capacitor Random Access Memory (Z-RAM)

3.1 Introduction

A conventional dynamic random access memory (DRAM) [12] consists of a transistor and a storage capacitor as shown in Fig. 2.4. The storage capacitor should have a minimum capacitance value of $30\ fF$ [18] to obtain a sufficient signal to noise ratio. The minimum capacitance requirements are met using trench capacitor [20, 21] or stack capacitor [24, 25] technologies for sub $100\ nm$ technology nodes. The limitations of trench and stack capacitor DRAM cells are discussed in Chapter-2. Another big challenge for a conventional 1T-1C DRAM is the migration from $6F^2$ to $4F^2$ (where F is the bit line half-pitch) cell size. The majority of the foundries manufacture DRAM cells with $6F^2$ cell size. The most recent 1T-1C DRAM (LPDDR4X is being manufactured using '$1y$' nm technology [123] . '$1y$' nm is defined as $14 - 16$ nm [124]. Gate length is approximately 20 nm [125] for the '$1y$' nm technology node chip manufactured by SAMSUNG and this has a cell size of $6F^2$. The most promising way to increase the number of bits per chip for the next generation technology node is by scaling the cell size factor 'a' (where a = [DRAM cell size]/[DRAM half pitch]^2). Migration from $6F^2$ (a = 6) to $4F^2$ (a = 4) cell size is very challenging. Z-RAM cell [49, 50, 59] can be a good alternative. The size of a Z-RAM cell is $4F^2$ due to its capacitor-less structure. This increases the number of bits per chip for future technology nodes. This is one of

the major advantages of Z-RAM cells over conventional 1T-1C DRAM cells. In the first generation of Z-RAM cell [49, 50], the floating-body of n-channel silicon on insulator (SOI) MOSFET is used as the storage node. The presence of excess majority carriers (holes in this case) at the body can be assigned the logic state '1' and the absence of excess holes can be assigned the logic state '0'. These excess majority carriers at the body are created by impact ionization at the drain-body junction and they are stored at the lower part of the body near the body-buried oxide (BOX) interface. Due to the presence of excess majority carriers at the floating body, the threshold voltage is changed which confirms the two states (logic state '1' (I_1) and logic state '0' (I_0)) of the first generation of Z-RAM cell. The difference between the logic state '1' (I_1) and logic state '0' (I_0) read currents is called the sense margin ($SM = \Delta I = I_1 - I_0$). As per our device design, the retention time is defined as the time when the sense margin reaches it's 50% value. The retention time specification is one of the most important figure of merit for a DRAM cell and it should be a minimum of $64\ ms$ at $T = 358\ K$ as per ITRS [17]. But, due to over the barrier leakage, an all-Si Z-RAM cell faces low retention time [64, 65] problem.

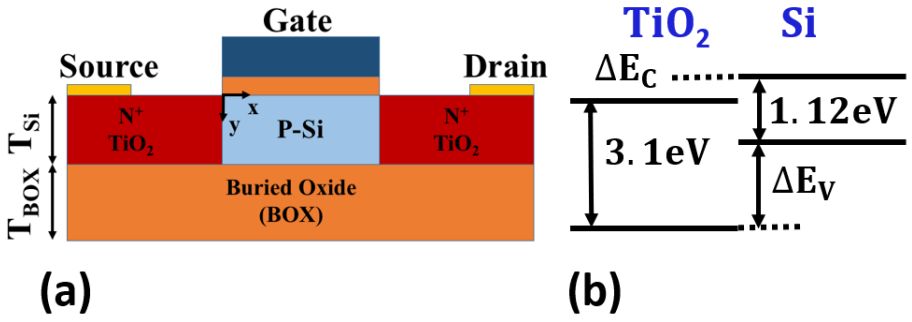

FIGURE 3.1: (a) Schematic of the proposed TiO_2 source/drain Z-RAM cell. (b) TiO_2 and Si band line up. $\Delta E_C \approx 0.05\ eV$ and $\Delta E_V \approx 2\ eV$ [126].

In this chapter, we propose a TiO_2 based Z-RAM cell (Fig. 3.1) in which Si is replaced by high bandgap TiO_2 ($E_g = 3.1\ eV$ [127]) in the source and drain in an n-channel PD-SOI MOSFET. TiO_2 is a n-type semiconductor due to different types of defects like oxygen vacancies and titanium interstitials [128]. Depending on the deposition technique, the carrier concentration in TiO_2 varies over $10^{17} - 10^{20} cm^{-3}$ [129, 130]. In the recent past, excellent surface passivation of Si surface using TiO_2 with surface recombination velocities of $2.8\ cm/s$ and $8.3\ cm/s$ for n-type and p-type wafers have been demonstrated [131]. Silicon solar cells using TiO_2 as electron selective contact have been demonstrated with a power conversion efficiency of 22.1% [132].

This implies that high effective carrier lifetime in excess of 100 μs can be obtained in Si bounded by TiO_2.

Large valance band offset between TiO_2 and Si ($\Delta E_V \approx 2\ eV$ [126]) is utilized for storing larger number of excess holes at the body for a longer time than is possible with an all-Si Z-RAM cell. An improvement in retention time as well as sense margin at both $T = 300\ K$ and $T = 358\ K$ are reported for the proposed Z-RAM cell through well calibrated TCAD simulations. The extracted retention time for the proposed TiO_2 source/drain Z-RAM cell is 3.5 s and 160 ms at $T = 300\ K$ and 358 K respectively and for all-Si Z-RAM cell, it is 1.5 ms and 150 μs at $T = 300\ K$ and 358 K respectively.

3.2 Device Design

Fig. 3.1(a) shows the schematic of the proposed TiO_2 S/D Z-RAM cell. TCAD simulations were carried out to study the operation and retention characteristics of the device. For calibration of the simulation models, an all-Si n-channel PD-SOI MOSFET, described in reference [133] was simulated. The device structural details are given in Table 3.1.

TABLE 3.1: PD-SOI MOSFET parameters used in TCAD simulations, used for calibration of simulation models.

Parameter (Unit)	**Value**
Buried oxide thickness (nm)	400
Device layer thickness (nm)	150
Gate oxide thickness (nm)	4.5
S/D doping (cm^{-3})	10^{20}
Body doping (cm^{-3})	2×10^{17}
Gate length (nm)	150
Gate metal work function (eV)	4.4

The match between the simulated I_{DS}-V_{GS} characteristics and experimental data are shown in Fig. 3.2.

To calibrate the simulation models for the TiO_2-Si diode, simulation was carried out on n-TiO_2 - p-Si heterostructure diode and the diode I-V characteristics (Fig. 3.3) was matched with the experimental data from reference [126].

The parameters used [127, 129, 134, 135] for TCAD simulations of n-TiO_2 - p-Si heterostructure diode are listed in Table 3.2. Same device dimensions and parameters

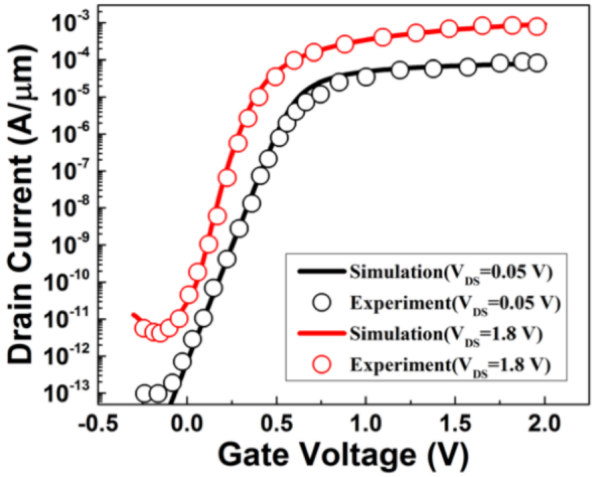

FIGURE 3.2: Comparison of simulated drain current (I_{DS}) vs. gate voltage (V_{GS}) characteristics of an all Si n-channel PD-SOI MOSFET with published result [133], demonstrating the calibration of the simulation models used.

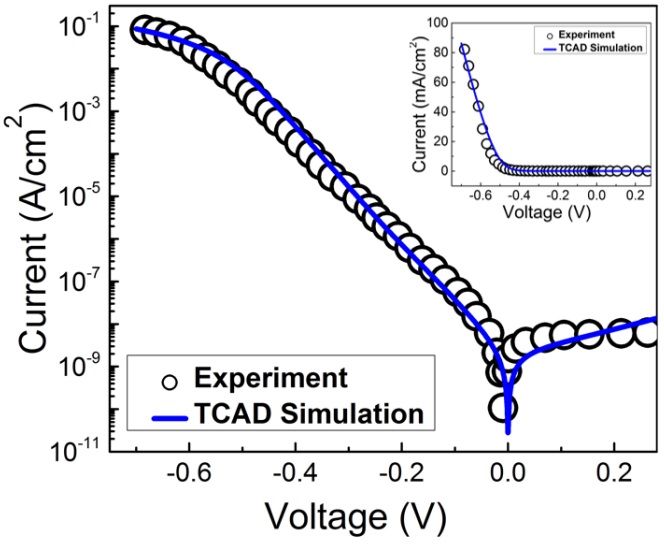

FIGURE 3.3: Comparison of simulated diode characteristics for n-TiO_2 - p-Si heterostructure diode with published result [126], demonstrating the calibration of the simulation models used.

listed in Table 3.1 and Table 3.2 were used while simulating the TiO_2 source/drain Z-RAM cell. In our simulations, we have assumed the carrier lifetime in Si to be $1\mu s$, and the programming time was taken as $1ns$.

3.2.1 Model Calibration

In the study of TiO_2 source/drain Z-RAM, TCAD Sentaurus simulator [122] is used. To capture the real device phenomena, proper device models are chosen while simulating

TABLE 3.2: TiO_2 parameters used in TCAD simulation shown in Fig. 3.3.

Parameters of TiO_2	**Value**	**References**
Band-gap E_g (eV)	3.1	[127]
Electron affinity (eV)	4.1	[135]
Dielectric constant	91	[134]
N_C (cm^{-3})	3.1×10^{21}	[134]
N_V (cm^{-3})	3.3×10^{21}	[134]
Doping concentration (cm^{-3})	10^{19} (n-type)	[129]
Electron mobility $(cm^2/V.s)$	0.1	[127]
Electron effective mass	$10m_0$	[135]

the Z-RAM cell. Different models used are described in brief in the below section.

Electrostatic potential: Immobile ionized impurities and mobile electrons/holes play key roles in any semiconductor devices. Depending on the charge distributions inside a device, the electrostatic potential varies. The dependence of electrostatic potential on charge distribution inside a device can be given by the Poisson's equation, which is:

$$\nabla(\epsilon\nabla\phi + P) = -q(p - n + N_D - N_A) - \rho_{trap} \tag{3.1}$$

Where:
ϵ is the electrical permittivity of the material in $F.m^{-1}$,
ϕ is the electrostatic potential in V,
n and p are the electron and hole densities in cm^{-3},
N_D and N_A are the concentration of ionized donors and acceptors cm^{-3},
q is the elementary electronic charge in C,
P is the ferroelectric polarization in $C.m^{-2}$ and ρ_{trap} is the interface trap charge density in $cm^{-2}.eV^{-1}$.

Hydrodynamic transport model: When device dimensions shrink below 100 nm, the traditional drift-diffusion transport model can not properly capture the real device phenomena like impact ionization, velocity overshoot, etc. It overestimates the impact ionization rate and can not predict the velocity overshoot phenomena properly. Since for lower dimensions, the hydrodynamic model with carrier temperature and the lattice temperature as the driving forces accurately predict the device phenomena, this model is used in the device simulation.

Semiconductor band structure: The most fundamental property of a semiconductor material is its band structure. Realistic band structures can only be fully captured in full-band Monte Carlo simulations. The band structure is simplified to several quantities: the conduction and valence band edge energies, electrons, and holes effective mass. The intrinsic carrier concentration in silicon is determined by the band-gap narrowing model. In this simulation, the "OldSlotboom" model is selected, which is based on measurements of in n-p-n transistors. Since there is a parasitic BJT inside the transistor, it plays an important role in device characteristics.

Doping and temperature dependent SRH recombination model: Recombination through deep defect levels in the energy bandgap is usually labelled as SRH recombination. SRH lifetimes depend on doping and temperature. Doping dependent SRH lifetime is modelled using Scharfetter relation.

$$\tau_{dop}(N_A) = \tau_{min} + \frac{(\tau_{max} - \tau_{min})}{1 + (N_A/N_{ref})^{\gamma}} \tag{3.2}$$

Where, $N_{ref} = 10^{16}\ cm^{-3}$ and $\gamma = 1$. In our device design, we have assumed, N_A = $10^{17}\ cm^{-3}$, $\tau_{max} = 1\ \mu s$, and $\tau_{min} = 0$.

Doping and temperature dependent mobility model: In the device simulation, we have used doping and temperature dependent mobility models so that the most accurate electron and hole mobilities can be predicted. Both phonon scattering and impurity scattering determine the mobility . For undoped materials, constant mobility model (lattice temperature mobility model) is used. For doped materials, impurities are the main cause of carrier scattering. This leads to a degradation of mobility.

Impact ionization model: Since impact ionization plays an important role in Z-RAM programming, impact ionization phenomena is included through the avalanche generation model. Since hydrodynamic simulation is performed, the driving force here is temperature instead of the electric field.

Non-local band to band tunnelling model: Since band to band tunnelling plays an important role in Z-RAM operation, a non-local band to band tunnelling model is

included in the device simulation. In the non-local BTBT model, electron effective mass for Si and TiO_2 were taken as $0.65m_0$ [136] and $10m_0$ [135] respectively.

Thermionic emission model: As there are both valence and conduction band offsets between Si and TiO_2, the thermionic model is also included to account for the transport of electrons across the conduction bands of the two sides of the junction.

Interface Trap Density: For proper calibration of the Si-TiO_2 diode current-voltage (I-V) characteristics, a Gaussian distribution of interface trap density with a maximum value of $10^{10}eV^{-1}cm^{-2}$, energetically located at $50\ meV$ below the conduction band edge and with a standard deviation of $50\ meV$ is included at the interface between TiO_2 and Si (Fig. 3.4).

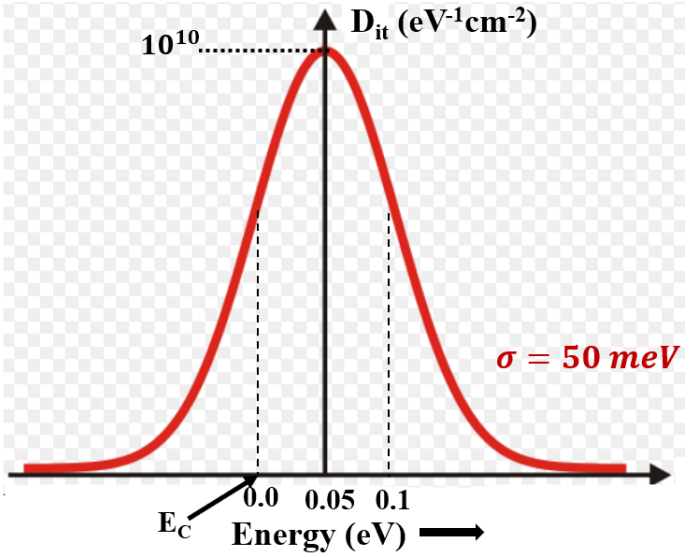

FIGURE 3.4: Gaussian distribution of interface trap density (D_{it}) with a maximum value of $10^{10}\ eV^{-1}cm^{-2}$, energetically located at $50\ meV$ (mean position) below the conduction band and with a standard deviation of $50\ meV$.

The Table 3.3 shows the calibrated values against the default values of different parameters used in the TCAD simulations.

3.3 Simulation Results

Fig. 3.5 shows the electrostatic potential of TiO_2 S/D Z-RAM cell ($y = 100\ nm$ i.e., $50\ nm$ above the body-BOX interface. See Fig. 3.1 for the definition of the coordinate

TABLE 3.3: Default and Calibrated values of Parameters Used in the TCAD Simulations of PD-SOI n-channel TiO_2 source/drain MOSFET.

Parameters	**Default Value**		**Calibrated Value**	
	Si	**TiO_2**	**Si**	**TiO_2**
SRH Recom. τ_{max} & τ_{min}	10 μs, 0 for e 3 μs, 0 for h	10 μs, 0 for e 3 μs, 0 for h	1 μs, 0 for e 1 μs, 0 for h	1 μs, 0 for e 1 μs, 0 for h
Auger Recom. (cm^6/s) A & B	$6.7e^{-32}$, $7.2e^{-32}$ e $2.4e^{-31}$, $4.5e^{-33}$ h	$6.7e^{-32}$, $7.2e^{-32}$ e $2.4e^{-31}$, $4.5e^{-33}$ h	$6.7e^{-32}$, $7.2e^{-32}$ $2.4e^{-31}$, $4.5e^{-33}$	$6.7e^{-32}$, $7.2e^{-32}$ $2.4e^{-31}$, $4.5e^{-33}$
Mobility $(cm^2/V.s)$ μ_{emax} & μ_{hmax}	1417, 470	1417, 470	1417, 470	0.1, 0.1
BTBT A $(cm^{-1}s^{-1}V^{-2})$ & B $(V.cm^{-1}eV^{-1.5})$	$8.9e^{20}$, $2.1e^{7}$	$8.9e^{20}$, $2.1e^{7}$	$8.9e^{19}$, $2.1e^{7}$	$8.9e^{19}$, $2.1e^{7}$
Tunnelling mass m_e^* & m_h^*	$0.5m_0$, $0.5m_0$	$0.5m_0$, $0.5m_0$	$0.65m_0$, $0.5m_0$	$10m_0$, $10m_0$
Impact Ionization n_1_f & p_1_f	1, 1	1, 1	0.95, 0.95	0.95, 0.95
Surface Recom. S_0 for e & h	100, 100		50, 50	

system.) under read '1' biasing condition and it has been compared with the all-Si Z-RAM cell. Due to the large valance band offset between TiO_2 and Si as shown in Fig. 3.1(b), the hole barrier at the body of TiO_2 S/D cell will be higher by $2\ eV$ as compared to that of all-Si cell as can be seen from Fig. 3.5.

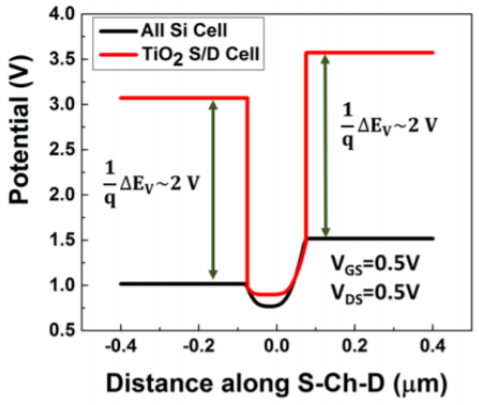

FIGURE 3.5: Comparison of the simulated potential between TiO_2 S/D cell and all-Si cell under read '1' condition.

The barrier height for holes at the valence band for both the transistors can be given by equation 3.3 and 3.4.

$$E_{B\_Si-Si} = k_B T \ln\left(\frac{N_{A-body}}{n_{i-Si}}\right) + k_B T \ln\left(\frac{N_{D-source}}{n_{i-Si}}\right) \tag{3.3}$$

$$E_{B\_TiO_2-Si} = \Delta E_V + k_B T \ln\left(\frac{N_{A-body}}{n_{i-Si}}\right) + k_B T \ln\left(\frac{N_{D-source}}{n_{i-TiO_2}}\right) - \Delta E_i \tag{3.4}$$

Where:
$\Delta E_i = \frac{E_{g-TiO_2} - E_{g-Si}}{2} - \Delta E_C$,
N_{A-body} is the body doping concentration in cm^{-3},
n_{i-Si} is the intrinsic carrier concentration of silicon in cm^{-3},
$N_{D-source}$ is the source/drain doping concentration in cm^{-3},
n_{i-TiO_2} is the intrinsic carrier concentration in TiO_2 in cm^{-3},
ΔE_C and ΔE_V are conduction and valence band offsets respectively in eV,
E_{g-TiO_2} and E_{g-Si} correspond to the energy band gap of TiO_2 and Si respectively in eV.

Due to large barrier height for excess generated holes at the body, floating body effects are more prominent in TiO_2 source/drain cell as compared to the all-Si cell.

The biasing scheme for both the TiO_2 source/drain and all-Si cells are the same and it is shown in Fig. 3.6. The impact ionization mechanism was used for writing state

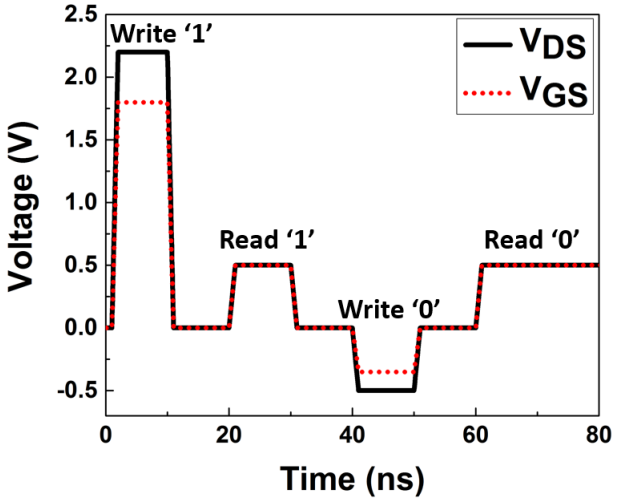

FIGURE 3.6: Biasing scheme for both TiO_2 source/drain and all-Si Z-RAM cell. Time in the x-axis is shown for reference.

'1'. Forward biasing of the drain-body junction is used for erasing or to write state '0'. For reading '1' or '0', a comparatively low drain bias is applied at the drain terminal to avoid impact ionization.

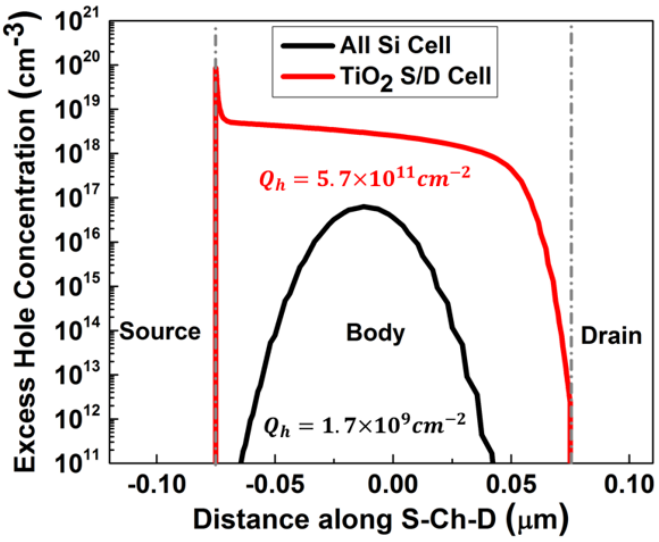

FIGURE 3.7: Comparison of the excess hole concentration between TiO_2 S/D cell and all-Si cell at time $t = 25\ ns$ i.e., $4\ ns$ after the read '1' operation starts (time reference is shown in Fig. 3.6). Q_h is the excess hole density in the body, obtained by integrating the hole concentration in the body.

Excess hole concentration (cm^{-3}) at $50\ nm$ above the body - BOX interface is plotted during the read '1' operation i.e., at $t = 25\ ns$ as shown in Fig. 3.7 for both the cells. The extracted excess hole density (integral of the hole concentration in the body in cm^{-2}) of TiO_2 S/D cell ($5.7 \times 10^{11} cm^{-2}$) is more than two orders of magnitude higher than that of the all-Si cell ($1.7 \times 10^9 cm^{-2}$). Since the stored holes at the body cannot

escape easily for TiO_2 source/drain cell due to large valence band offset between TiO_2 and Si, floating body effects will be more prominent here and more number of excess holes will accumulate at the body and stay there for a comparatively longer time than all-Si cell.

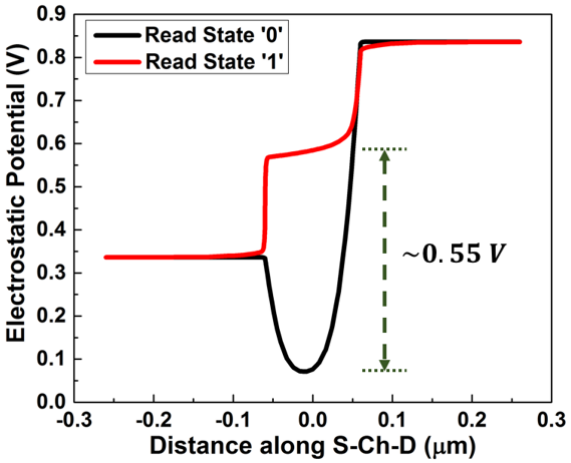

FIGURE 3.8: Electrostatic potential at the body of the TiO_2 S/D cell during read '1' and read '0'.

Due to the higher hole density of TiO_2 source/drain cell, the difference in threshold voltage between state '1' and state '0' will be higher. Hence, the sense margin of the TiO_2 S/D cell is anticipated to be higher. Fig. 3.8 shows the electrostatic potential of the TiO_2 S/D cell during read '1' and read '0'. Between the two states, there is a $0.55V$ of potential difference at the body which leads to two different threshold voltages for state '1' and state '0'.

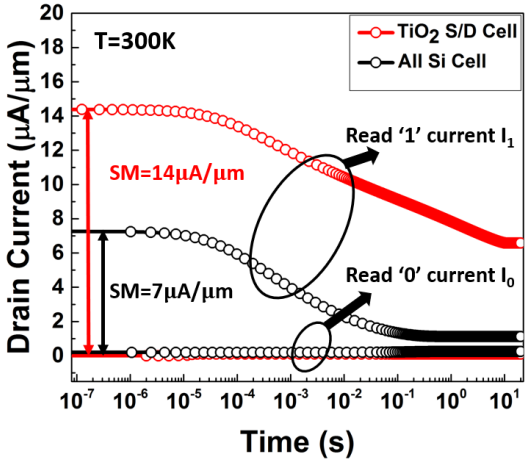

FIGURE 3.9: Comparison of the transient characteristics of TiO_2 S/D cell and all-Si Z-RAM cell at $T = 300K$.

Fig. 3.9 shows the comparison of retention characteristics for both the memory cells at $T = 300K$. As shown in the figure, there is an improvement in sense margin by $7\mu A/\mu m$ at $T = 300K$ for TiO_2 S/D cell.

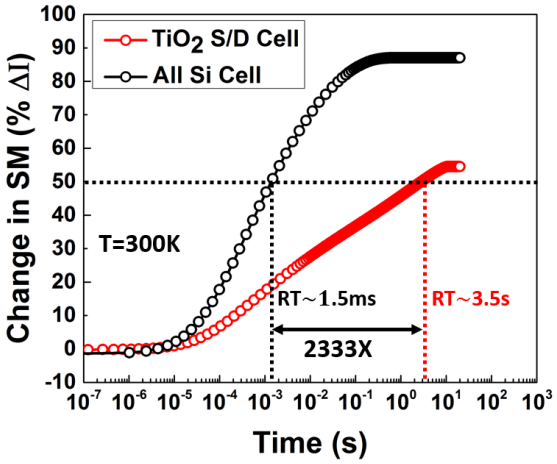

FIGURE 3.10: Comparison of the change in sense margin as a function of time between TiO_2 source/drain cell and all-Si cell at $T = 300K$.

Fig. 3.10 shows the change in sense margin with time at $T = 300K$ for both the cells. The extracted retention time for TiO_2 S/D cell and all-Si cell are $3.5\ s$ and $1.5\ ms$ respectively.

We have also calculated the retention times and sense margins for both the Z-RAM cells at $358K$ as shown in 3.11 and 3.12. SRH recombination rate increases with temperature leading to a faster decay of read '1' current (I_1). As a consequence, the retention time for both the memory cells decrease. But even at $358\ K$, the retention time is $160\ ms$ for the TiO_2 S/D cell which is $2.5X$ that of the ITRS specification.

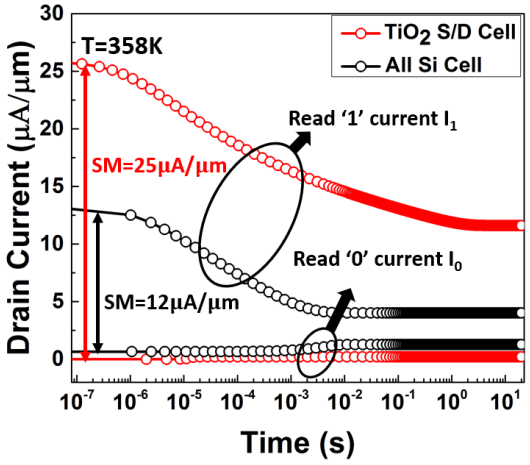

FIGURE 3.11: Comparison of the transient characteristics of TiO_2 S/D cell and all-Si Z-RAM cell at $T = 358K$.

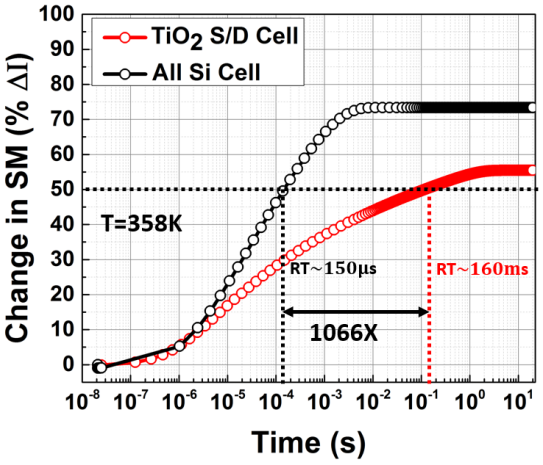

FIGURE 3.12: Comparison of the change in sense margin as a function of time between TiO_2 source/drain cell and all-Si cell at $T = 358K$.

The proposed cell has a retention time of $1066X$ that of the all Si cell at $358K$. For TiO_2 cell, as the barrier for excess holes at the body is large, the excess holes can not escape easily even at higher temperature, leading to the improved performance of this cell compared to the all-Si cell.

3.4 Conclusion

Due to the large valance band offset between TiO_2 and Si, TiO_2 S/D Z-RAM cell showed higher sense margin and longer retention time. The retention time predicted using TCAD simulations are seen to be better than the specifications for all-Si Z-RAM cell. In the context of the recent advancements in deposition of TiO_2 and excellent passivation of Si surface using TiO_2, the proposed device architecture is very much realizable for dynamic memory applications.

Chapter 4

TCAD Implementation of TiO_2 S/D FD-SOI Based Zero Capacitor Random Access Memory (Z-RAM)

4.1 Introduction

In the previous chapter, we have demonstrated a TiO_2 source/drain PD-SOI MOSFET based first generation of Z-RAM cell where impact ionization based programming method is used. There are certain drawbacks in PD-SOI MOSFET based first generation of Z-RAM cell which is programmed by impact ionization. A high drain bias is applied during programming to cause impact ionization at the drain body junction. This high drain bias can be a concern. Due to several cycles of erase/program, it can degrade the gate oxide quality through hot carrier injection, which leads to the generation of interface states at the Si-SiO_2 interface, and eventually to premature oxide breakdown. More the density of the interface states at the Si-SiO_2 interface more will be the degradation of the retention time due to higher recombination. Furthermore, programming time is an important factor in deciding the speed of a DRAM cell. To achieve a faster programming, the impact ionization rate should be increased. To increase the impact ionization rate, a higher drain voltage needs to be applied at the drain terminal which leads to higher power consumption. Furthermore, in PD-SOI MOSFET based Z-RAM cells, the threshold voltage difference between state '1' and state '0' (ΔV_{TH}) can be approximated by $\Delta V_B \times (C_D/C_{OX})$ [137], where the body potential variation (ΔV_B) is proportional to the change in hole density at the body between state '1' and state '0'.

From the ΔV_{TH} and C_D relation, we can say that to increase the sense margin, we need to increase the C_D. C_D can be increased by reducing the body thickness (t_{Si}). We can not reduce the body thickness of a PD-SOI MOSFET otherwise it will lose it's "partially depleted" property. So scaling is also a big challenge for PD-SOI based Z-RAM cells.

These problems are solved up to a certain extent in the second generation of Z-RAM cells [54, 55] where the parasitic BJT inside an n-channel FD-SOI MOSFET is used to write state '1'. In this case, source (N^+) acts as an emitter, body (P) as a base and drain (N^+) as a collector of the parasitic BJT inside the MOSFET. This parasitic BJT is turned on by increasing body potential through the accumulation of holes which is initiated by impact ionization at the drain-channel junction. Once the BJT is turned on, more electrons will enter the channel and impact ionization rate increases and this is a regenerative process that eventually leads to latch-up of current. The increase in the current can be used for reading logic state '1'. As per our device design, the retention time is defined as the time when the read state '1' current (I_1) reaches 2 $\mu A/\mu m$. The retention time specification is one of the most important figure of merit for a DRAM cell and it should be a minimum of 64 ms at $T = 358\ K$ as per ITRS [17]. All-Si Z-RAM cells [64, 65] does not meet the ITRS specification of retention time. The reasons being the leakage of excess holes from the floating body to source/drain due to low barrier height for holes and Shockley-Read-Hall (SRH) recombination in the body. To overcome this problem, few hetero-structure based Z-RAM cells [66–69] have been reported. Write '0' time is an important parameter for hetero-structure based Z-RAM cell as with the increase in valence band offset, write '0' time increases [67]. Write '0' process and the time required for the same are not elaborated in the literature.

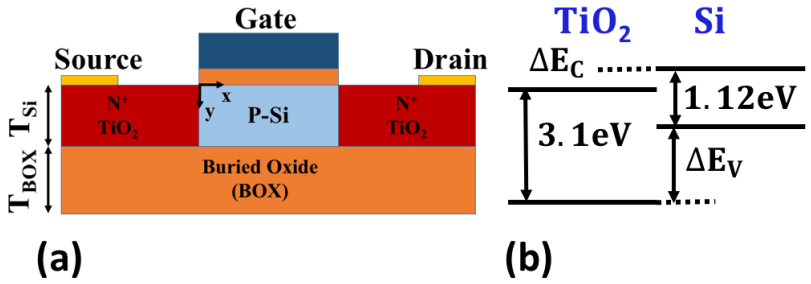

FIGURE 4.1: (a) Schematic of the proposed TiO_2 source/drain FD-SOI Z-RAM cell. (b) TiO_2 and Si band line up. $\Delta E_C \approx 0.05\ eV$ and $\Delta E_V \approx 2\ eV$ [126].

In this chapter, we propose a TiO_2 based Z-RAM cell (Fig. 4.1) in which Si is replaced by high bandgap TiO_2 ($E_g = 3.1\ eV$) in the source and drain in an n-channel FD-SOI MOSFET.

Parasitic BJT based programming method has been used for the proposed memory cell. We have demonstrated low voltage programming and have shown that, during programming, hole storage is initiated by band to band tunnelling and subsequently it is taken over by impact ionization. We have extracted the retention time for $30\ nm$ channel length device for both $T = 300\ K$ and $T = 358\ K$ through well-calibrated TCAD simulations using Sentaurus Device simulator [122]. We report retention times of $2\ s$ and $70\ ms$ at $T = 300\ K$ and $T = 358\ K$ respectively for the proposed design. We have also optimized the write '0' time to $6\ \mu s$ for the proposed Z-RAM cell.

4.2 Device Design & TCAD Model Calibration

The device structural details are given in Table 4.1. Unlike logic transistors, the device layer thickness and gate oxide thickness are taken comparatively higher to increase the volume of charge storage and to make sure that gate oxide leakage is as low as possible.

TABLE 4.1: Transistor parameters used in TCAD simulations.

Parameter (Unit)	Value
Buried oxide thickness (nm)	50
Device layer thickness (nm)	50
Gate oxide thickness (nm)	4.5
S/D doping (cm^{-3})	10^{19}
Body doping (cm^{-3})	10^{17}
Gate length (nm)	30
Gate metal work function (eV)	4.4

In general, logic transistors are designed to achieve good electrostatics. Good electrostatics is achieved by having a thin gate oxide ($1\ nm$) and thin body thickness (generally $< 30\ nm$ for thin body SOI and $< 10\ nm$ for ultra-thin body SOI). However, this design strategy is not suitable from a Z-RAM point of view. Logic transistors can allow gate leakage to some extent. But gate leakage in Z-RAM cell has to be minimal to prevent the leakage of the charge stored in the transistor body (or in the capacitor in conventional 1T-1C DRAM). This criterion leads to the design constraint of relatively thicker gate oxide ($4 - 5\ nm$). This constraint of thicker T_{ox} in the Z-RAM transistor is similar to the conventional DRAM technology as per ITRS guidelines.

Again, in the Z-RAM cell, the excess charge is stored inside the transistor body. Thus, scaling of the body thickness (T_{body}) also seriously impacts its charge storage

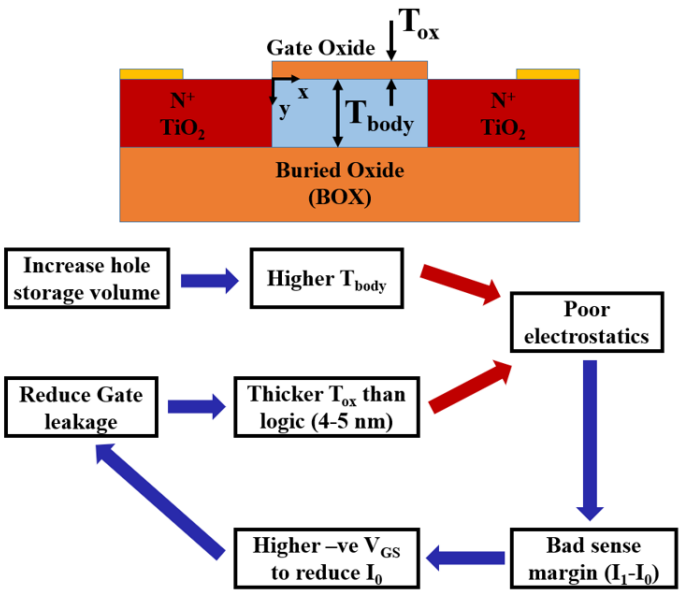

FIGURE 4.2: Design constraints of the transistor parameter for zero capacitor random access memory (Z-RAM) cell.

capability. Hence, the Z-RAM transistor should be optimized to have the maximum body thickness possible. These design guidelines make Z-RAM transistors susceptible to poor electrostatics. Fig. 4.2 above shows the optimization strategy of the Z-RAM transistor parameters. Further, as the parasitic BJT based programming is used for the proposed Z-RAM cell, the transistor operates in the accumulation region. Once the excess hole generation is initiated by BTBT, the barrier for electrons from source to body reduces due to these excess holes at the body and this leads to an increase in impact ionization rate. As this is a positive feedback process, the drain current latches up after some time. If we denote the impact ionization factor M as forward gain and source electron injection efficiency β as feedback factor then the latch-up condition is satisfied when $M\beta \approx 1$. At the start of the programming, β is quite small, thus $M\beta << 1$. The time required for the transistor to go into latch-up mode depends on the transistor parameters and the operating voltages. As an example, if the gate oxide thickness is too small then the gate will have good control over the transistor body, which will prevent the source to inject more electrons in the channel. This will increase the latch-up time of the Z-RAM transistor. Similarly having a thicker body also helps in reducing the latch-up time as a greater number of excess holes can be stored. With a thin body and thin gate oxide, the latch-up time can be reduced by increasing the drain bias but that will increase the power dissipation. So, the transistor parameter optimization strategy is different for the Z-RAM cell which is unlike the logic transistor.

We used $10^{17}\ cm^{-3}$ body doping for the FD-SOI MOSFET. As the excess holes

generation during programming is initiated by band to band tunnelling, band to band tunnelling is an important factor for the proposed memory cell. Band to band tunnelling has to be high at the initial phase of the programming. We know that the tunnelling probability of an $n^+ - p^+$ junction $>$ tunnelling probability of an $n^+ - p$ junction $>$ tunnelling probability of an $n^+ - p^-$ junction where, p^+, p and p^- represent highly, moderately and low doped p-type Si [138]. If the body is made very low doped the programming will not start at such low bias. That is why the body of the proposed memory cell is moderately doped.

Hydrodynamic simulations are carried out to study the promise of the proposed TiO_2 source/drain FD-SOI MOSFET based Z-RAM cell. The simulation models and calibrated parameters (Table 3.3) are already shown in chapter 3. The TiO_2 parameters used in the TCAD simulations are listed in Table 3.2.

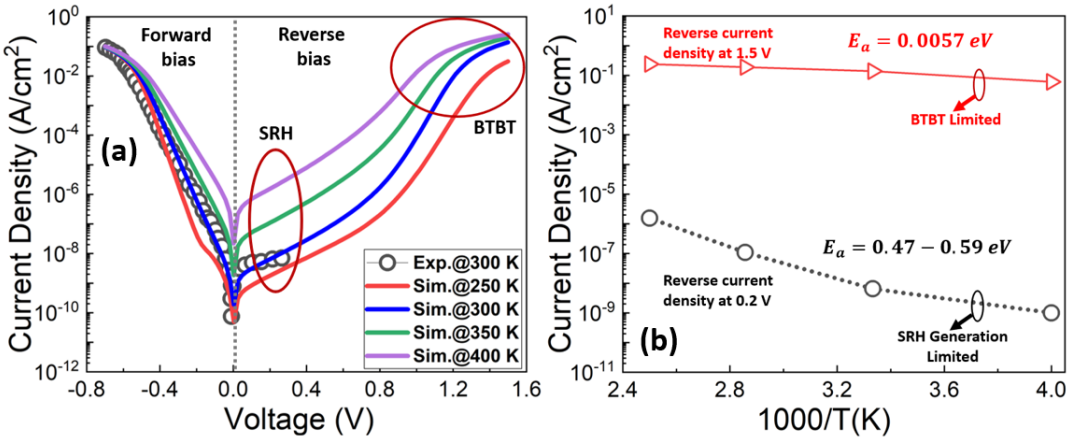

FIGURE 4.3: (a) Comparison of simulated diode characteristic for n-TiO_2 - p-Si heterostructure diode with published result [126] at $T = 300\ K$, demonstrating the calibration of the simulation models used. I-V characteristics are shown for $T = 250\ K, 300\ K, 350\ K$ and $400\ K$ respectively and the simulation is extended upto $1.5\ V$ in reverse direction to check the current conduction mechanism at high reverse bias. (b) Current density is plotted as a function of inverse temperature for two different reverse biases ($0.2\ V$ and $1.5\ V$).

For the validation of our simulation, TiO_2 - Si heterostructure diode I-V characteristic is matched with the experimental data from [126] at $T = 300\ K$ as can be seen in Fig. 3.3 from chapter 3. To understand the temperature effects on the high reverse bias, the simulation is extended up to $1.5\ V$ in the reverse direction and the diode I-V characteristics are simulated for four different temperatures starting from $T = 250\ K$ to $400\ K$ with a gap of $50\ K$ as shown in Fig. 4.3(a). The Arrhenius plot (Fig. 4.3(b))is also shown where the current density is plotted as a function of inverse temperature for two different reverse biases $0.2\ V$ and $1.5\ V$. It can be seen that the current is varying

as a function of temperature for a reverse bias of $0.2\ V$. The extracted activation energy ranges from 0.47 - $0.59\ eV$ which is almost half the mid-gap energy of Si. This indicates that current at low reverse bias is dominated by SRH generation-recombination. On the other hand, for a reverse bias of $1.5\ V$, the current density is almost independent of temperature with activation energy of $0.0057\ eV$. This confirms that the current at high reverse bias is dominated by tunnelling because tunnelling has a weak dependence on temperature. So, as reverse bias increases, the current conduction mechanism switches from SRH generation-recombination to tunnelling.

4.3 Device Operating Principle

Fig. 4.1(a) shows the schematic of the proposed TiO_2 source/drain Z-RAM cell. In parasitic BJT based Z-RAM, the state '1'/state '0' are recognized by turning on/off of the parasitic BJT in the transistor. The device physics of the programming operation is investigated by carrying out transient simulations. Simulations are carried out with and without BTBT to elucidate the role of this mechanism in programming. The gate voltage (V_{GS}) is fixed at $-0.8\ V$ and simulations were carried out at drain voltages (V_{DS}) of $0.5\ V$ to $0.8\ V$. The results of the transient simulations with BTBT are shown in Fig. 4.4(a). It is seen that the drain current latches up for $V_{DS} \geq 0.6$ V, with the time to latch-up decreasing with the increase in V_{DS}.

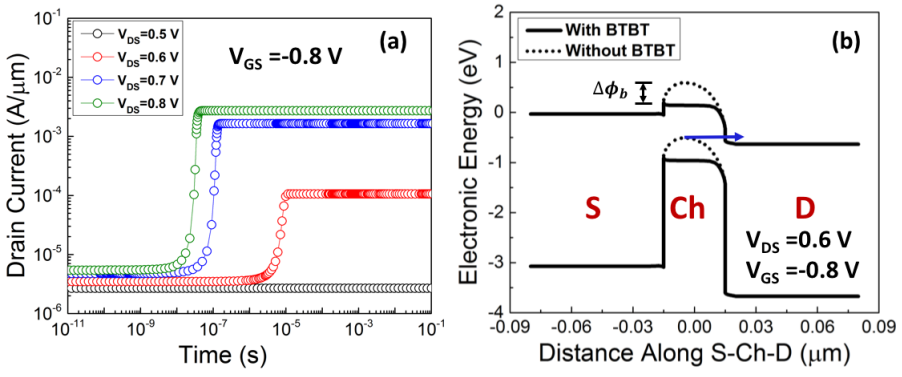

FIGURE 4.4: (a) Latch-up characteristics of the $30\ nm$ channel length device under different values of V_{DS} and $V_{GS} = -0.8V$. It is evident from this figure that the device latches up for $V_{DS} \geq 0.6\ V$ and $V_{GS} = -0.8\ V$. (b) Energy band diagram along the source-channel-drain ($1\ nm$ below the gate oxide - body interface) with and without band to band tunnelling of the TiO_2 source/drain cell taken at $t = 1\ s$.

Fig. 4.4(b) shows the band diagram over a cut line along source - channel - drain, 1 nm below the gate dielectric - silicon interface, when the simulations were carried out

with and without BTBT. It is seen that when BTBT is not turned on in the simulations, the valence band of the channel region aligns with the conduction band of the drain, suggesting the possibility of tunneling across the channel to drain junction.

Fig. 4.5(a) shows the impact ionization rate with and without BTBT. The impact ionization rate is seen to be much higher when BTBT is turned on. Fig. 4.5(b) shows the temporal evolution of the BTBT rate for $V_{DS} = 0.8\ V$ and $V_{GS} = -0.8\ V$. Initially the BTBT rate is high. However with time, the BTBT rate decreases.

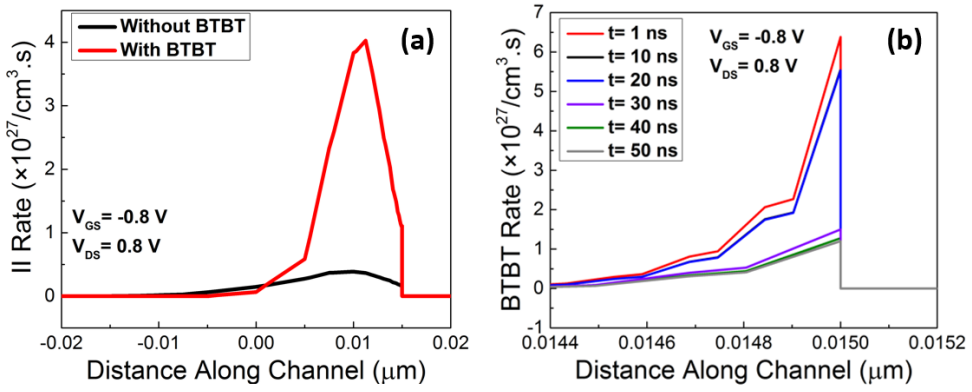

FIGURE 4.5: (a) Impact ionization rate at the drain junction with and without BTBT (taken at $t = 1\ s$). (b) BTBT rate taken at different time at the drain-channel junction ($1\ nm$ below the gate oxide-body interface).

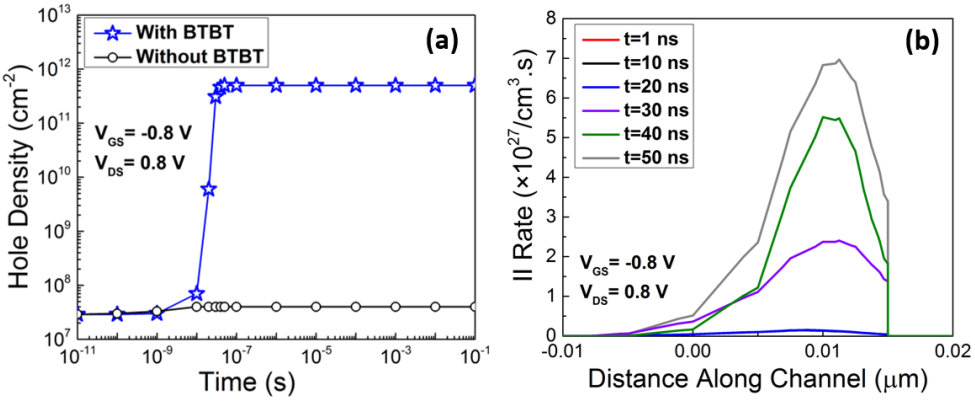

FIGURE 4.6: (a) Hole density ($20\ nm$ below the gate oxide-body interface) as a function of time at $V_{GS} = -0.8\ V$ and $V_{DS} = 0.8\ V$ with and without BTBT. (b) Impact ionization rate at the drain-channel junction taken at different time instances.

Fig. 4.6(a) shows the temporal evolution of the hole density integrated over the cut line with and without BTBT. Fig. 4.6(b) shows the temporal evolution of the impact ionization rate. From these figures, the following can be inferred. As the programming bias in such low values is applied to the device, the valence band of the channel region is aligned to the conduction band of the drain, leading to electron tunnelling from channel

to the drain. This increases the hole concentration in the channel region, leading to a lowering of the bands in the channel. This leads to higher electron injection to the channel from the source. This leads to an increase in impact ionization rate, which in turn increases the hole concentration in the channel region. In the meanwhile, the BTBT rate decreases as the channel valence band to drain conduction band overlap decreases.

There are several experimental demonstration of sub-band gap impact ionization [139–142] and low voltage BTBT [51–53] in silicon devices. Anil et al. reported impact ionization in n-channel MOSFETs for drain voltage as low a $0.6\ V$ [139], and Das et al. reported impact ionization in n-i-p-i-n diodes at $0.2\ V$ [141]. Huaung et al. demonstrated silicon tunnel FETs operating at low voltages [51].

4.4 Z-RAM Biasing

There are four operations in a Z-RAM cell. These are write '1' (programming), write '0' (erasing), read, and hold. The source and substrate are grounded for all operations. The biasing scheme for the proposed TiO_2 source/drain Z-RAM cell is shown in Fig. 4.7.

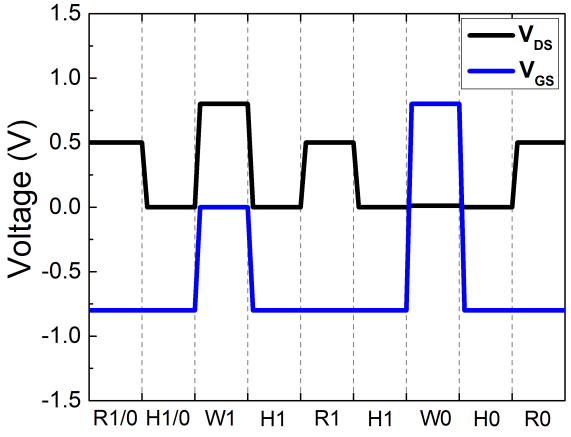

FIGURE 4.7: Biasing scheme for the proposed TiO_2 source/drain Z-RAM cell. 'W','R' and 'H' stands for 'Write','Read' and 'Hold'.

4.4.1 Programming

The parasitic BJT based programming method [54, 55] is used for the proposed memory cell. Initially, just before programming, the device operates in the accumulation mode with $V_{GS} = -0.8\ V$ and $V_{DS} = 0.8\ V$. Excess hole storage at the body is initiated by BTBT and later on, it is taken over by impact ionization as discussed in the previous section. At the onset of programming, to increase the body potential, V_{GS} is lifted up from $-0.8\ V$ to $0\ V$ as shown in Fig. 4.7. Once the BJT is turned on, impact ionization rate increases and current starts increasing due to the regenerative nature of this process.

4.4.2 Erasing

To write '0', we need to remove or erase the excess holes from the floating body. There are two ways by which the excess holes at the body can be removed: (i) by forward biasing the drain to body junction [143], and (ii) by gate coupling [54]. In our design, the gate coupling method is used for erasing or writing state '0'. The biasing scheme for write '0' is $V_{GS} = 0.8\ V$ and $V_{DS} = 10\ mV$.

4.4.3 Read

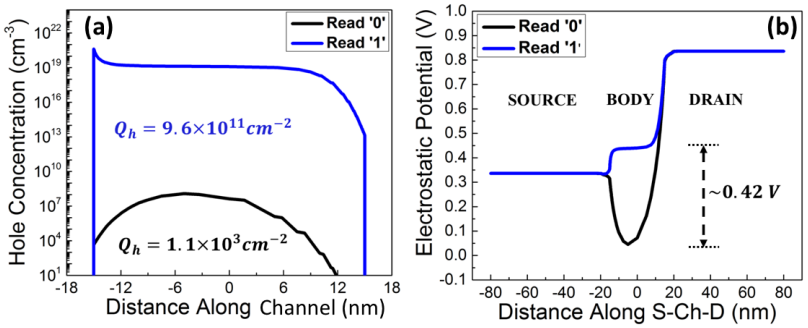

FIGURE 4.8: (a) Comparison of the excess hole concentration (cm^{-3}) at the body (at $y = 20\ nm$ i.e., $20\ nm$ below the body-gate oxide interface) of the proposed memory cell during read '1' and read '0' operation. Q_h is the excess hole density (cm^{-2}) at the body and it has been calculated by integrating the corresponding hole concentration curve. (b) Comparison of electrostatic potential at the body (at $y = 20\ nm$ i.e., $20\ nm$ below the body-gate oxide interface) of the proposed memory cell during read '1' and read '0' operation. There is a difference in body potential by $420\ mV$ between the two states.

For read '1' or '0', a comparatively low drain bias is applied so that, impact ionization is avoided. As the excess holes generated during write '1' operation will still

be present in the body, the body potential would be higher than the steady-state value. This reduces the threshold voltage leading to a higher drain current than the steady-state value. The biasing scheme for read operation is $V_{GS} = -0.8\ V$ and $V_{DS} = 0.5\ V$. Excess hole concentration (cm^{-3}) is plotted along the body ($y = 20\ nm$) during read state'1' and read state '0' as shown in Fig. 4.8(a). We have integrated the respective hole concentration curves to calculate the excess hole density (cm^{-2}) and it is seen that the excess hole density during read state '1' ($9.6 \times 10^{11}\ cm^{-2}$) is more than eight orders of magnitude higher than that of the excess hole density during read state '0' ($1.1 \times 10^{3}\ cm^{-2}$). Due to higher hole density at the body, the electrostatic potential at the body ($y = 20\ nm$) during read state '1' will be higher (by $420\ mV$) than the electrostatic potential during read state '0' as shown in Fig. 4.8(b). Due to a change in the electrostatic potential, there will be a difference in threshold voltage between state '1' and state '0'. Hence, the sense margin of the proposed TiO_2 source/drain cell is anticipated to be much higher.

4.5 Transient Analysis

To get the retention characteristics, transient simulations are performed for the proposed Z-RAM cell.

4.5.1 Optimization of Drain Bias For Reading Operation

V_{DS} is optimized for read operation by keeping V_{GS} fixed at $-0.8\ V$ as shown in Fig. 4.9.

As can be seen from Fig. 4.9, For a drain bias $V_{DS} \leq 0.5\ V$, the read state '1' current dies down very quickly. So, here we have chosen $V_{DS} = 0.5\ V$ as the optimized value for the read operation.

4.5.2 Optimization of Write '0' Time

Due to a large valence band offset ($\Delta E_V \approx 2\ eV$) between TiO_2 and Si, the stored holes during programming cannot be removed easily. So, writing state '0' will take a longer time than writing state '1'. We have optimized the writing state '0' time (t_{w0}) as shown in Fig. 4.10.

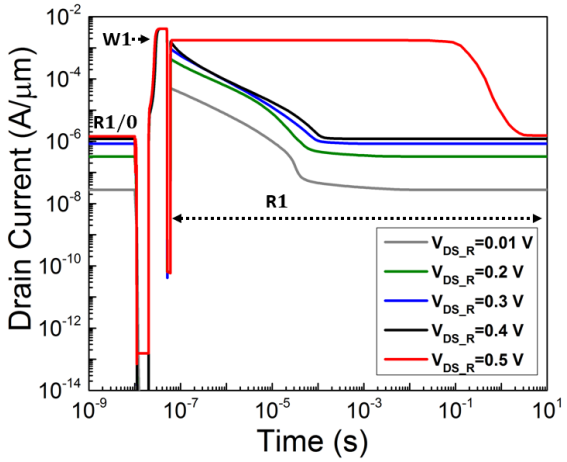

FIGURE 4.9: Optimization of V_{DS} for reading operation. V_{GS} is kept fixed at $-0.8\ V$.

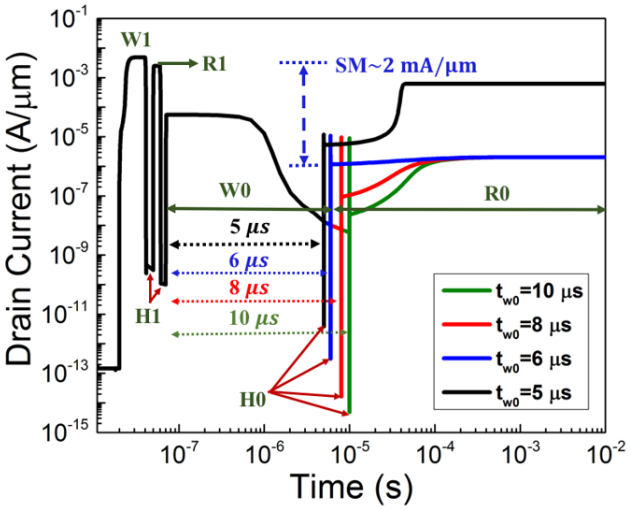

FIGURE 4.10: Drain current transients in the sequence of operation with the given biasing scheme under different write '0' time (t_{w0}): Write '1': $V_{GS} = 0V$, $V_{DS} = 0.8V$. Hold: $V_{GS} = -0.8V$, $V_{DS} = 0V$ Read: $V_{GS} = -0.8V$, $V_{DS} = 0.5V$.

As shown in the figure, for $t_{w0} < 6\ \mu s$, state '0' read current latches up. So, the minimum write state '0' time is taken as $t_{W0} = 6\ \mu s$. It is seen that write '0' time increases dramatically when $\Delta E_V > 0.3\ eV$ [67] as shown in Fig. 4.11 below. This is because, once the valance band offset between body-source/drain increases, the stored holes at the body can't escape. SRH recombination then becomes the only mechanism of charge removal from the body.

So, it takes longer time to remove those charges from the body and write '0' time increases. SRH recombination rate (R) can be given as,

$$R = \frac{np - {n_i}^2}{\tau_{p0}(n + n') + \tau_{n0}(p + p')} \tag{4.1}$$

Where,
$n^{'} = n_i e^{\frac{E_T - E_i}{k_B T}}$
$p^{'} = n_i e^{\frac{E_i - E_T}{k_B T}}$
n and p are the electron and hole concentrations in cm^{-3},
n_i is the intrinsic carrier concentration in cm^{-3},
τ_{p0} and τ_{n0} are the hole minority carrier lifetime and electron minority carrier lifetime respectively, and
E_T is the trap energy level.

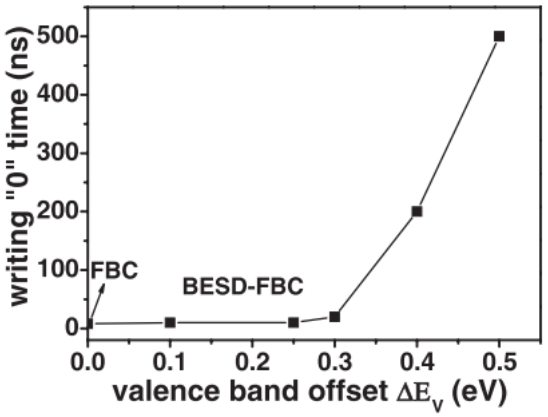

FIGURE 4.11: Writing '0' time dependence of valence band offset (ΔE_V). $\Delta E_V = 0$ eV means normal floating body cell (FBC). The writing time increase dramatically for band-gap engineered source/drain FBC when $\Delta E_V > 0.3\ eV$ [67].

If excess charge at the body is denoted by ΔQ, and change in the body potential due to that excess of charge is denoted by ΔV_B, then, ΔV_B = ΔQ/ C_{Si}, where, C_{Si} is the silicon body capacitance. But once we reduce the device dimensions and go in the lower technology nodes, the excess charge at the body (ΔQ) comes under an increased influence of source/drain potential which results in drain induced barrier lowering (DIBL) [144]. Due to DIBL, a greater number of electrons enter the body from the source. As a result, electron concentration at the body increases. Since the SRH recombination rate is directly proportional to the product of the electron and hole concentrations, the SRH recombination rate increases at the body due to the increased effect of DIBL as we go to lower technology nodes. So, the excess charge (ΔQ) at the body reduces faster due to an increase in the SRH recombination rate as we go to lower technology nodes. As a result, write '0' time also goes down as we go to lower technology nodes.

As shown in the Fig. 4.10, there is a sense margin of 2 $mA/\mu m$ at $T = 300\ K$ for the proposed TiO_2 source/drain Z-RAM cell.

4.5.3 Retention Characteristics

Fig. 4.12(a) shows the retention characteristics for the proposed Z-RAM cell for both $T = 300\ K$ and $T = 358\ K$. In our memory design, the retention time (RT) is defined by the time by which the state '1' current (I_1) reaches $2\mu A/\mu m$ from it's initial value. As seen in the figure, the extracted retention time for the proposed Z-RAM cell is 2 s and 70 ms at $T = 300\ K$ and $T = 358\ K$ respectively.

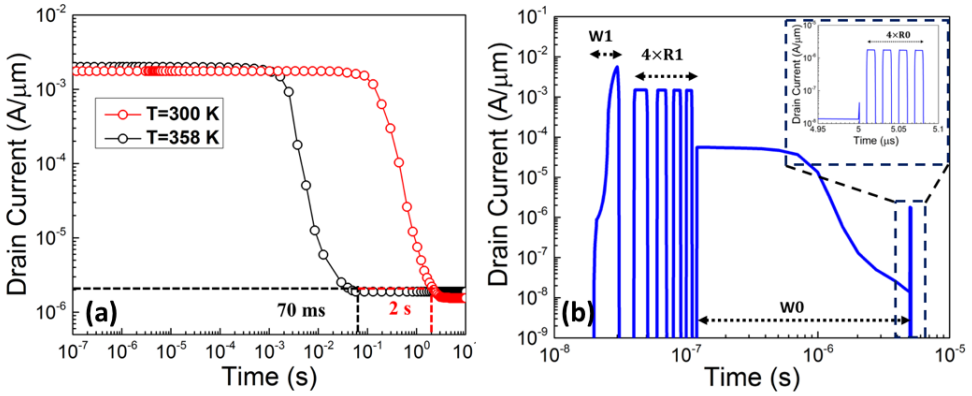

FIGURE 4.12: (a) Variation of read '1' current as a function of time at both $T = 300\ K$ and $T = 358\ K$ respectively. (b) Multiple reading operation of the proposed cell at $T = 300\ K$. This shows the non-destructive read-out mechanism in the proposed Z-RAM cell.

The read '0' currents are seen to be nearly constant at both the temperatures. However, Shockley-Read-Hall recombination rate increases with temperature leading to a faster decay of I_1 at $T = 358\ K$. As a consequence, the retention time of the memory cell at $T = 358\ K$ decreases. But even at $T = 358\ K$, the retention time is 70 ms for our proposed memory cell which is 1.1 times the ITRS specification of 64 ms.

In a conventional DRAM cell, after every read operation, the capacitor has to be charged again. One of the advantages of Z-RAM cells is that the read operation is non-destructive for several read cycles. Fig. 4.12(b) shows the multiple non-destructive reading operation for both state '1' and state '0'. Results from literature reports are summarised in Table 4.2 for comparison.

TABLE 4.2: Performance of the TiO_2 S/D 1T-DRAM cell and comparison to some of the results from the literature.

References	Year	Lg(nm)	Prog. Mech.	Prog. Bias	Sense Margin ($\mu A/\mu m$)	Retention Time	
						300 K	358 K
TiO_2 S/D Z-RAM (Simulation) (Present Work)	**2019**	**30**	**Parasitic BJT**	**V_{DS}= 0.8 V, V_{GS}= 0 V**	**2000**	**2 s**	**70 ms**
All-Si Z-RAM (Simulation) (Present Work)	2019	30	Parasitic BJT	V_{DS}= 0.8 V, V_{GS}= 0 V	6	20 ms	4 ms
All-Si 1T-DRAM[65] (Experiment)	2004	75	II	V_{DS}=1.2 V, V_{FG}= 0.8 V V_{BG}= -10 V	47	100 ms	—-
Si_xC_{1-x} S/D 1T-DRAM[67] (Simulation)	2010	100	II	V_{DS}= 1.8 V, V_{GS}= 0.6	99.8	100 ms	—-
Si_{1-x} Ge_x QW 1T-DRAM[68] (Simulation)	2010	250	II	V_{DS}=1.2 V, V_{FG}=1 V V_{BG}= -1.5 V	60	500 ms	10 ms
GaP S/D 1T-DRAM[66] (Simulation)	2013	55	Parasitic BJT	V_{DS}= 2 V, V_{GS}= -1.5 V	140	10 s	1.5 s
1T-1C DRAM (LPDDR4X by Samsung[123])	2019	20 (1y Node)	—-	1.8 / 1.1 / 0.6 V	—-	Minimum 64 ms	

As per our simulations, the proposed structure would result in lower programming voltage and better retention characteristics than other hetero-structure based Z-RAM cells reported in the literature.

For a scaled device down to sub-30 nm, random dopant fluctuations (RDF) is the main source of variability [145] in threshold voltage and this can also create a fluctuation in the retention time of the Z-RAM cell. But appropriate channel engineering may also help to reduce the random dopant fluctuation effects. It has already been demonstrated in the literature, that the introduction of a low-doped epitaxial layer in the device channel significantly suppresses the threshold voltage fluctuation [145] .

4.5.4 Disturbance Analysis

We have carried out disturbance analysis due to the influence of the neighbouring cells. Fig. 4.13 shows the TiO_2 S/D Z-RAM cell array where the drain and gate of every single cell are connected to the bit line and word line respectively.

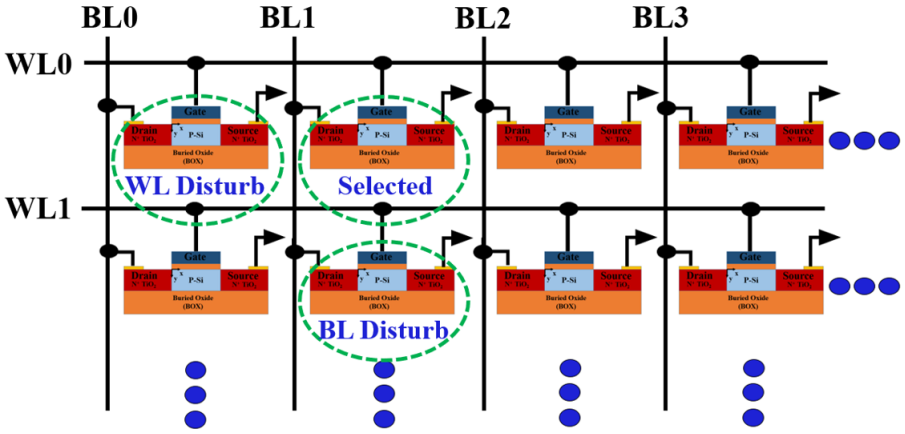

FIGURE 4.13: TiO_2 S/D Z-RAM cell array and disturbances among neighbouring cells.

Bit line disturb and word line disturb are the two types of disturbances from the neighbouring cells that can affect the performance of a selected cell. We have selected one Z-RAM cell and did the disturbance analysis due to the neighbouring cells as shown in Fig. 4.13. In Table 4.3, we have also compared the retention time of a disturbed cell of the TiO_2 S/D Z-RAM cell array with the same from an all-Si Z-RAM cell array at $T = 300$ K and 358 K.

The disturbance analysis is done individually when the selected cell is either in Hold '1' or Hold '0' state. The disturbing cells (BL disturb and WL disturb) are kept in

TABLE 4.3: Simulated retention time of the disturbed cell is represented as a percentage of the undisturbed cell retention time for $T = 300\ K$ and $358\ K$. Green color represents greater than or equal to 90%, yellow color represents greater than 50% but less than 90% and red color represents less than 50%.

Device Condition	**Disturb Source**	**Disturb Operation**	**Retention Time %**			
			TiO_2 S/D Cell		**All-Si Cell**	
			300 K	**358 K**	**300 K**	**358 K**
Hold 1	BL	Program	93.7	92	74	69
		Erase	85	83.5	41	35
		Read	94.1	93	71	70
	WL	P/E/R	96	95.2	82	77
Hold 0	BL	Program	98.1	97	65	63
		Erase	95	94.5	74	71
		Read	96.8	95.1	48	47
	WL	P/E/R	94.7	94.3	49	49

any one of the program, erase, or read conditions. This implies, if the disturbing cells are in the programmed state, the corresponding programming voltages are applied at the WL and the BL respectively.

It can be seen from Table 4.3 that, except for bit line 'erase' disturbance, the retention time of the disturbed cell is more than 90% of the undisturbed cell for TiO_2 S/D Z-RAM cell array. It is also seen that, for all-Si Z-RAM cell array, the retention time of the disturbed cell for most of the disturbances is less than 90% and for some cases, it is even less than 50% of the undisturbed cell. It clearly shows the superiority of the TiO_2 S/D Z-RAM cell array as compared to the all-Si Z-RAM cell array.

The main reason for the better retention characteristic of the TiO_2 S/D Z-RAM cell as compared to the all-Si Z-RAM cell is the large valence band offset between TiO_2 and Si ($\Delta E_V \approx 2\ eV$). This is due to the fact that the charge dynamics inside the body region does not change much in TiO_2 S/D Z-RAM cell as a result of high valence band offset between TiO_2 and Si. The retention time here mainly depends on the SRH recombination inside the body region unlike in an all-Si Z-RAM cell where the retention time depends on both SRH recombination at the body as well as over-the-barrier leakage between body and source/drain. Charge loss is more prominent for an all-Si Z-RAM cell as compared to TiO_2 S/D Z-RAM cell.

4.6 Possible Fabrication Process Steps

The proposed device may be fabricated using the scheme shown in Fig. 4.14. The process flow is similar to Intel's replacement gate process for 45 nm technology node [146].

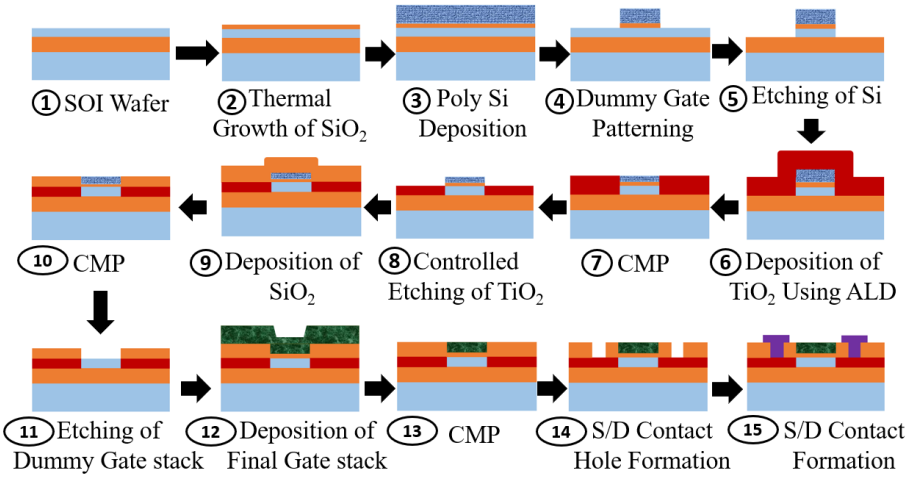

FIGURE 4.14: Proposed fabrication process flow of the TiO_2 S/D Z-RAM Cell.

First, an SOI wafer with desired specifications is taken. Then depending on the doping requirement of the p-type transistor body, ion implantation is done. To activate the dopant atoms annealing is done next. Then SiO_2 is grown using thermal oxidation process. As poly-Si is better than metal for self-aligned gate due to its high melting temperature, poly-Si is deposited on top of thermally grown SiO_2 which is followed by gate patterning using lithography. After lithography, Si is etched out from the selective areas by the anisotropic etching process. Once etching is done, TiO_2 is deposited by the atomic layer deposition process to form source/drain . To planarize the top surface, the CMP process is done which is followed by controlled etching of TiO_2 to make source, body, and drain of the same thickness. Then SiO_2 is deposited throughout and again CMP is done planarize the top surface. Dummy poly-Si gate stack is removed by etching and the final gate stack is deposited which is followed by the CMP process. Then source/drain contact holes are formed using lithography and etching. Finally, source/drain contacts are made.

4.7 Conclusion

A TiO_2 S/D Z-RAM cell on FD-SOI device architecture with low programming voltages, better retention characteristics, and better sense margin than heterojunction based Z-RAM reported in the literature, is proposed. Using well-calibrated TCAD simulations, the performance of the proposed structure are extracted and compared to literature data. The retention time predicted for a $30\ nm$ channel length device is seen to be better than the ITRS specification. Also, low voltage programming has been demonstrated for the proposed cell. In the context of the recent advancements in the deposition of ultra-thin TiO_2 and excellent passivation of Si surface using TiO_2, the proposed device architecture is very realizable for low power dynamic memory applications.

Chapter 5

TCAD Implementation of Bulk finFET Based Artificial Neuron For Spiking Neural Networks

5.1 Introduction

The main goal of neuromorphic computing or artificial neural network is to mimic the functionalities of a human brain. Spiking neural network (SNN) [92] is the new generation of artificial neural networks which is more energy-efficient than other neural networks as well as a traditional computer. An artificial neuron and an artificial synapse are the two fundamental building blocks of a spiking neural network. Here, we mainly focus on the realization of an electronic neuron for spiking neural networks. For the realization of an electronic neuron, it is essential to understand the basic functionalities of a biological neuron. In the next section, we discuss the operating principle of a biological neuron. Later, we demonstrate through well-calibrated TCAD [122] simulations, a highly scalable bulk FinFET based analog implementation of the integrate block of a LIF neuron. The charge integration mechanism is shown at the body of the proposed device through the addition of a buried n^+ layer. We also demonstrate that the proposed neuron is seen to have a spiking frequency in the MHz range which is 5 order of magnitude higher than that of a biological neuron ($f_0 = 10\ Hz$) and the lowest energy consumption for the integrate block till date i.e., 6.3×10^{-15} J/spike.

5.2 Operating Principle of a Biological Neuron

A neuron is the fundamental unit of human brain (Fig. 5.1(a)). Human brain consists of 10^{11} number [147] of neurons. To understand how brain functions, it is essential to know the working principle of a neuron. Each neuron consists of three parts (Fig. 5.1(b)) namely the cell body or soma, axon, and dendrite.

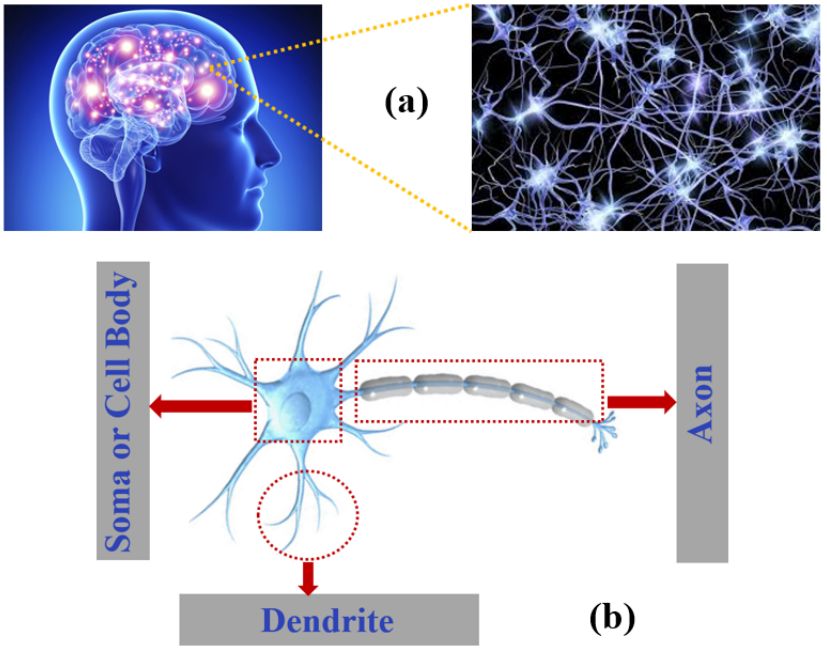

FIGURE 5.1: (a) Human brain consists of billions of neurons. (b) Each neuron consists of three parts and they are cell body, axon, and dendrite. (Images are taken from [148]).

Like other cell bodies, a neuron's soma contains a nucleus and specialized organelles. It's enclosed by a membrane that protects it and allows it to interact with its immediate surroundings. An axon is a long, tail-like structure that joins the cell body at a specialized junction called the axon hillock. Many axons are insulated with a fatty substance called myelin. Myelin helps axons to conduct an electrical signal. Neurons generally have one main axon. Dendrites are fibrous roots that branch out from the cell body. Like antennae, dendrites receive and process signals from the axons of other neurons. Neurons can have more than one set of dendrites, known as dendritic trees. The number of dendrites generally depends on their role. For instance, Purkinje cells are a special type of neuron found in the cerebellum. These cells have highly developed dendritic trees which allow them to receive thousands of signals. Neurons send signals using action potentials. An action potential is a shift in the neuron's electric potential caused by the flow of ions in and out of the neural membrane.

Each neuron is connected to 10^4 number [149] of other neurons. The connection between two neurons is called a synapse (Fig. 5.2). So the total number of synapses in a human brain is 10^{15}. Information is transferred from one neuron to another in the form of a current spike via a synapse. These 10^{11} neurons together with 10^{15} synapses form a spiking neural network (SNN).

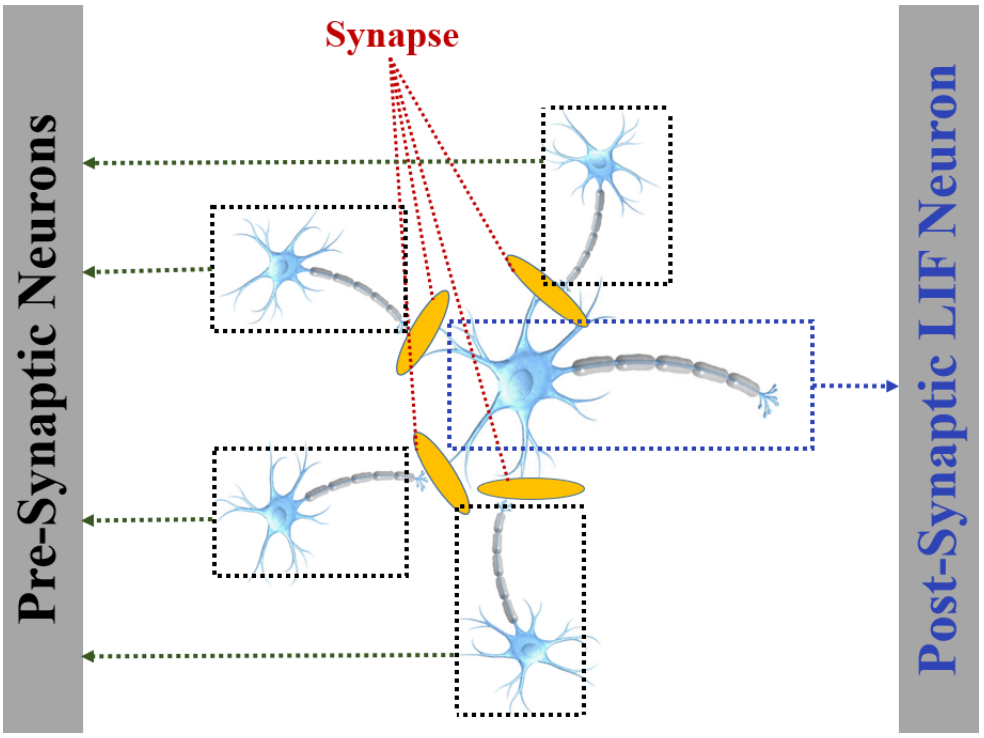

FIGURE 5.2: Pre-synaptic and post-synaptic neurons are shown inside black and blue dotted rectangle respectively. Each post-synaptic neuron is connected to many other pre-synaptic neurons as can be seen. The connection between two neurons is called synapse. (Neuron images are taken from [148]).

Fig. 5.3 shows the schematic diagram of a SNN. As shown in the diagram, several weighted input signals from the pre-synaptic neurons are coming to the post-synaptic LIF neuron through the synapse. These signals are summed up at the post-synaptic LIF neuron. Fig. 5.4(a) shows the simplest circuit model [103] of a LIF neuron. $I_{in}(t)$ is the summed-up input current at the LIF neuron at any given time. This $I_{in}(t)$ increases the potential of the LIF neuron with time and this is modelled by a leaky capacitor i.e., a capacitor in parallel with a resistor. At certain $I_{in}(t) = I_{th}$, the capacitor voltage ($V_c(t)$) reaches a threshold ($V_c(t) = V_{th}$). At, $V_c(t) \geq V_{th}$, $V_c(t)$ resets itself to resting potential and a spike is generated as shown in Fig. 5.4(b). With the increase in $I_{in}(t)$, $V_c(t)$ reaches V_{th} faster and spike frequency increases. Fig. 5.4(c) shows the output spike frequency (f_0) vs. input current curve. This is a signature characteristic of a biological neuron and this is to be mimicked artificially.

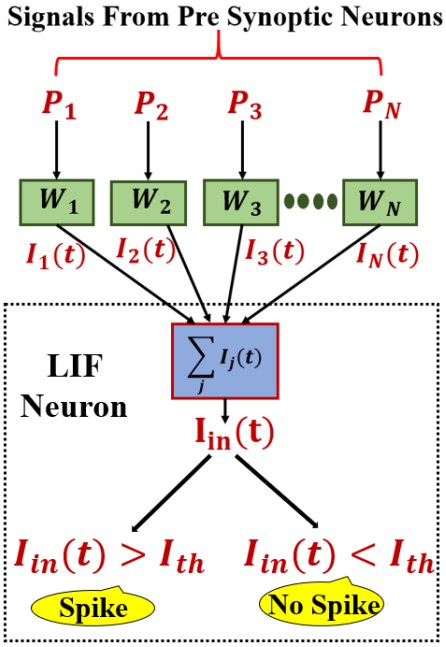

FIGURE 5.3: Schematic representation of a spiking neural network (SNN) with several pre-synaptic neurons and one post synaptic LIF neuron. Weighted signals are coming from pre-synaptic neurons through synapse and integrated in the LIF neuron. The algorithm for SNN is taken from reference [150].

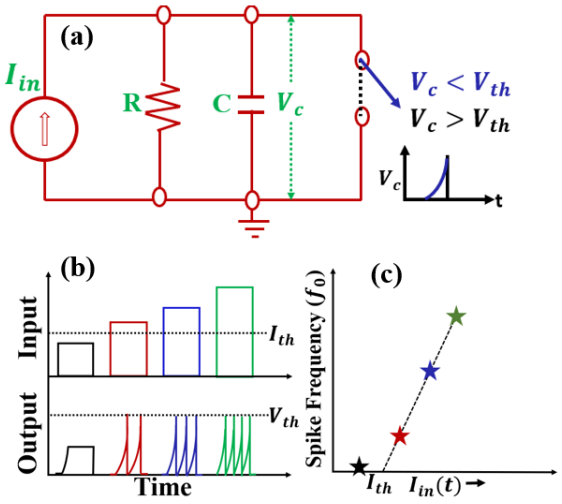

FIGURE 5.4: (a) Simplest model [103] of the LIF neuron with first order R-C circuit. (b) $V(t)$ will never exceed V_{th} as long as $I_{in}(t) < I_{th}$. Hence, there will not be any spike. But when $I_{in}(t)$ crosses I_{th}, the LIF neuron fires and creates a spike the moment $V(t) \geq V_{th}$ and immediately resets itself to resting potential after that. (c) As long as $I_{in}(t) < I_{th}$, spiking frequency (f_0) is zero. But with the increase in $I_{in}(t)$ after I_{th}, f_0 increases. This output spiking frequency (f_0) versus input curve is the signature of a biological neuron and it is to be mimicked artificially.

5.3 Bulk FinFET Based LIF Neuron

The schematics of the proposed device are shown in Fig. 5.5. The device would implement the integration function of the neuron to produce the current spikes. It is

assumed that appropriate circuits for the summing of the weighted outputs of the pre-synaptic neurons are added by a summing amplifier circuit as can be seen in Fig. 5.9.

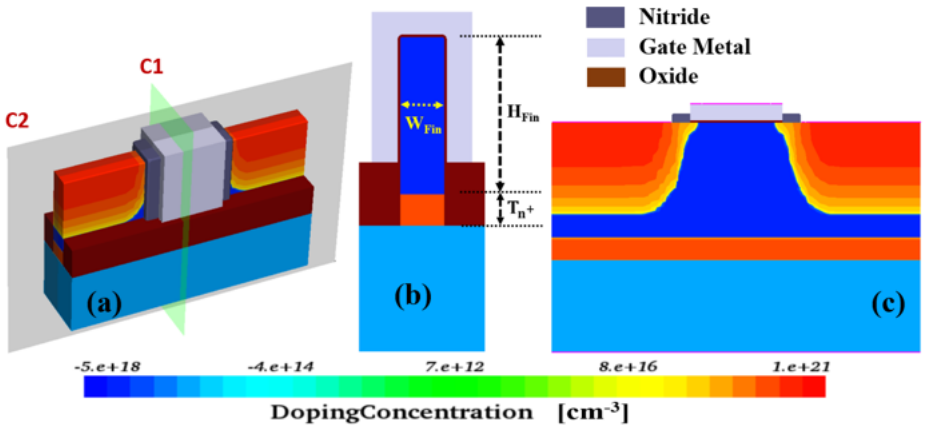

FIGURE 5.5: (a) Simulated bulk FinFET with n^+ buried layer. (b) $2D$ structure along $c1$. (c) $2D$ structure along $c2$. Gaussian doping profile with a peak concentration of 10^{21} cm^{-3} is used at source/drain and the channel is uniformly doped. Doping gradient along source/drain to channel is kept at 2 $nm/decade$. The +ve (-ve) sign indicates n-type (p-type) doping.

5.3.1 Device Design and TCAD Validation

The simulation models used are calibrated by matching (as can be seen in Fig. 5.6) the experimental $I_{DS} - V_{GS}$ characteristics taken from the gate-all-around nano-sheet FET device as described in the reference [151]. The device design parameters used to simulate the proposed device are listed in Table 5.1.

TABLE 5.1: Bulk FinFET parameters used in TCAD simulation.

Parameter	Value
Gate length (L_g) (nm)	100
Fin height (H_{Fin}) (nm)	100
Fin width (W_{Fin}) (nm)	30
Buried n^+ layer thickness (T_{n^+}) (nm)	25
Gate oxide thickness (T_{ox}) (nm)	2
Channel p-doping (cm^{-3})	5×10^{18}
Buried n^+ layer doping (cm^{-3})	5×10^{19}
Source/Drain n^+-doping (cm^{-3})	10^{21}
Gate metal work-function (eV)	4.6

Table 5.2 shows the calibrated values of different parameters used in the TCAD simulation against the default values.

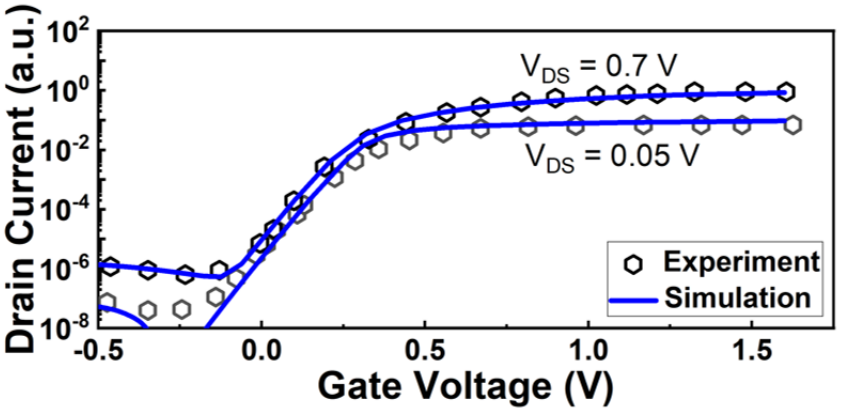

FIGURE 5.6: Comparison of simulated $I_{DS} - V_{GS}$ characteristics with experimental data [151], demonstrating the calibration of the simulation models used.

TABLE 5.2: Default and Calibrated parameters used in TCAD simulations.

Parameters	**Default Value**	**Calibrated value**
SRH Recombination τ_{max} and τ_{min}	10 μs, 0 for e 3 μs, 0 for h	1 μs, 0 for e 1 μs, 0 for h
Mobility $(cm^2/V-s)$ μ_{emax} and μ_{emin}	1417, 470	1417, 470
Band to Band Tunnelling A $(cm^{-1}s^{-1}V^{-2})$ and B $(Vcm^{-1}eV^{-1.5})$	$8.9e^{20}$, $2.1e^{7}$	$8.9e^{19}$, $2.1e^{7}$
Tunnelling Mass $m_e{}^*$ and $m_h{}^*$	$0.5m_0$, $0.5m_0$	$0.5m_0$, $0.5m_0$
Impact Ionization Avalanche Factors: for e and h	1, 1	0.95, 0.95
Velocity Saturation $Vsat0$ (for e and h)	$1.07e^{7}$, $8.37e^{6}$	$2.1e^{7}$, $8.37e^{6}$

The proposed device can be fabricated using the typical process flow reported in [152].

5.3.1.1 Functionality of Buried n^+ Layer

The purpose of the buried n^+ layer in this device is to create a barrier (Fig. 5.7(b)) for the excess majority carriers to confine them within the p-doped body region. When a comparatively large drain bias is applied, electron-hole-pairs are generated by impact ionization at the drain-body junction. Electrons are swept away to the drain and the excess holes come to the body region.

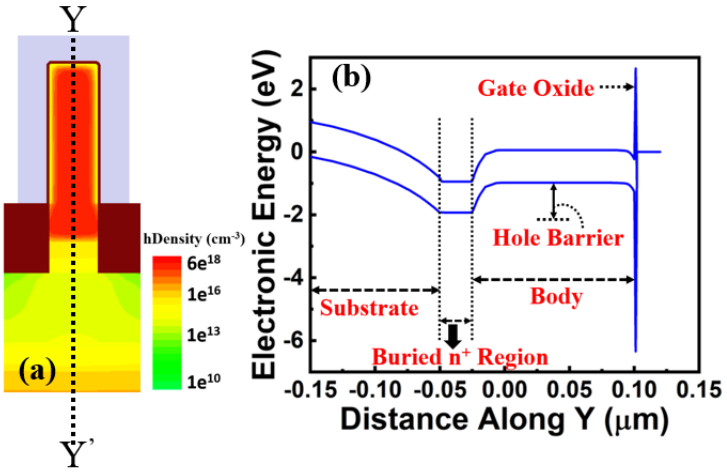

FIGURE 5.7: (a) Simulated contour plot of excess hole density at the body along $c1$. Corresponding biases are $V_{DS} = 2\ V$ and $V_{GS} = -1\ V$. (b) Energy band diagram along YY'. Buried n^+ layer, creates a barrier for excess majority carriers at the body.

Fig. 5.7(a) shows the contour plot of excess hole density at the body which clearly shows that the holes are being stored at the body.

5.3.1.2 Signature of Hole Storage

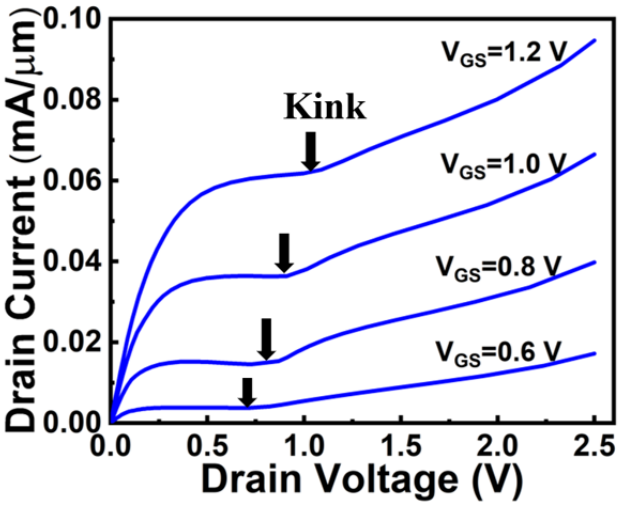

FIGURE 5.8: Kink arises in the output characteristics of the proposed device. This is a signature of hole storage at the body of the transistor.

Floating body effects [43–45] induced by impact ionization (II) at the drain-body junction causes the excess hole integration overtime at the body and this results in kink in the output characteristics of the proposed device as shown in Fig. 5.8.

5.3.2 Working Principle of Bulk FinFET Based LIF Neuron

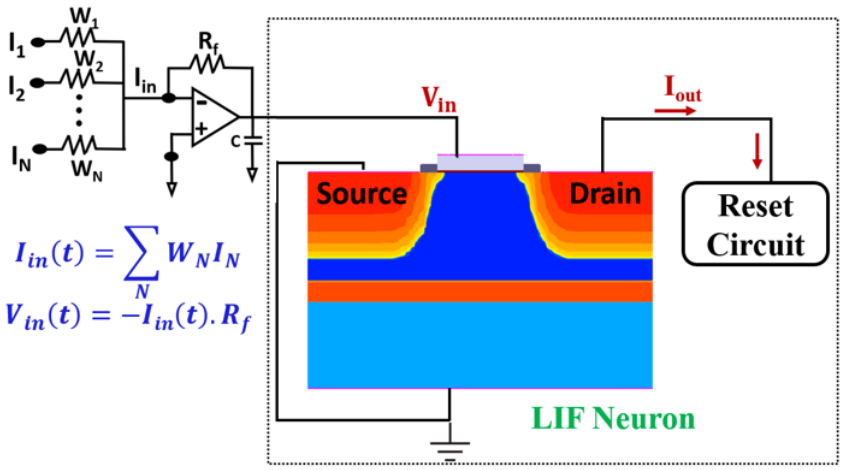

FIGURE 5.9: Simulated LIF neuron. Signals (I_1, I_2, ... I_N) from the pre-synaptic neurons are coming to the post-synaptic LIF neuron through synapses with synaptic weights $W_1, W_2, W_3, \ldots W_N$. As the proposed bulk FinFET based device takes voltage as input, the current is converted to a proportional voltage. $V_{in}(t) = -I_{in}(t)R_f$, where $I_{in}(t)$ is the summed-up current from the pre-synaptic neurons. I_{out} starts increasing at some $V_{in}(t) \geq V_{th}$ and $V_{DS} = V_{Integrate} = 3\ V$. As soon as I_{out} reaches I_{th}, the reset circuit resets V_{DS} to V_{Reset} for $t = (t_{Erase} + t_{RS})\ s$.

Fig. 5.9 shows the simulated LIF neuron with input voltage (V_{in}) and output current (I_{out}). As shown in Fig. 5.9, increasing I_{in}(t) causes V_{in}(t) to increase. As the proposed bulk FinFET based device takes voltage as input, the current has first to be converted to a proportional voltage. As can be seen in Fig. 5.9, I_1, I_2, ... I_N are the inputs from the pre-synaptic neurons and these signals pass through the synapses with synaptic weights W_1, W_2, W_3, ... W_N. This summed up current (I_{in}(t)) is then converted to a voltage using an inverting amplifier. The output of this inverting amplifier is the input (V_{in}(t)) of the LIF neuron.

V_{DS} is controlled by the reset circuit. Fig. 5.10 shows the schematics of the biasing scheme and output drain current for the LIF functionality. For charge integration by impact ionization, a high drain bias ($V_{DS} = V_{Integrate}$) is applied for $t_{Int}\ seconds$ and V_{in} is kept above V_{th}. I_{out} starts increasing due to positive feedback effect by excess hole storage at the body region and as soon as it reaches I_{th}, V_{DS} is reset by the reset circuit for $(t_{Erase} + t_{RS})\ seconds$ and a spike is generated. The same integrate and reset cycle repeat again to make each LIF cycle identical. Fig. 5.11 shows the energy band diagram of the proposed device along source-channel-drain at $t = t_1$ and t_4, i.e., for initial and reset conditions when there is no charge integration taking place, and Fig. 5.12 shows the energy band diagram at $t = t_2$ and t_3. It describes the barrier lowering

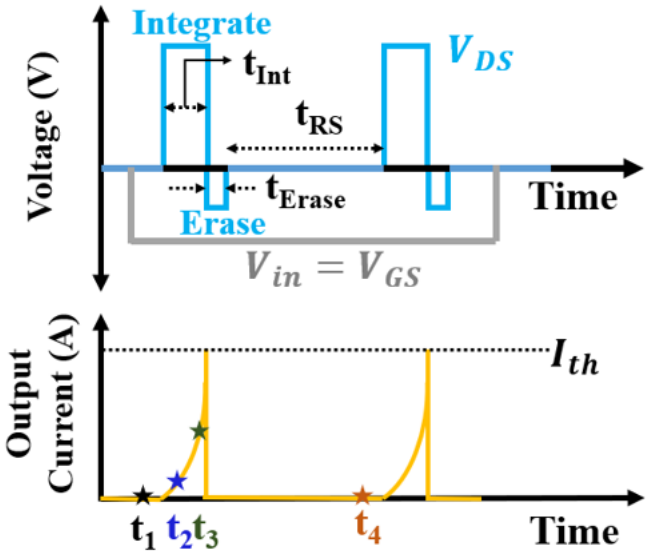

FIGURE 5.10: Schematics of the biasing mechanism for LIF functionality and corresponding output current I_{out}.

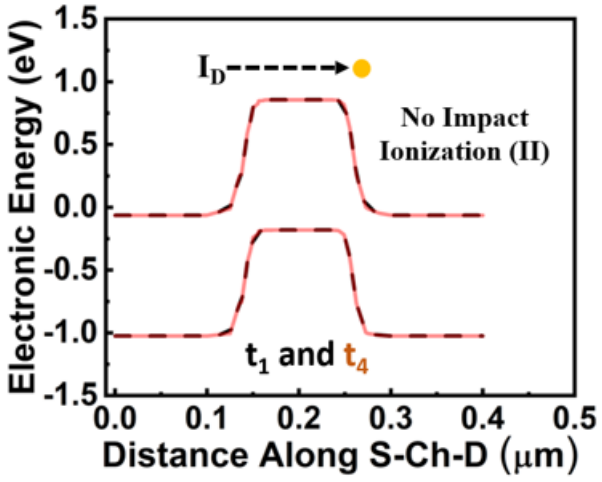

FIGURE 5.11: Energy band diagram of the proposed device along source-channel-drain at $t = t_1$ and t_4 i.e., for initial and reset conditions.

by charge integration mechanism which takes place due to impact ionization and charge leak by diffusion process from the body to source and SRH recombination at the body. It is important to note is that, while resetting, the drain-body junction can not be forward biased for long because this can also program the device as discussed in [59]. That is why a negative drain bias is applied for a small time then it is set to zero bias for a larger time. This drain pulse is generated by the reset circuit i.e., the drain bias is controlled by the reset circuit. The pulse width is set depending on the excess charge dynamics inside the device.

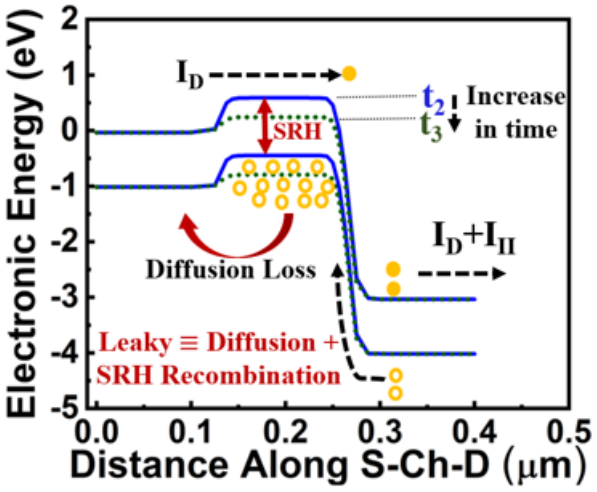

FIGURE 5.12: Energy band diagram of the proposed device along source-channel-drain at $t = t_2$ and t_3 which describes charge integration and leak phenomena.

5.3.3 Results and Discussion

There is a parasitic n^+ - p - n^+ BJT associated with the proposed n-channel bulk FinFET where source (n^+) acts as an emitter, body (p) as a base and drain (n^+) as a collector. When a comparatively large drain bias (V_{DS}) is applied along with a negative gate bias (V_{GS}), the transistor operates in accumulation mode and impact ionization (II) takes place at the drain-body junction due to the presence of a large electric field. The direction of the electric field is from drain to body. Electron-hole-pairs are generated at the depletion region of the drain-body junction. Electrons are swept away to the drain side and the holes come to the body. These excess holes at the body region increase the body potential or the base potential of the parasitic BJT. As a result, there is a barrier lowering for electrons from source to body or emitter to base. Hence a greater number of electrons enter into the body and take part in the impact ionization process at the drain-body junction causing an increase in impact ionization rate. More impact ionization means more electron-hole-pair generation and a greater number of hole accumulation at the body. This is a positive feedback process. Within a very short time, the base potential increases to a level which forward biases the base (body) – emitter (source) junction and turn the parasitic BJT on. We see a steep rise in drain current and eventually, the drain current latches up as shown in Fig. 5.13.

The impact ionization rate, hole density, and electrostatic potential along the source-channel-drain for different times are simulated as shown in Fig. 5.14.

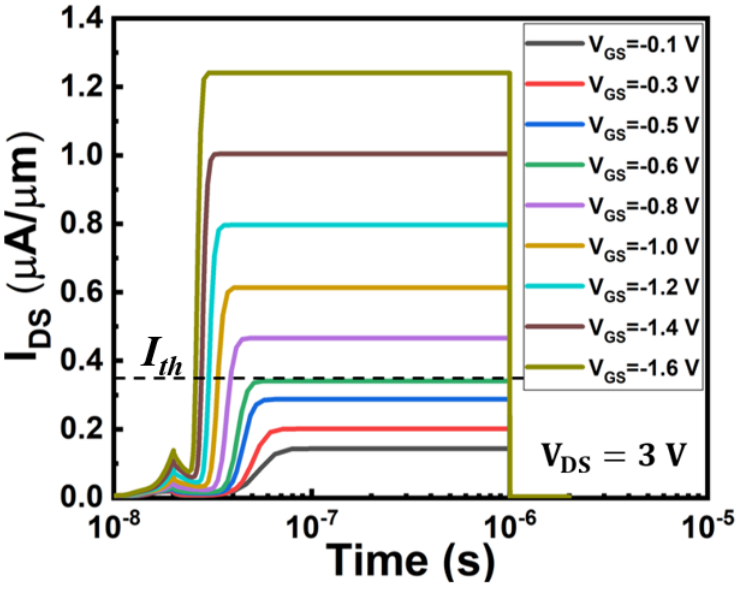

FIGURE 5.13: Transient simulation shows the I_{DS} - $Time$ characteristics under different V_{in}. The I_{th} is set at 0.35 μA/μm.

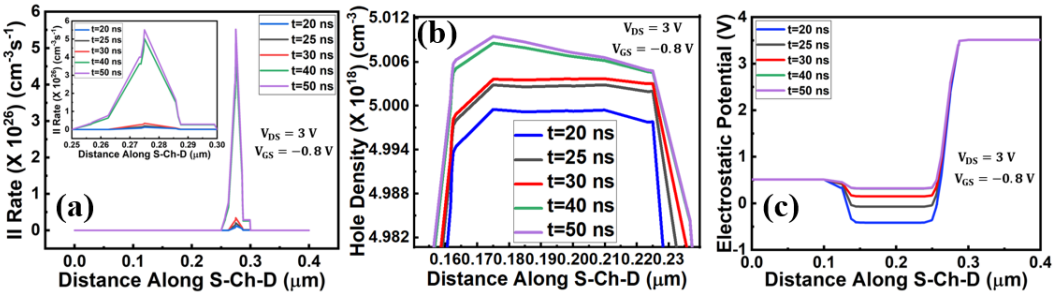

FIGURE 5.14: (a) Impact ionization (II) rate is plotted along source-channel-drain for different time. Inset is the zoomed figure. (b) Hole density is plotted along source-channel-drain for different time. (c) Variation of electrostatic potential along source-channel-drain for different time. The time reference is the same as in Fig. 5.13.

As shown in Fig. 5.14, due to the positive feedback process, impact ionization rate at the drain-body junction, hole density at the body, and the electrostatic potential at the body increases with time and saturates at some point.

Fig. 5.13 shows the temporal evolution of drain current overtime under different V_{GS} when V_{DS} is fixed at 3 V. It can be clearly seen that, with the increase in $|V_{GS}|$, the drain current reaches I_{th} faster. So, the triggering time depends on the input voltage of the proposed bulk FinFET based artificial neuron. The input voltage can be given as follows.

$$V_{in}(t) = V_{GS}(t) = -I_{in}(t) \times R_f = R_f.W_N I_N \tag{5.1}$$

From Fig. 5.13, and Fig. 5.14(a) and (b), it can be seen that, with the increase in the input voltage ($V_{in}(t)$), the triggering time decreases and the drain current reaches threshold level in less time. Hence, the frequency of spike generation increases. More the signals from the pre-synaptic neurons, less the time it requires to reach the threshold level and more the frequency of spike generation. Here the gate bias is negative (V_{in}(t)= V_{GS}(t) = I_{in}(t). R_f) which means the transistor operates in accumulation mode. A high number of holes are present at the body-gate oxide interface. These accumulation layer holes repel the excess holes at the body which are generated by impact ionization at the drain-body junction. As a result, the excess holes will accumulate away from the gate dielectric-body interface and SRH recombination is reduced and the retention time of excess holes is increased. This is unlike the case when a positive gate bias is applied and the transistor operates in inversion mode. Under inversion mode, a large number of electrons are present at the inversion layer which attracts these excess holes present at the body, and SRH recombination takes place.

For the proposed bulk-FinFET based artificial neuron, I_{th} is set at $0.35\ \mu A/\mu m$. V_{th} corresponding to I_{th} is $0.6\ V$. As long as $|V_{in}| = |V_{GS}| \leq V_{th} = 0.6\ V$ (Fig. 5.15(a)), drain current (I_{out}) saturates (Fig. 5.15(b)) before it reaches I_{th} = 0.35 μA/μm and no spike is generated.

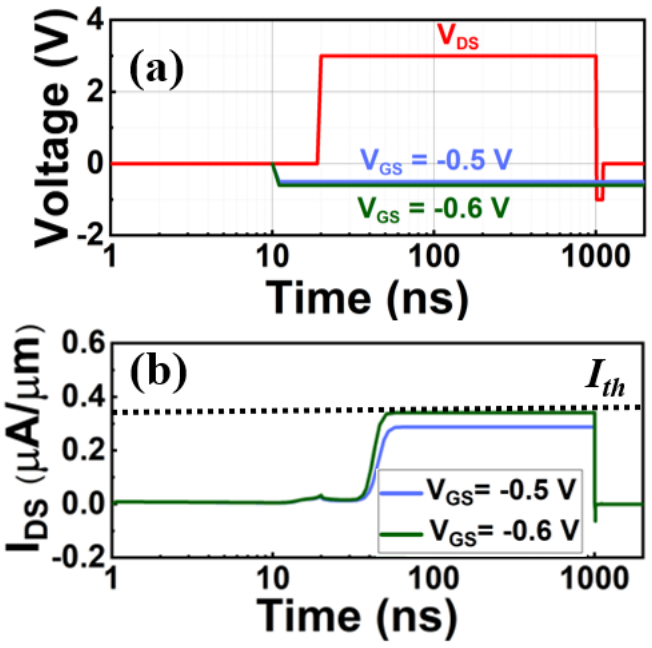

FIGURE 5.15: (a-b) Output drain current does not make any spike as long as $V_{in} \leq V_{th} = 0.6\ V$. Drain current is taken for $V_{in} = 0.5$ and $0.6\ V$ and it saturate before reaching I_{th}.

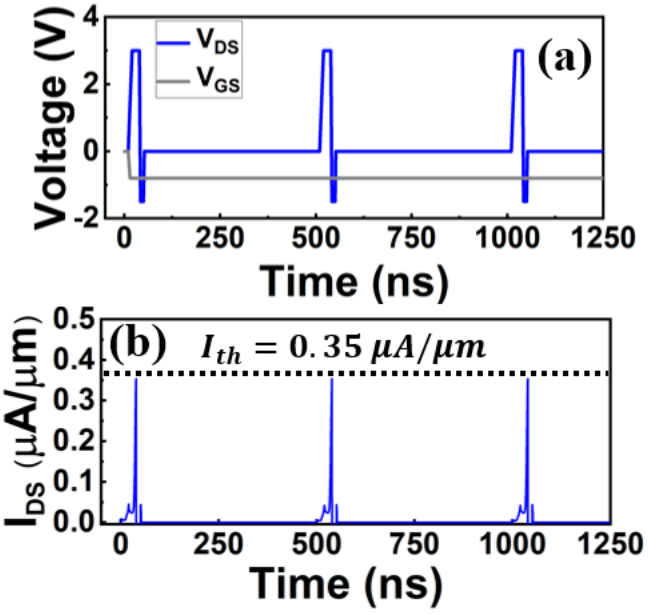

FIGURE 5.16: (a-b) Biasing scheme for neuron firing and reset of the proposed neuron. For $V_{in} = 0.8\ V$, when I_{DS} reaches $I_{th} = 0.35 \mu A/\mu$m, V_{DS} is reset by the reset circuit and a spike is generated.

But when $V_{in} = |V_{GS}| > V_{th} = 0.6\ V$ as shown in Fig. 5.16(a), as soon as the drain current reaches $I_{th} = 0.35\ \mu A/\mu m$, V_{DS} is reset to $-1.5\ V$ and a spike is generated (Fig. 5.16(b)). V_{DS} is kept at $-1.5\ V$ for $8\ ns$ and then changed to V_{DS} = 0 V for $500\ ns$ to remove all the excess holes present at the body, then V_{DS} is set back to 3 V again (Fig. 5.16(a)) and from the same initial point the current starts after each reset which essentially makes each LIF cycle identical. As shown in the Fig. 5.16(b), a spike is generated after every $520\ ns$ for the given biasing condition. Spiking frequency (f_0) is plotted against input voltage (V_{in}) in Fig. 5.17. As in biology, input signals from the pre-synaptic neurons control the spiking frequency, in the proposed bulk FinFET based electronic neuron, the spiking frequency is controlled by the input signal $|V_{GS}|$. The achieved spiking frequency for the proposed LIF neuron is in the range of MHz which enables attractive hardware acceleration [153] for neuromorphic computing.

Maximum energy/spike of the integrate block for the proposed LIF neuron is calculated using the following equation:

$$E_{spike} = V_{spike} \times I_{th} \times t_{spike} \tag{5.2}$$

It gives $E_{spike} = 3\ V \times 0.35\ \mu A/\mu m \times 6\ ns$ = $6.3 \times 10^{-15}\ J$ which is the lowest as compared to any other nano scale device based integrate block of artificial neurons

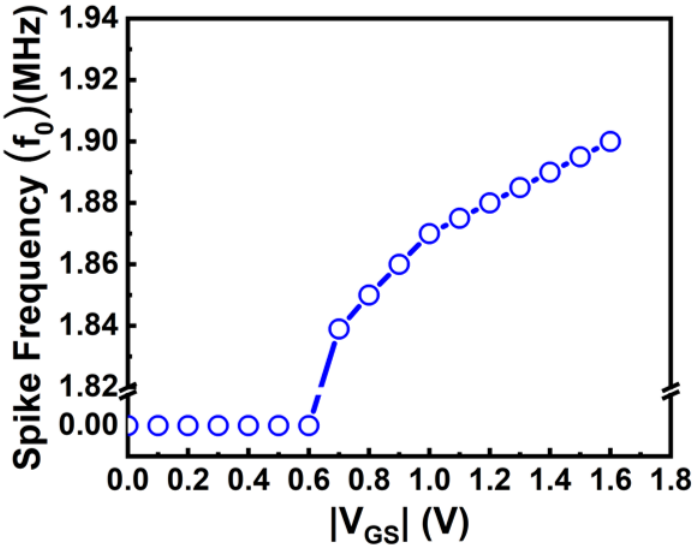

FIGURE 5.17: Spiking frequency (f_0) versus input ($|V_{GS}|$) shows that, for $V_{in} \leq V_{th} = 0.6\ V$, the frequency is zero while for $V_{in} \geq V_{th}$, f_0 increases with input bias.

reported in the literature [117, 119, 120].

5.3.4 Benchmarking

Integrate and reset are the two circuit blocks of a LIF neuron. Integrate block is replaced by the nano scale devices [119, 120] with an overall improvement in the area of at least $10X$ compared to CMOS circuit based neurons [108, 110]. The reset block is a circuit design challenge from an energy point of view. Sub-threshold design and operation can be one of the ways to improve the reset circuit which has been discussed extensively in [154].

TABLE 5.3: Comparison of the energy/spike and area of the nano-Scale devices for the integration function in neurons

References	**Experiment /Simulation**	**Device Type**	**Area (μm^2)**	**Energy/Spike (Integrator)**	**Energy/Spike (Reset circuit)**
T. Tuma et. al.[119]	Experiment	PCM	0.0132	$5 \times 10^{-12}\ J$	—-
S. Lashkare et. al.[117]	Experiment	PCMO RRAM	25	$4.8 \times 10^{-12}\ J$	—-
S. Dutta et. al.[120]	Experiment	SOI MOSFET	0.4	$13 \times 10^{-12}\ J$	$41 \times 10^{-12}\ J$
This Work	Device Simulation	Bulk FinFET	0.04	$6.3 \times 10^{-15}\ J$	—-

In the comparison Table 5.3, the proposed bulk FinFET based neuron is compared with the three other nano-scale device-based neurons namely, phase change memory (PCM) [119], $Pr_{0.7}Ca_{0.3}MnO_3$ (PCMO) RRAM [117], and PD-SOI MOSFET [120]. In phase change memory-based neuron, the integrate and fire functionality is realized using chalcogenide-based phase-change memory. The reset circuit is not fabricated along with the PCM based integrator block but is designed and implemented using off-the-shelf circuit components as described in the supplimentary information of reference [119]. In PCMO RRAM-based neuron, the integrate and fire functionality is realized by changing the resistance (high resistance state low resistance state) of the PCMO material with the application of proper bias. In this case also the reset operation is performed through an external circuitry as can be seen in reference [117]. In PD-SOI MOSFET-based neuron, the integrator functionality is experimentally demonstrated using a PD-SOI MOSFET. In this case the reset block is not experimentally demonstrated. To check the reset functionality, a compact model of the PD-SOI MOSFET (derived from the experimentally measured characteristics) along with a circuit model of a reset circuit is simulated.

To the best of my knowledge, no single device-based neuron which demonstrates all the neuronal functionalities (Integration, reset) is reported in the literature. The reset circuit is always external. In this thesis, integration functionality of a neuron is demonstrated using a bulk FinFET with a buried n^+ layer. In Appendix-B, a possible reset circuit is shown for the reset operation of the proposed bulk FinFET based neuron. But there is a scope to design a better reset circuit in terms of area and energy efficiency.

In the comparison Table 5.3, we compare only the energy and area efficiency of the nano scale devices replacing the leaky integration function of neuron (i.e. RRAM [117], PCM [119], PD-SOI-MOSFET [120]) with the proposed bulk FinFET based neuron. It can be clearly seen that, as compared to the other nano scale devices, the proposed bulk FinFET based integrator is more energy efficient with the lowest energy/spike of 6.3×10^{-15} J which transfers the bottleneck from low energy integrator design to low energy reset circuit design.

5.4 Conclusion

A highly scalable CMOS compatible n-channel bulk FinFET based ultra-low energy integrate block of an artificial neuron is proposed and demonstrated by well-calibrated

TCAD simulations. The proposed device consumes an energy of $6.3\ fJ$/spike which is almost three orders of magnitude lower than other nano scale device based LIF neurons reported. The proposed bulk FinFET based LIF neuron follows the signature of a biological neuron, i.e., spike frequency increases with the input voltage with a maximum frequency in the MHz range. Therefore, the proposed bulk FinFET based LIF neuron can be a potential building block for spiking neural network.

Chapter 6

Summary, Conclusion and Future Work

6.1 Thesis Summary

This chapter summarizes the important conclusions drawn from our work and lists out some of the open areas to explore further which could not be covered in the scope of this thesis.

In Chapter 2, we briefly introduced the hierarchy and classification of memory in a computer. Our main focus is on dynamic random access memory (DRAM). The challenges and limitations faced in traditional one transistor one capacitor DRAM are discussed. Z-RAM is introduced as a possible replacement of conventional DRAM. The drawbacks of an all-Si Z-RAM cell were discussed. To solve the problem of all-Si Z-RAM cell, TiO_2 source/drain Z-RAM cell is proposed. A brief introduction is given on neuromorphic computing. We discussed about the limitations of traditional von-Neumann architecture based computing system. To overcome the limitations of von-Neumann architecture based computing system, neural network based neuromorphic computing system is introduced briefly.

Since this thesis is based on the application of floating body effects, Chapter 2 provided a comprehensive overview of the floating body effects (FBE) which are seen in SOI MOSFETs. Kink effect, parasitic BJT effect and hysteresis effect are the three types of floating body effects which are seen in a SOI MOSFET and they are discussed in a greater detail.

In Chapter 3, an n-channel TiO_2 source/drain PD-SOI MOSFET based Z-RAM is introduced as a replacement of conventional DRAM. An improvement in both retention time and sense margin is shown through well calibrated TCAD simulations.

Using the framework of Chapte 3, in Chapter 4, TCAD simulations for n-channel TiO_2 source/drain FD-SOI based Z-RAM were performed based on the calibrated model parameters against measured data. Overcoming the limitations of PD-SOI based Z-RAM, FD-SOI based Z-RAM cell is shown to be more power efficient with a better sense margin and retention characteristics.

Since neuron is one of the two building blocks of a neural network, it is important to realise the functionality of a neuron in an electronic system. A bulk FinFET with buried n^+ layer based electronic neuron is proposed and investigated in Chapter 5.

6.2 Conclusion

In conclusion we can say that, two major aspects of the modern computing system are studied and investigated using well calibrated TCAD simulations. The first one is dynamic random access memory (DRAM) which is the primary memory of any computing system and the second one is the electronic neuron, a fundamental building block of an artificial neural network which drives the concept of neuromorphic computing. We have realized both the dynamic memory and electronics neuron by utilizing the MOSFET's floating body effects.

Overcoming the capacitor scalability problem of conventional DRAM cell, we have proposed TiO_2 source/drain based Z-RAM cell where instead of using physical capacitor, transistor body is used as a storage node. We have also shown that the proposed TiO_2 source/drain based Z-RAM cell exhibits higher retention time and non-destructive readability.

On the other hand, to make a potential building block for spiking neural networks, a highly scalable CMOS compatible n-channel bulk FinFET based ultra-low energy integrate block of an artificial neuron is proposed and demonstrated. The proposed device consumes an energy of $6.3\ fJ/spike$ which is almost three orders of magnitude lower than other nano scale device based LIF neurons reported in the literature. The proposed bulk FinFET based LIF neuron follows the signature of a biological neuron,

i.e., spike frequency increases with the input voltage with a maximum frequency in the MHz range.

In today's data driven world, realization of Z-RAM cell overcoming the storage capacitor scaling challenges in a conventional DRAM and realization of electronic neuron using highly scalable bulk-FinFET will help further in advancing the digital computation using artificial intelligence.

6.3 Future Work

The proposed n-channel TiO_2 source/drain SOI MOSFET based Z-RAM cell is based on TCAD study. Since there are variability issues for sub-20 nm technology nodes, variability study can be done on the proposed device. This TCAD study of TiO_2 Z-RAM can be further illustrated and investigated by experimental demonstration with the same technological parameters as used in simulations. This will lead to more concrete proof of acceptability of the proposed device as a possible replacement of the conventional DRAM cell.

In the second part of the thesis, a bulk FinFET based electronic neuron with lowest energy per spike is proposed and investigated as one of the building blocks of a spiking neural network which leads the way of neuromorphic computing. Since neuron and synapse are the two building blocks of a spiking neural network, it is necessary to realise a synapse along with the neuron. As scope for future work, an artificial electronic synapse should be realised using bulk FinFET to completely describe a spiking neural network. Besides that, experimental demonstrations are required for the bulk FinFET based neuron and synapse.

Appendix A

Codes for Simulations

A.1 Structure Editor code for TiO_2 S/D Z-RAM Cell:

```
(sde:clear)

(define DD1 "ArsenicActiveConcentration")
(define SD1 "ArsenicActiveConcentration")
(define CD "BoronActiveConcentration")

(define nm 1e-3)
(define W 0.1)
(define tsub 0.5)
(define tbox 0.40)
(define tsi 0.150)
(define Lg 0.05)
(define tox 0.0045)
;(define Tsd 0.040)
```

```
(define DopSD1 1e20)
(define DopSub 2e17)
(define DopSubstrate 1e19)

(define a (* 0.5 W))
(define b (* 1.0 tsub))
(define c (+ b tbox))
(define d (+ c tsi))
(define e (+ d tox))

(define f (* 1 (/ Lg 2)))
;(define g (+ e Tsd))

(define h (- f  tox))
(define i (+ f  0.02))

(define j (- f  0.002))
(define z (+ f  0.01))

;*************************** Creation of structure ********************************************

;(sdegeo:set-default-boolean "ABA")
(sdegeo:set-default-boolean "ABA")
(sdegeo:create-rectangle (position (* -1 a) 0  0) (position a (* -1 b) 0) "Silicon" "Substrate")
(sdegeo:create-rectangle (position (* -1 a) (* -1 b)  0) (position a (* -1 c) 0) "SiO2" "Oxide2")

(sdegeo:create-rectangle (position (* -1 a) (* -1 c) 0) (position (* -1 j) (* -1 d)  0) "GaP" "Source")
(sdegeo:create-rectangle (position (* -1 j) (* -1 c)  0) (position j (* -1 d) 0) "Silicon" "Channel")
(sdegeo:create-rectangle (position (* 1 a) (* -1 c)  0) (position (* 1 j) (* -1 d) 0) "GaP" "Drain")

(sdegeo:create-rectangle (position (* -1 f) (* -1 d)  0) (position f (* -1 e) 0) "SiO2" "Oxide1")


;(sdegeo:create-rectangle (position (* -1 f) (* -1 e)  0) (position (* -1 h) (* -1 g)  0) "SiO2" "SpacerL")
;(sdegeo:fillet-2d (find-vertex-id (position (* -1 m) (* -1 l) 0.0)) 0.03)

;(sdegeo:create-rectangle (position h (* -1 e)  0) (position f (* -1 g) 0) "SiO2" "SpacerR")
;(sdegeo:fillet-2d (find-vertex-id (position (* 1 m) (* -1 l) 0.0)) 0.03)

;(sdegeo:create-rectangle (position (* -1 f) (* -1 e)  0) (position (* 1 f) (* -1 l)  0) "Aluminum" "Gate")
```

```
;********************************** Doping ****************************************

(sedr:define-constant-profile "DDP1" DD1 DopSD1)
(sedr:define-constant-profile-region "driandopprofile1" "DDP1"  "Drain")


(sedr:define-constant-profile "SDP1" SD1 DopSD1)
(sedr:define-constant-profile-region "sourcedopprofile1" "SDP1" "Source")


(sedr:define-constant-profile "CDP" CD  DopSub)
(sedr:define-constant-profile-region "chdopprofile" "CDP" "Channel")



(sedr:define-constant-profile "SubDP" CD  DopSubstrate)
(sedr:define-constant-profile-region "subdopprofile" "SubDP" "Substrate")


;********************************** Contacts ****************************************

(sdegeo:define-contact-set "gate"      4.0 (color:rgb 1.0 0.0 0.0 ) "##")
(sdegeo:define-contact-set "drain"     4.0 (color:rgb 0.0 1.0 0.0 ) "||")
(sdegeo:define-contact-set "source"    4.0 (color:rgb 0.0 0.0 1.0 ) "||")
(sdegeo:define-contact-set "substrate" 4.0 (color:rgb 1.0 0.0 0.0 ) "<><>")


(sdegeo:set-current-contact-set "source")
(sdegeo:set-contact-edges (find-edge-id (position (* -1 i) (* -1 d)  0)) "source")

(sdegeo:set-current-contact-set "drain")
(sdegeo:set-contact-edges (find-edge-id (position i (* -1 d)  0)) "drain")

(sdegeo:set-current-contact-set "gate")
(sdegeo:set-contact-edges (find-edge-id (position 0.0 (* -1 e) 0)) "gate")

(sdegeo:set-current-contact-set "substrate")
(sdegeo:set-contact-edges (find-edge-id (position 0.0 0.0 0)) "substrate")
```

```
;********************************* Meshing ******************************************************

(define mesh4 (* 0.1 ))

(sdedr:define-refeval-window "SDMesh" "Rectangle"  (position (* -1 a) (* -1 e) 0 )  (position a 0 0) )
(sdedr:define-refinement-size "SDMeshDef" mesh4 mesh4 mesh4 mesh4 mesh4 mesh4 )
(sdedr:define-refinement-placement "Ref3" "SDMeshDef" "SDMesh" )

(sdedr:define-refeval-window "NWMesh.Silicon" "Rectangle" (position (* -1 z) (+ (* -1 c) 0.01) 0) (position z (* -1 e)
(sdedr:define-refinement-size "NWMeshDef.Silicon" 0.01 0.01 0 0.001 0.001 0)
(sdedr:define-refinement-placement "Ref5" "NWMeshDef.Silicon" "NWMesh.Silicon")
(sdedr:define-refinement-function "NWMeshDef.Silicon" "MaxLenInt" "Silicon" "GaP" 0.001 1.5 "DoubleSide")
(sdedr:define-refinement-function "NWMeshDef.Silicon" "MaxLenInt" "Silicon" "Oxide" 0.001 1.5 "DoubleSide")

;(sdedr:define-refinement-function "SDMeshDef" "DopingConcentration" "MaxTransDiff" 1)


(sde:build-mesh "snmesh" "" "n4_soifet_msh")
```

A.2 SDEVICE code for TiO_2 S/D Z-RAM Cell:

```
########### All the input and output files ##########

File {
        Grid=       "n@node|-2@_soifet_msh.tdr"
        Current=    "SOINMOS_@node@.plt"
        Plot=       "SOINMOS_@node@_IdVg.tdr"
        Output=     "SOINMOS_@node@_IdVg.log"
        Parameter="@parameter@"

      }

############### Electrode Section ###############

Electrode {

    { name="gate" Voltage=0.0 Workfunction=4.4}
    { name="drain" Voltage=0.0 }
    { name="source" Voltage=0.0 }
    { name="substrate" Voltage=0 }

       }

############### Global Physics Section#################
Physics {
    AreaFactor=1.0
   Hydrodynamic( )
   Temperature= 300
   EffectiveIntrinsicDensity( OldSlotboom )
        }
######### Silicon Physics #########

Physics(Material="Silicon") {
   Mobility(
      PhuMob
      eHighFieldsaturation(CarrierTempDrive )
      hHighFieldsaturation( GradQuasiFermi )
      Enormal
   )
```

```
    Recombination(
        SRH( DopingDep TempDep)
        Auger(WithGeneration)
        Band2Band (E2)
        Avalanche(CarrierTempDrive)
    )
}

########### TiO2 Physics #########
Physics(Material=TiO2) {
    Mobility(
        DopingDependence
        eHighFieldsaturation(CarrierTempDrive )
        hHighFieldsaturation( GradQuasiFermi )
          )
    EffectiveIntrinsicDensity( NoBandGapNarrowing)
    Recombination(
        Radiative
        Avalanche(CarrierTempDrive)
                    )
          }
############ interface Physics #####
Physics (MaterialInterface="Silicon/TiO2") {
Thermionic
HeteroInterface
Traps((Donor Gaussian fromCondBand Conc=@DT@ EnergyMid=0.1
EnergySig=0.1 eXsection=1e-12 hXsection=1e-12)  )
Recombination(
          Radiative
          SRH
          eBarrierTunneling(Band2band TwoBand)
          hBarrierTunneling(Band2band TwoBand)
                )
  }

```

```
74 ############## Plot Section #################
75
76 Plot {
77 *--Density and Currents, etc
78    eDensity hDensity
79    TotalCurrent/Vector eCurrent/Vector hCurrent/Vector
80    eMobility hMobility
81    eVelocity hVelocity
82    eQuasiFermi hQuasiFermi
83
84 *--Temperature
85    eTemperature * Temperature hTemperature
86    hTemperature
87    Temperature
88
89 *--Fields and charges
90    ElectricField/Vector Potential SpaceCharge
91
92 *--Doping Profiles
93    Doping DonorConcentration AcceptorConcentration
94
95 *--Generation/Recombination
96    SRH Band2Band * Auger
97    AvalancheGeneration eAvalancheGeneration hAvalancheGeneration
98
99 *--Driving forces
100   eGradQuasiFermi/Vector hGradQuasiFermi/Vector
101   eEparallel hEparallel eENormal hENormal
102
103 *--Band structure/Composition
104   BandGap
105   BandGapNarrowing
106   Affinity
107   ConductionBand ValenceBand
108
109 *--Traps
110   * eTrappedCharge  hTrappedCharge
111   * eGapStatesRecombination hGapStatesRecombination
112 }
```

```
114 ################ Math Section#####################
115
116 Math {
117    Extrapolate
118    Avalderivatives
119    RelErrControl
120    Digits=5
121    ErRef(electron)=1.e10
122    ErRef(hole)=1.e10
123    Notdamped=50
124    Iterations=20
125    DirectCurrent
126 #    BreakCriteria{ Current(Contact="drain" AbsVal=1.443e-3) }
127 }
128
129 ############ Interface Math ##############
130
131 Math (MaterialInterface="Silicon/TiO2") {
132
133 Nonlocal(Length=15e-7)
134 Digits(NonLocal)=4
135 EnergyResolution(NonLocal)=0.0001
136
137 }
138
139 ############## Solve Section ####################
140 Solve {
141
142         Coupled(Iterations= 100 LineSearchDamping= 1e-4) { Poisson Electron }
143         Coupled { Poisson Electron Hole eTemperature}
144         Coupled {Poisson  Electron Hole eTemperature}
145
146          save(FilePrefix="initial")
147
148 ### Initial Boundary Condition************
149
150         QuasiStationary
151              ( InitialStep=1e-3 Increment=1.3 Maxstep=1e-2 MinStep=1e-4
152               Goal { name="drain" voltage=0.05 }
153               Goal { name="gate" voltage=-1 }
154
```

```
155             )
156             { Coupled { Poisson  Electron Hole  eTemperature } }
157         NewCurrentPrefix="Idvg=2_T=300_Lg=@Lg@_"
158
159     #************ Sweeping *************
160
161         QuasiStationary
162             ( InitialStep=1e-3 Increment=1.3 Maxstep=1e-2 MinStep=1e-4
163               Goal { name="gate" voltage=2 }
164             )
165             { Coupled { Poisson  Electron Hole  eTemperature } }
166
167         Plot (FilePrefix= "nr@node@_Vg" Time=( 44e-9; 1e-7; 1e-6 ) NoOverwrite )
168
169         }
170
171
172
```

A.3 Structure Editor code for bulk FinFET with buried n^+ layer:

```
; Reinitializing SDE
(sde:clear)

; Selecting default Boolean expression
(sdegeo:set-default-boolean "BAB")

;-----------------------------------------------------
; Setting parameters
(define Ymin -0.15)   ; [um] Substrate height
(define Ybox -0.05)   ; [um] BOX height
(define Ybar -0.025)   ; [um] BOX height
;(define Ybar -0.03)   ; [um] BOX height
(define Ymax  0.25)   ; [um] Total BEOL length

(define FPitch 0.1) ;[um] Fin Pitch


(define Lch  0.1)   ; [um] Channel Length
(define Lsd  0.15)   ; [um] Source Drain Legth
(define Lext  0.05)   ; [um] Channel Length
(define Lsp  0.02)   ; [um] Spacer Legth
;(define Lsp  0.005)   ; [um] Spacer Legth
(define Hsp  0.01)   ; [um] Spacer Height

;(define Wfin  0.03)   ; [um] Fin Width
(define Wfin  0.035)   ; [um] Fin Width
(define Hfin  0.1)   ; [um] Fin Height

;(define Tox  0.0006)   ; [um] Oxide thickness
(define Tox  0.002)   ; [um] Oxide thickness
(define Thk  0.00203)   ; [um] HK thickness

(define M (+ (/ Wfin 2) 0.004))

(define RviaT  0.01)   ; [um] Via Radius @ Top
(define RviaB  0.0075)   ; [um] Via Radius @ Bottom
(define Rcorner  0.002)   ; [um] Corner Radius
(define Rsp  0.003)   ; [um] Corner Radius

```

```
;-----------------------------------------------------------------------
; Derived quantities
(define Zmin  0.0)   ; [um]
(define Zmax  (+ (+ Lsd Lch) Lsd))   ; [um] Width 2
(define Xmin (* -1 (/ FPitch 2)))   ; [um] Lateral extend divided by 2
(define Xmax  (/ FPitch 2))   ; [um] Lateral extend divided by 2
(define GPitch (+ Lch (* Lsd 2)))

;-----------------------------------------------------------------------


;********** Creating n+ region*********

(sdegeo:create-cuboid
  (position (* -1 (/ Wfin 2)) Ybox Zmin)
  (position (/ Wfin 2) Ybar Zmax) "Silicon" "R.Well")


;********** Creating Fin channel region**********

(sdegeo:create-cuboid
  (position (* -1 (/ Wfin 2)) Ybar Zmin)
  (position (/ Wfin 2) Hfin Zmax) "Silicon" "R.Fin")
(sdegeo:fillet (list
                (car (find-edge-id (position (/ Wfin 2) Hfin (+ Lsd (/ Lch 2)))))
                (car (find-edge-id (position (* -1 (/ Wfin 2)) Hfin (+ Lsd (/ Lch 2)))))) Rcorner



;********* Creating BOX region*********

(sdegeo:create-cuboid
  (position Xmin Ybox Zmin )
  (position Xmax 0.0  Zmax ) "Oxide" "R.BOX" )

;********* Creating substrate region******

(sdegeo:create-cuboid
  (position Xmin Ymin Zmin )
  (position Xmax 0.0  Zmax ) "Silicon" "R.Substrate" )
```

```
;************ Creating Fin (channel+SiO2)region********

(sdegeo:create-cuboid
  (position (* -1 (+ (/ Wfin 2) Tox)) 0 Lsd)
  (position (+ (/ Wfin 2) Tox) (+ Hfin Tox) (+ Lsd Lch)) "Oxide" "R.Fin1")
(sdegeo:fillet (list
                (car (find-edge-id (position (+ (/ Wfin 2) Tox) (+ Hfin Tox) (+ Lsd (/ Lch 2)))))
                (car (find-edge-id (position (* -1 (+ (/ Wfin 2) Tox)) (+ Hfin Tox) (+ Lsd (/ Lch 2)))))) (+ Tox Rcorner))




;********** Creating spacer1 region***********

(sdegeo:create-cuboid
  (position (* -1 (+ (/ Wfin 2) Hsp)) 0 (- Lsd Lsp))
  (position (+ (/ Wfin 2) Hsp) (+ Hfin Hsp) Lsd) "Nitride" "R.Spacer1")
(sdegeo:fillet (list
                (car (find-vertex-id (position (* -1 (+ (/ Wfin 2) Hsp)) (+ Hfin Hsp) (- Lsd Lsp))))
                (car (find-vertex-id (position (+ (/ Wfin 2) Hsp) (+ Hfin Hsp) (- Lsd Lsp))))) Rsp)

;********** Creating spacer2 region***********

(sdegeo:create-cuboid
  (position (* -1 (+ (/ Wfin 2) Hsp)) 0 (+ Lsd Lch))
  (position (+ (/ Wfin 2) Hsp) (+ Hfin Hsp) (+ (+ Lsd Lch) Lsp)) "Nitride" "R.Spacer2")
(sdegeo:fillet (list
                (car (find-vertex-id (position (* -1 (+ (/ Wfin 2) Hsp)) (+ Hfin Hsp) (+ Lsd (+ Lch Lsp)))))
                (car (find-vertex-id (position (+ (/ Wfin 2) Hsp) (+ Hfin Hsp) (+ Lsd (+ Lch Lsp)))))) Rsp)

;******************** Creating Gate region **********************

(sdegeo:create-cuboid
  (position (+ Xmin 0.01) 0 Lsd)
  (position (- Xmax 0.01) (+ Hfin (* 2 Hsp)) (+ Lsd Lch)) "Tungsten" "R.Gate")

```

```
; ------------ n+ Region---------
(sdedr:define-constant-profile "Const.Well" "ArsenicActiveConcentration" 1e20 )
(sdedr:define-constant-profile-region "PlaceCD.Well" "Const.Well" "R.Well" )

; ------------------ Drain extension implant----------
; -- base line definition----

(sdedr:define-refinement-window "BaseLine.DrainExt" "Rectangle"
(position  (* -1 (/ Wfin 2)) 0.0 (+ Lch Lsd) )
(position  (/ Wfin 2) 0.0 (+ Lch (* Lsd 2)) ) )

; ---- implant definition-----

(sdedr:define-gaussian-profile "Gauss.DrainExt"
 "ArsenicActiveConcentration"
 "PeakPos" Hfin "PeakVal" 1e19
 "ValueAtDepth" 5e17 "Depth" 0.08 "Gauss" "Factor" 0.8)

; --------- implant placement--------

(sdedr:define-analytical-profile-placement "PlaceAP.DrainExt"
 "Gauss.DrainExt" "BaseLine.DrainExt" "Positive" "NoReplace" "Eval")

; ---------- Drain implant---------
; -- base line definition---

(sdedr:define-refinement-window "BaseLine.Drain" "Rectangle"
  (position  (* -1 (/ Wfin 2)) 0.0 (- (+ Lch (* Lsd 2)) 0.095) )
  (position  (/ Wfin 2) 0.0 (+ Lch (* Lsd 2)) ) )

; --------- implant definition--------

(sdedr:define-gaussian-profile "Gauss.Drain"
 "ArsenicActiveConcentration"
 "PeakPos" Hfin  "PeakVal" 1e21
 "ValueAtDepth" 5e18 "Depth" 0.0 "Gauss"  "Factor" 0.5)
```

```
198 ; -------- implant placement---------
199
200 (sdedr:define-analytical-profile-placement "PlaceAP.Drain"
201  "Gauss.Drain" "BaseLine.Drain" "Positive" "NoReplace" "Eval")
202 ; ----------******* Source extension implant******----------
203 ; -- base line definition----
204
205 (sdedr:define-refinement-window "BaseLine.SourceExt" "Rectangle"
206 (position (* -1 (/ Wfin 2)) 0.0 0.0) (position (/ Wfin 2) 0.0 Lsd ) )
207
208 ; -------- implant definition-----
209
210 (sdedr:define-gaussian-profile "Gauss.SourceExt"
211  "ArsenicActiveConcentration"
212  "PeakPos" Hfin "PeakVal" 1e19
213  "ValueAtDepth" 5e17 "Depth" 0.08 "Gauss" "Factor" 0.8)
214
215 ; ------- implant placement------
216
217 (sdedr:define-analytical-profile-placement "PlaceAP.SourceExt"
218  "Gauss.SourceExt" "BaseLine.SourceExt" "Positive" "NoReplace" "Eval")
219
220 ; --------- Source implant--------
221 ; -- base line definition-----
222
223 (sdedr:define-refinement-window "BaseLine.Source" "Rectangle"
224   (position (* -1 (/ Wfin 2)) 0.0 0.0)  (position (/ Wfin 2) 0.0 0.095 ) )
225
226 ; -------- implant definition-------
227
228 (sdedr:define-gaussian-profile "Gauss.Source"
229  "ArsenicActiveConcentration"
230  "PeakPos" Hfin  "PeakVal" 1e21
231  "ValueAtDepth" 5e18 "Depth" 0.0 "Gauss"  "Factor" 0.5)
232
233 ; ------- implant placement--------
234
235 (sdedr:define-analytical-profile-placement "PlaceAP.Source"
236  "Gauss.Source" "BaseLine.Source" "Positive" "NoReplace" "Eval")
237
```

```
;-----------------------------------------------------------------------------
;-- Specify mesh refinements -------------------------------------------------
;-----------------------------------------------------------------------------

;************ Defining the global refinement window *************

(sdedr:define-refinement-window "RefWin.all" "Cuboid" (position Xmin Ymin Zmin) (position Xmax Ymax Zmax))
(sdedr:define-refinement-size "RefDef.all"
    0.015 0.025 0.03
    0.015 0.025 0.03)
(sdedr:define-refinement-placement "PlaceRF.all" "RefDef.all" "RefWin.all")
(sdedr:define-refinement-function "RefDef.all" "DopingConcentration" "MaxTransDiff" 1)


;************ Interface Meshing *************



(sdedr:define-refinement-window "RefWin.all" "Cuboid"
(position (* M -1) Ybar (- Lsd 0.07)) (position M (+ Hfin 0.005) (- (+ Lch (* Lsd 2)) 0.07)))
(sdedr:define-refinement-size "RefDef.all"
    0.015 0.025 0.025
    0.015 0.025 0.025)
(sdedr:define-refinement-placement "PlaceRF.all" "RefDef.all" "RefWin.all")
(sdedr:define-refinement-function "RefDef.all" "DopingConcentration" "MaxTransDiff" 1)
(sdedr:define-refinement-function "RefDef.all" "MaxLenInt" "Silicon" "Oxide" 0.0015 1.8 "DoubleSide")
(sdedr:define-refinement-function "RefDef.all" "MaxLenInt" "Silicon" "Silicon" 0.003 2 "DoubleSide")

;(sdedr:define-refinement-function "RefDef.all" "MaxLenInt" "Oxide" "Metal" 0.001 2 "DoubleSide")
;(sdedr:define-refinement-function "RefDef.all" "MaxLenInt" "Oxide" "Nitride" 0.0008 1.5 "DoubleSide")
;(sdedr:define-refinement-function "RefDef.all" "MaxLenInt" "Silicon" "Nitride" 0.0008 1.5 "DoubleSide")

(sde:build-mesh "snmesh" "-AI" "n@node@_FinFET_msh")
```

A.4 SDEVICE code for bulk FinFET with buried n^+ layer:

```
#************* All the input and output file names *****

File {
      Grid=       "n@node|-1@_FinFET_msh.tdr"
      Current=    "FinFET_@node@.plt"
      Plot=       "FinFET_@node@_IdVgsub.tdr"
      Output=     "FinFET_@node@_IdVg.log"
      Parameter="@parameter@"

      }

#************* Electrodes *************

Electrode {

        { name="gate" Voltage=0.0 Workfunction=4.6}
        { name="drain" Voltage=0.0  }
        { name="source" Voltage=0.0 }
        { name="substrate" Voltage=0 }

       }

#********************* Physical models ****************

Physics {
    AreaFactor=1.0
   Hydrodynamic( )
   Temperature= 300
   EffectiveIntrinsicDensity( OldSlotboom )
   Recombination( SRH(DopingDep TempDependence) )

}

```

```
Physics(Material="Silicon") {
   Mobility(
      PhuMob
      HighFieldsaturation
      Enormal
   )
   Recombination(
      Auger(WithGeneration)
      Band2Band (E2)
      Avalanche(CarrierTempDrive)
   )
}

#***************** The parameters that you want to visualize *******

Plot {
*--Density and Currents, etc
   eDensity hDensity
   TotalCurrent/Vector eCurrent/Vector hCurrent/Vector
   eMobility hMobility
   eVelocity hVelocity
   eQuasiFermi hQuasiFermi

*--Temperature
   eTemperature * Temperature hTemperature
   hTemperature
   Temperature

*--Fields and charges
   ElectricField/Vector Potential SpaceCharge

*--Doping Profiles
   Doping DonorConcentration AcceptorConcentration

*--Generation/Recombination
   SRH Band2Band * Auger
   AvalancheGeneration eAvalancheGeneration hAvalancheGeneration

*--Driving forces
   eGradQuasiFermi/Vector hGradQuasiFermi/Vector
   eEparallel hEparallel eENormal hENormal
```

```
76 *--Band structure/Composition
77    BandGap
78    BandGapNarrowing
79    Affinity
80    ConductionBand ValenceBand
81
82 *--Traps
83    * eTrappedCharge  hTrappedCharge
84    * eGapStatesRecombination hGapStatesRecombination
85 }
86
87 #***************** Math Section ****************
88
89 Math {
90    Extrapolate
91    Avalderivatives
92    RelErrControl
93    Digits=5
94    ErRef(electron)=1.e10
95    ErRef(hole)=1.e10
96    Notdamped=50
97    Iterations=20
98    DirectCurrent
99 #   BreakCriteria{ Current(Contact="drain" AbsVal=1.443e-3) }
100 }
101
102 #***************** Solve Section ****************
103
104 Solve {
105         # initial gate voltage Vgs=0.0V
106     Poisson
107         Coupled(Iterations= 100 LineSearchDamping= 1e-4) { Poisson Electron }
108         Coupled { Poisson Electron Hole }
109         Coupled {Poisson  Electron Hole  }
110          save(FilePrefix="initial")
111
```

```
#   drain voltage Vds=0.05V
      QuasiStationary
           ( InitialStep=1e-3 Increment=1.3 Maxstep=2e-1 MinStep=1e-4
            Goal { name="drain" voltage=0 }
            Goal { name="gate" voltage=@Vg@ }
           )
     { Coupled { Poisson  eTemperature Electron Hole  } }
       save(FilePrefix="vg_@Vg@")


 # third curve
   load(FilePrefix="vg_@Vg@")
   NewCurrentPrefix="idvd_vg_@Vg@_withBTBTwithII1af@AvF@LT6_H100nmW35nmLDD1e19_"
      QuasiStationary
          ( InitialStep=1e-3 Increment=1.3 Maxstep=2e-1 MinStep=1e-4
            Goal { name="drain" voltage=2.5 }
          )
    {Coupled { Poisson  eTemperature Electron Hole  }}
      Plot (FilePrefix="idvd_vg_@Vg@_withBTBTwithII1af@AvF@LT6_H100nmW35nmLDD1e19_" Time=(0;0.
                     Coupled { Poisson eTemperature Electron Hole }


     }
```

Appendix B

Reset Circuit For Bulk FinFET Based Neuron

Bulk FinFET together with the reset circuit will act as an electronic LIF neuron. Reset circuit controls the drain bias. In biology, more the number of signals from the pre-synaptic neurons, more quickly a spike is generated. Similarly, in the proposed device, more the number of pre-synaptic signals, I_1, I_2, $\ldots I_N$ coming through the synapses with synaptic weights W_1, W_2, $\ldots W_N$, more will be the input bias (V_{in}(t)) as shown in Fig. B.1 and faster will be the spike generation.

Integrate and reset are the two circuit blocks of a LIF neuron. Integrate circuit block is replaced by the nano scale devices with an overall area improvement of at least $10X$ compared to CMOS circuit-based neurons. The reset block is a circuit design challenge from an energy point of view. Sub-threshold design and operation is one of the ways to improve the reset circuit which has been discussed extensively in reference [154]. So, here, we compare only the area and energy efficiency of the nano-scale devices replacing the leaky integration function of neuron (i.e. PCM, PD-SOI-MOSFET, RRAM) with the proposed bulk FinFET integrator and it is seen that, the proposed bulk FinFET integrator is more energy efficient with the lowest energy/spike of 6.3×10^{15} J which transfers the bottleneck from low energy integrator design to low energy reset circuit design. The main interest of this paper is not to design an efficient reset circuit but to replace the area and energy inefficient CMOS circuit based integrate block with a single nano scale device for both area and energy reduction. The purpose of the reset circuit is to control the drain bias. As soon as the drain current reaches the threshold, the drain bias is reset to -1.5 V for 8 ns and then 0 V for 500 ns by the reset

circuit. The width of the drain pulses is determined by the excess charge dynamics of the proposed device. Even though, the main focus of this paper is to design an energy and area efficient integrate block with a single nano scale device, for a better understanding of the proposed neuron we have designed a reset circuit as shown below.

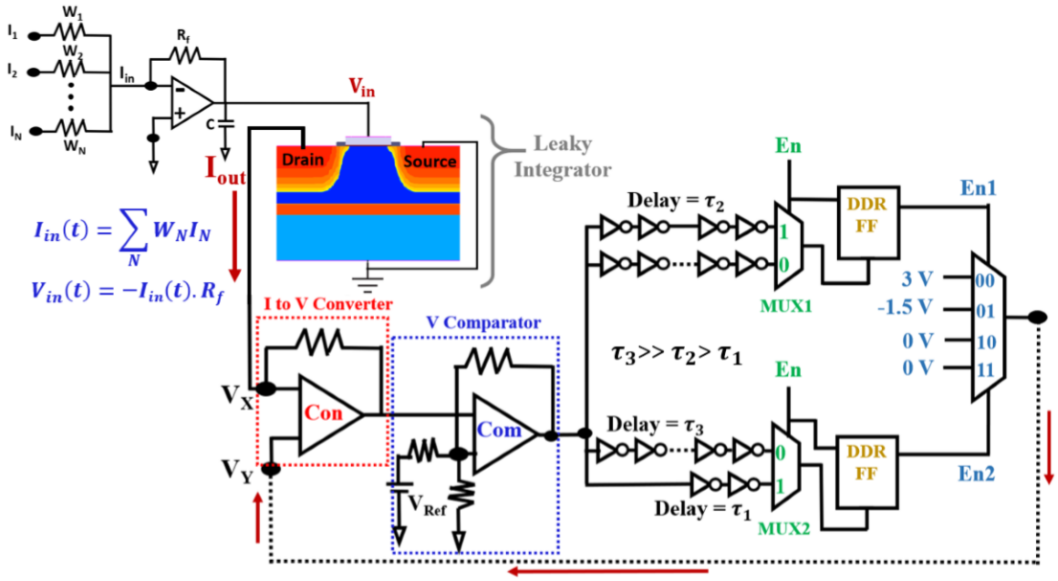

FIGURE B.1: Integrate block (bulk FinFET with n^+ buried layer) together with the reset circuit makes a LIF neuron.

Explanation of the reset circuit: Let's consider initially, En_1 and En_2 are "0" and "0". So, 3 V goes to the output of the $4:1$ MUX and fed to V_Y which is the non-inverting terminal of the I to V converter. The output of the $4:1$ MUX is not directly fed into the drain terminal of the integrate block because in that case, the output current will be splitting into two components and the functionality of the integrate block may be lost. As op-amp is a differential amplifier, $V_Y = V_X$. So, the drain voltage will be the same as the $4:1$ MUX output. Drain current or the output current is coming out from the integrate block and enters into the I to V converter. The output of the I to V converter is then fed into the input of a voltage comparator to compare with the threshold voltage. If it is greater than threshold voltage, the output of the voltage comparator will be "1" or V_{dd} and if less than threshold voltage, then the output of the voltage comparator will be "0". Enable (En) signal of the two $2:1$ MUX is the same and are connected to the voltage comparator output. So, when output of the voltage comparator becomes "1"/"0", the En also becomes "1"/"0". As $V_Y = V_Y = 3$ V, due to positive feedback process, I_{OUT} increases which increases the output voltage of the I to V converter. After some time (t_{Int}), it crosses the threshold voltage and the comparator output becomes "1" or V_{dd}. Hence En becomes "1". So, "1" comes from

τ_1 delay path and makes $En_2 =$ "1". Now, En_1 and En_2 are "0" and "1" and -1.5 V goes to the output of the $4:1$ MUX and it is fed to the non-inverting terminal (V_Y) of the I to V converter. In this way the drain pulse is generated by the reset circuit.

Better reset circuit can be designed using subthreshold design technique as mentioned above. The above circuit is shown for a better understanding of the functionality of the proposed neuron.

Bibliography

[1] W.H. Brattain and B. John. Three-electrode circuit element utilizing semiconductive materials. *US Patent*, 2(035):524, October 1950.

[2] L.J. Edgar. Three-electrode circuit element utilizing semiconductive materials. *US Patent*, 1(347):745, January 1930.

[3] J. S. Kilby. Miniature semiconductor integrated circuit. *US Patent*, 3(581):115, December 1963.

[4] R. H. Dennard. Field-Effect Transistor Memory. *US Patent*, 3(286):387, June 1968.

[5] J. von Neumann. First draft of a report on the EDVAC. *IEEE Annals of the History of Computing*, 15(4):27–75, 1993.

[6] M. Naylor and C. Runciman. The reduceron: Widening the von Neumann bottleneck for graph reduction using an FPGA. 5083:129–146, 2008.

[7] Chih-Yuan Lu. Future Prospects of NAND Flash Memory Technology—The Evolution from Floating Gate to Charge Trapping to 3D Stacking. *Journal of Nanoscience and Nanotechnology*, 12(10):7604–7618, October 2012.

[8] Cai Qingchao, Rajesh Vellore Arumugam, Quanqing Xu, and Bingsheng He. Understanding the Behavior of Solid State Disk. In *Proceedings of the Asia Pacific Symposium on Intelligent and Evolutionary Systems*, volume 1, pages 341–355, 2015.

[9] Alavi Seyed Rashid, Rahmati Mehdi, and Ziaei Rad Saeed. Optimization of passive control performance for different hard disk drives subjected to shock excitation. *Journal of Central South University*, 24(4):891–899, January 2017.

[10] B. S. Haran, A. Kumar, L. Adam, J. Chang, V. Basker, S. Kanakasabapathy, D. Horak, S. Fan, J. Chen, J. Faltermeier, S. Seo, M. Burkhardt, S. Burns, S. Halle, S. Holmes, R. Johnson, E. McLellan, T. M. Levin, Y. Zhu, J. Kuss, A. Ebert, J. Cummings, D. Canaperi, S. Paparao, J. Arnold, T. Sparks, C. S. Koay, T. Kanarsky, S. Schmitz, K. Petrillo, R. H. Kim, J. Demarest, L. F. Edge, H. Jagannathan, M. Smalley, N. Berliner, K. Cheng, D. LaTulipe, C. Koburger, S. Mehta, M. Raymond, M. Colburn, T. Spooner, V. Paruchuri, W. Haensch, D. McHerron, and B. Doris. 22 nm technology compatible fully functional 0.1 μm^2 6T-SRAM cell. In *Technical Digest of the IEEE International Electron Device Meeting*, pages 1–4, December 2008.

[11] Apollos Ezeogu. Performance Analysis of 6T and 9T SRAM. *CoRR*, May 2019.

[12] J. A. Mandelman, R. H. Dennard, G. B. Bronner, J. K. DeBrosse, R. Divakaruni, Y. Li, and C. J. Radens. Challenges and future directions for the scaling of dynamic random-access memory (DRAM). *IBM Journal of Research and Development*, 46(2.3):187–212, March 2002.

[13] S. S. Iyer, J. E. Barth, P. C. Parries, J. P. Norum, J. P. Rice, L. R. Logan, and D. Hoyniak. Embedded DRAM: Technology platform for the Blue Gene/L chip. *IBM Journal of Research and Development*, 49(2.3):333–350, 2005.

[14] RESEARCH BULLETIN. Memory ICs to Account for 53% of Total 2018 Semi Capex. *IC INSIGHTS, INC*, pages 1–3, August 2018.

[15] Seong Keun Kim and Mihaela Popovici. Future of dynamic random-access memory as main memory. *MRS Bulletin*, 43(5):334–339, 2018.

[16] K. Roy, S. Mukhopadhyay, and H. Mahmoodi-Meimand. Leakage current mechanisms and leakage reduction techniques in deep-submicrometer CMOS circuits. *Proceedings of the IEEE*, 91(2):305–327, February 2003.

[17] International technology roadmap for semiconductors. 2011.

[18] A. Nitayama, Y. Kohyama, and K. Hieda. Future directions for DRAM memory cell technology. In *Technical Digest of the IEEE International Electron Device Meeting*, pages 355–358, December 1998.

[19] Kanghoon Yim, Youn Yong, Joohee Lee, Kyuhyun Lee, Ho-Hyun Nahm, Jiho Yoo, Chanhee Lee, Cheol Seong Hwang, and Seungwu Han. Novel high-

dielectrics for next-generation electronic devices screened by automated *ab initio* calculations. *NPG Asia Materials*, 7(6):e190–e190, June 2015.

[20] L. Nesbit, J. Alsmeier, B. Chen, J. DeBrosse, P. Faheyk, M. Gall, J. Gambino, S. Gernhard, H. Ishiuchi, R. Kleinhenz, J. Mandelman, T. Mii, M. Morikado, A. Nitayama, S. Parke, H. Wong, and G. Bronner. A 0.6 μm^2 256 Mb trench DRAM cell with self-aligned buried strap (BEST). In *Technical Digest of the IEEE International Electron Device Meeting*, pages 627–630, December 1993.

[21] G. Bronner, H. Aochi, M. Gall, J. Gambino, S. Gernhardt, E. Hammerl, H. Ho, J. Iba, H. Ishiuchi, M. Jaso, R. Kleinhenz, T. Mii, M. Narita, L. Nesbit, W. Neumueller, A. Nitayama, T. Ohiwa, S. Parke, J. Ryan, T. Sato, H. Takato, and S. Yoshikawa. A fully planarized 0.25 μm CMOS technology for 256 Mbit DRAM and beyond. In *Proceedings of the Symposium on VLSI Technology*, pages 15–16, June 1995.

[22] K. P. Muller, B. Flietner, C. L. Hwang, R. L. Kleinhenz, T. Nakao, R. Ranade, Y. Tsunashima, and T. Mii. Trench storage node technology for gigabit DRAM generations. In *Technical Digest of the IEEE International Electron Device Meeting*, pages 507–510, December 1996.

[23] Erwin Hammerl and Herb Lei Ho. Nitride cap formation in a DRAM trench capacitor. *US Patent*, (5717628):1–5, February 1998.

[24] K. Itabashi, S. Tsuboi, H. Nakamura, K. Hashimoto, W. Futoh, K. Fukuda, I. Hanyu, S. Asai, T. Chijimatsu, E. Kawamura, T. Yao, H. Takagi, Y. Ohta, T. Karasawa, H. Iio, M. Onods, F. Inoue, H. Nomura, Y. Satoh, M. Higashimoto, M. Matsumiya, T. Miyabo, T. Ikeda, T. Yamazaki, M. Miyajima, K. Watanabe, S. Kawamura, and M. Taguchi. Fully planarized stacked capacitor cell with deep and high aspect ratio contact hole for gigs-bit DRAM. In *Proceedings of the Symposium on VLSI Technology*, pages 21–22, June 1997.

[25] M. Tsukamoto, H. Kuroda, and Y. Okamoto. 0.25 μm w-polycide dual gate and buried metal on diffusion layer (BMD) technology for DRAM-embedded logic devices. In *Proceedings of the Symposium on VLSI Technology*, pages 23–24, June 1997.

[26] T. Lim and Y. Kim. Effect of band-to-band tunnelling leakage on 28 nm MOSFET design. *Electronics Letters*, 44(2):157–158, January 2008.

[27] A. Rjoub, N. R. Al-Taradeh, and M. F. Al-Mistarihi. Accurate subthreshold leakage model for nanoscale MOSFET transistor. In *Proceedings of the International Conference on Electronics, Circuits, and Systems*, pages 711–714, December 2013.

[28] Dongwoo Lee, D. Blaauw, and D. Sylvester. Gate oxide leakage current analysis and reduction for VLSI circuits. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 12(2):155–166, February 2004.

[29] V. Nathan and N. C. Das. Gate-induced drain leakage current in MOS devices. *IEEE Transactions on Electron Devices*, 40(10):1888–1890, October 1993.

[30] T. Y. Chan, J. Chen, P. K. Ko, and C. Hu. The impact of gate-induced drain leakage current on MOSFET scaling. In *Technical Digest of the IEEE International Electron Device Meeting*, pages 718–721, December 1987.

[31] Y. Taur and T. H. Ning. Fundamentals of Modern VLSI Devices. *New York: Cambridge University Press*, pages 94–95, 1998.

[32] Y. Taur and T. H. Ning. Fundamentals of Modern VLSI Devices. *New York: Cambridge University Press*, pages 120–128, 1998.

[33] A. Keshavarzi, K. Roy, and F. Hawkins. Intrinsic leakage in low power deep submicron CMOS ICs. In *Proceedings of the International Test Conference*, pages 146–155, December 1997.

[34] Y. Taur and T. H. Ning. Fundamentals of Modern VLSI Devices. *New York: Cambridge University Press*, pages 143–144, 1998.

[35] S. Okhonin, M. Nagoga, J. M. Sallese, and P. Fazan. A capacitor-less 1T-DRAM cell. *IEEE Electron Device Letters*, 23(2):85–87, February 2002.

[36] S. Okhonin, M. Nagoga, E. Carman, R. Beffa, and E. Faraoni. New generation of Z-RAM. In *Technical Digest of the IEEE International Electron Device Meeting*, pages 925–928, December 2007.

[37] E. Yoshida and T. Tanaka. A capacitorless 1T-DRAM technology using gate-induced drain-leakage (GIDL) current for low-power and high-speed embedded memory. *IEEE Transactions on Electron Devices*, 53(4):692–697, April 2006.

[38] S. Ryu, J. Han, C. Kim, and Y. Choi. Investigation of Isolation-Dielectric Effects of PDSOI FINFET on Capacitorless 1T-DRAM. *IEEE Transactions on Electron Devices*, 56(12):3232–3235, December 2009.

[39] L.M. Almeida, K.R.A. Sasaki, M. Aoulaiche, E. Simoen, C. Claeys, and J.A. Martino. One Transistor Floating Body RAM Performances on UTBOX Devices Using the BJT Effect. *Journal of Integrated Circuits and Systems*, 7(1):113–120, 2012.

[40] T. Ohsawa, K. Fujita, T. Higashi, Y. Iwata, T. Kajiyama, Y. Asao, and K. Sunouchi. Memory design using one-transistor gain cell on SOI. In *Proceedings of the IEEE International Solid-State Circuits Conference*, volume 1, pages 152–455 vol.1, February 2002.

[41] C. Kuo, Tsu-Jae King, and Chenming Hu. A capacitorless double gate DRAM technology for sub-100-nm embedded and stand-alone memory applications. *IEEE Transactions on Electron Devices*, 50(12):2408–2416, December 2003.

[42] N. Collaert, M. Aoulaiche, M. Rakowski, B. De Wachter, K. Bourdelle, B. . Nguyen, F. Boedta, D. Delprat, and M. Jurczak. Analysis of sense margin and reliability of 1T-DRAM fabricated on thin-film UTBOX substrates. In *Proceedings of the IEEE International SOI Conference*, pages 1–2, October 2009.

[43] K. Kato, T. Wada, and K. Taniguchi. Analysis of kink characteristics in silicon-on-insulator MOSFET's using two-carrier modeling. *IEEE Transactions on Electron Devices*, 32(2):458–462, February 1985.

[44] C. D. Chen, M. Matloubian, R. Sundaresan, B. Y. Mao, C. C. Wei, and G. P. Pollack. Single-Transistor Latch in SOI MOSFETs. *IEEE Electron Device Letters*, 9(12):636–638, December 1988.

[45] M. Matloubian, C. D. Chen, B. Y. Mao, R. Sundaresan, and G. P. Pollack. Modeling of the subthreshold characteristics of SOI MOSFETs with floating body. *IEEE Transactions on Electron Devices*, 37(9):1985–1994, September 1990.

[46] A. Wei, M. J. Sherony, and D. A. Antoniadis. Transient behavior of the kink effect in partially-depleted SOI MOSFET's. *IEEE Electron Device Letters*, 16(11):494–496, November 1995.

[47] J. Luo, J. Chen, J. Zhuo, Q. Wu, Z. Chai, and X. Wang. Temperature dependence of hysteresis effect in partially depleted silicon-on-insulator MOSFETs. *IEEE Transactions on Device and Materials Reliability*, 12(1):63–67, March 2012.

[48] J. G. Fossum and Z. Lu. Anomalous floating-body effects in SOI MOSFETs: Low-voltage CMOS? In *Proceedings of the IEEE International SOI Conference*, pages 1–2, October 2011.

[49] T. Hamamoto, Y. Minami, T. Shino, N. Kusunoki, H. Nakajima, M. Morikado, T. Yamada, K. Inoh, A. Sakamoto, T. Higashi, K. Fujita, K. Hatsuda, T. Ohsawa, and A. Nitayama. A floating-body cell fully compatible with 90-nm CMOS technology node for a 128-Mb SOI DRAM and its scalability. *IEEE Transactions on Electron Devices*, 54(3):563–571, March 2007.

[50] T. Hamamoto and T. Ohsawa. Overview and future challenges of floating body RAM (FBRAM) technology for 32nm technology node and beyond. In *Proceedings of the European Solid-State Device Research Conference*, pages 25–29, September 2008.

[51] Q. Huang, R. Huang, Z. Zhan, Y. Qiu, W. Jiang, C. Wu, and Y. Wang. A novel Si tunnel FET with 36mV/dec subthreshold slope based on junction depleted-modulation through striped gate configuration. In *Technical Digest of the IEEE International Electron Device Meeting*, pages 8.5.1–8.5.4, December 2012.

[52] A. M. Walke, A. Vandooren, R. Rooyackers, D. Leonelli, A. Hikavyy, R. Loo, A. S. Verhulst, K. Kao, C. Huyghebaert, G. Groeseneken, V. R. Rao, K. K. Bhuwalka, M. M. Heyns, N. Collaert, and A. V. Thean. Fabrication and analysis of a $\mathrm{Si}/\mathrm{Si}_{0.55}\mathrm{Ge}_{0.45}$ heterojunction line tunnel FET. *IEEE Transactions on Electron Devices*, 61(3):707–715, March 2014.

[53] I. A. Fischer, A. S. M. Bakibillah, M. Golve, D. Hahnel, H. Isemann, A. Kottantharayil, M. Oehme, and J. Schulze. Silicon tunneling field-effect transistors with tunneling in line with the gate field. *IEEE Electron Device Letters*, 34(2):154–156, February 2013.

[54] Ki-Whan Song, Hoon Jeong, Jae-Wook Lee, Sung In Hong, Nam-Kyun Tak, Young-Tae Kim, Yong Lack Choi, Han Sung Joo, Sung Hwan Kim, Ho Ju Song, Yong Chul Oh, Woo-Seop Kim, Yeong-Taek Lee, Kyungseok Oh, and

Changhyun Kim. 55 nm capacitor-less 1T DRAM cell transistor with non-overlap structure. In *Technical Digest of the IEEE International Electron Device Meeting*, pages 1–4, December 2008.

[55] S. Okhonin, M. Nagoga, C. W. Lee, J. P. Colinge, A. Afzalian, R. Yan, N. Dehdashti Akhavan, W. Xiong, V. Sverdlov, S. Selberherr, and C. Mazure. Ultra-scaled Z-RAM cell. In *Proceedings of the IEEE International SOI Conference*, pages 157–158, October 2008.

[56] K. Shimomura, H. Shimano, N. Sakashita, F. Okuda, T. Oashi, Y. Yamaguchi, T. Eimori, M. Inuishi, K. Arimoto, S. Maegawa, Y. Inoue, S. Komori, and K. Kyuma. A 1-V 46-ns 16-Mb SOI-DRAM with body control technique. *IEEE Journal of Solid-State Circuits*, 32(11):1712–1720, November 1997.

[57] J. P. Colinge. Silicon-on-Insulator Technology: Materials to VLSI. *Springer Science and Business Media New York*, 3, 1997.

[58] Min Hee Cho. *Thin-Body SOI Capacitorless DRAM Cell Design Optimization and Scaling*. PhD thesis, University of California Berkeley, July 2012.

[59] Maryline Bawedin, S. Cristoloveanu, Denis Flandre, and Florin Udrea. Floating-body memory: Concepts, physics and challenges. *ECS Transactions*, 19(4):243–256, 2009.

[60] M. Aoulaiche, N. Collaert, R. Degraeve, Z. Lu, B. De Wachter, G. Groeseneken, M. Jurczak, and L. Altimime. BJT-Mode Endurance on a 1T-RAM bulk FinFET Device. *IEEE Electron Device Letters*, 31(12):1380–1382, 2010.

[61] J. Han, S. Ryu, S. Choi, and Y. Choi. Gate-Induced Drain-Leakage (GIDL) Programming Method for Soft-Programming-Free Operation in Unified RAM (URAM). *IEEE Electron Device Letters*, 30(2):189–191, 2009.

[62] E. Yoshida and T. Tanaka. A capacitorless 1T-DRAM technology using gate-induced drain-leakage (GIDL) current for low-power and high-speed embedded memory. *IEEE Transactions on Electron Devices*, 53(4):692–697, 2006.

[63] S. Okhonin, M. Nagoga, C. . Lee, J. . Colinge, A. Afzalian, R. Yan, N. Dehdashti Akhavan, W. Xiong, V. Sverdlov, S. Selberherr, and C. Mazure. Ultra-scaled Z-RAM cell. In *Proceedings of the IEEE International SOI Conference*, pages 157–158, 2008.

[64] A. Hubert, M. Bawedin, G. Guegan, S. Cristoloveanu, T. Ernst, and O. Faynot. Experimental comparison of programming mechanisms in 1T-DRAM cells with variable channel length. In *Proceedings of the European Solid State Device Research Conference*, pages 150–153, September 2010.

[65] R. Ranica, A. Villaret, C. Fenouillet-Beranger, P. Malinge, P. Mazoyer, P. Masson, D. Delille, C. Charbuillet, P. Candelier, and T. Skotnicki. A capacitorless DRAM cell on 75nm gate length, 16nm thin fully depleted SOI device for high density embedded memories. In *Technical Digest of the IEEE International Electron Device Meeting*, pages 277–280, December 2004.

[66] A. Pal, A. Nainani, and K. C. Saraswat. Addressing key challenges in 1T-DRAM: Retention time, scaling and variability - using a novel design with GaP source-drain. In *Proceedings of the International Conference on Simulation of Semiconductor Processes and Devices*, pages 376–379, September 2013.

[67] Poren Tang, Ru Huang, and Dake Wu. Performance improvement of capacitorless dynamic random access memory cell with band-gap engineered source and drain. *Japanese Journal of Applied Physics*, 49(4S):04DD02, April 2010.

[68] M. G. Ertosun and K. C. Saraswat. Investigation of capacitorless double-gate single-transistor DRAM: With and without quantum well. *IEEE Transactions on Electron Devices*, 57(3):608–613, March 2010.

[69] Dibyendu Chatterjee and Anil Kottantharayil. An improved 1T-DRAM cell using TiO_2 as the source and drain of an n-channel PD-SOI MOSFET. In *Proceedings of the IEEE Device Research Conference*, pages 1–2, June 2018.

[70] M. Bawedin, S. Cristoloveanu, and D. Flandre. A capacitorless 1T-DRAM on SOI based on dynamic coupling and double-gate operation. *IEEE Electron Device Letters*, 29(7):795–798, July 2008.

[71] N. Rodriguez, S. Cristoloveanu, and F. Gamiz. Novel capacitorless 1T-DRAM cell for 22-nm node compatible with bulk and SOI substrates. *IEEE Transactions on Electron Devices*, 58(8):2371–2377, May 2011.

[72] J. Wan, C. Le Royer, A. Zaslavsky, and S. Cristoloveanu. A Compact Capacitor-Less High-Speed DRAM Using Field Effect-Controlled Charge Regeneration. *IEEE Electron Device Letters*, 33(2):179–181, December 2012.

[73] Ivan K. Schuller Rick Stevens Michael Pechan. Neuromorphic Computing – From Materials Research to Systems Architecture Roundtable. https://doi.org/10.2172/1283147, October 2015.

[74] R. A. Nawrocki, R. M. Voyles, and S. E. Shaheen. A Mini Review of Neuromorphic Architectures and Implementations. *IEEE Transactions on Electron Devices*, 63(10):3819–3829, 2016.

[75] Steve Furber. Large-scale neuromorphic computing systems. *Journal of Neural Engineering*, 13(5):051001, August 2016.

[76] Navnidhi K. Upadhyay, Hao Jiang, Zhongrui Wang, Shiva Asapu, Qiangfei Xia, and J. Joshua Yang. Emerging memory devices for neuromorphic computing. *Advanced Materials Technologies*, 4(4):1800589, 2019.

[77] W. Chen, W. Khwa, J. Li, W. Lin, H. Lin, Y. Liu, Y. Wang, Huaqiang Wu, Huazhong Yang, and M. Chang. Circuit design for beyond von Neumann applications using emerging memory: From nonvolatile logics to neuromorphic computing. In *Proceedings of the International Symposium on Quality Electronic Design*, pages 23–28, 2017.

[78] Michael Herh. Samsung and TSMC Racing to Become First to Start Volume Production of 3-nm Chips, BUSINESSKOREA. http://www.businesskorea.co.kr/news/articleView.html?idxno=40493, January 2020.

[79] Irfan Ahmad. How much data is generated every minute?, SocialMediaToday. https://www.domo.com/solution/data-never-sleeps-6, June 2018.

[80] Donald Michie. "Memo" Functions and Machine Learning. *Nature*, 218(5136):19–22, April 1968.

[81] F. Morgado-Dias and Alexandre Mota. Artificial neural networks: A review of commercial hardware. *Engineering Applications of Artificial Intelligence*, 17:945–952, 12 2004.

[82] D. Parra and C. Camargo. A Systematic Literature Review of Hardware Neural Networks. In *Proceedings of the Colombian Conference on Applications in Computational Intelligence*, pages 1–6, 2018.

[83] F. Akopyan, J. Sawada, A. Cassidy, R. Alvarez-Icaza, J. Arthur, P. Merolla, N. Imam, Y. Nakamura, P. Datta, G. Nam, B. Taba, M. Beakes, B. Brezzo, J. B. Kuang, R. Manohar, W. P. Risk, B. Jackson, and D. S. Modha. TrueNorth: Design and Tool Flow of a 65 mW 1 Million Neuron Programmable Neurosynaptic Chip. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 34(10):1537–1557, October 2015.

[84] Robert A. Nawrocki, Richard M. Voyles, and Sean E. Shaheen. A Mini Review of Neuromorphic Architectures and Implementations. *IEEE Transactions on Electron Devices*, 63(10):3819–3829, October 2016.

[85] S. Haykin. Neural Networks: A Comprehensive Foundation. *Upper Saddle River, NJ, USA: Prentice-Hall*, 1998.

[86] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(1522-9602):115–133, December 1943.

[87] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.

[88] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[89] Amirhossein Tavanaei, Masoud Ghodrati, Saeed Reza Kheradpisheh, Timothée Masquelier, and Anthony Maida. Deep learning in spiking neural networks. *Neural Networks*, 111:47–63, 2019.

[90] M. Abeles, H. Bergman, E. Margalit, and E. Vaadia. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *Journal of Neurophysiology*, 704:1629–1638, 1993.

[91] W. Bair. Reliable temporal modulation in cortical spike trains in the awake monkey. In *Proceedings of the symposium on dynamics of neural processing*, volume 112, pages 57–63, 1994.

[92] Wolfgang Maass. Networks of spiking neurons: The third generation of neural network models. *Neural Networks*, 10(9):1659–1671, 1997.

[93] Jesus L. Lobo, Javier Del Ser, Albert Bifet, and Nikola Kasabov. Spiking Neural Networks and online learning: An overview and perspectives. *Neural Networks*, 121:88 – 100, 2020.

[94] Wulfram Gerstner and Werner M. Kistler. Spiking Neuron Models: Single Neurons, Populations, Plasticity. *cambridge University Press*, 2002.

[95] Lei Deng, Yujie Wu, Xing Hu, Ling Liang, Yufei Ding, Guoqi Li, Guangshe Zhao, Peng Li, and Yuan Xie. Rethinking the Performance Comparison between SNNs and ANNs. *Neural Networks*, 121:294–307, 2020.

[96] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury. Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.

[97] G. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[98] G. Hinton, S. Osindero, and Y. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18(7):1527–1554, 2006.

[99] D. Pham, M. Packianather, and E. Charles. Control chart pattern clustering using a new selforganizing spiking neural network. In *Proceedings of the Institution of Mechanical Engineers*, volume 222, pages 1201–1211, 2008.

[100] R. Brette, M. Rudolph, M. Carnevale, T. Hines, M. Beeman, D. Bower, and A. Destexhe. Simulation of networks of spiking neurons: A review of tools and strategies. *Journal of Computational Neuroscience*, 23(3):349–398, 2007.

[101] W. Gerstner and W. M. Kistler. Spiking neuron models: Single neurons, populations, plasticity. *Cambridge University Press*, 2002.

[102] N. Kasabov, K. Dhoble, N. Nuntalid, and G. Indiveri. Dynamic evolving spiking neural networks for on-line spatio and spectro-temporal pattern recognition. *Neural Networks*, 41:188–201, 2013.

[103] L.F Abbott. Lapicque's introduction of the integrate-and-fire model neuron (1907). *Brain Research Bulletin*, 50(5):303–304, December 1999.

[104] E. M. Izhikevich. Simple model of spiking neurons. *IEEE Transactions on neural networks*, 14(6):1569–1572, 2003.

[105] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500–544, 1952.

[106] Misha Mahowald Rodney Douglas. A silicon neuron. *Nature*, 354(3):515–518, December 1991.

[107] J. V. Arthur and K. A. Boahen. Silicon-Neuron Design: A Dynamical Systems Approach. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 58(5):1034–1043, May 2011.

[108] A. Joubert, B. Belhadj, O. Temam, and R. Héliot. Hardware spiking neurons design: Analog or digital? In *Proceedings of the International Joint Conference on Neural Networks*, pages 1–5, June 2012.

[109] G. Indiveri, E. Chicca, and R. Douglas. A VLSI array of low-power spiking neurons and bistable synapses with spike-timing dependent plasticity. *IEEE Transactions on Neural Networks*, 17(1):211–221, January 2006.

[110] Jayawan H.B. Wijekoon and Piotr Dudek. Compact silicon neuron circuit with spiking and bursting behaviour. *Neural Networks*, 21(2):524–534, April 2008.

[111] A. Joubert, B. Belhadj, and R. Héliot. A robust and compact 65 nm LIF analog neuron for computational purposes. In *Proceedings of the International New Circuits and systems conference*, pages 9–12, June 2011.

[112] Jonghan Shin and C. Koch. Dynamic range and sensitivity adaptation in a silicon spiking neuron. *IEEE Transactions on Neural Networks*, 10(5):1232–1238, September 1999.

[113] K. M. Hynna and K. Boahen. Silicon neurons that burst when primed. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, pages 3363–3366, May 2007.

[114] Arindam Basu, Sun Shuo, Hongming Zhou, Meng Hiot Lim, and Guang-Bin Huang. Silicon spiking neurons for hardware implementation of extreme learning machines. *Neurocomputing*, 102:125 – 134, February 2013.

[115] P. Merolla, J. Arthur, F. Akopyan, N. Imam, R. Manohar, and D. S. Modha. A digital neurosynaptic core using embedded crossbar memory with 45 pJ per spike in 45nm. In *Proceedings of the IEEE Custom Integrated Circuits Conference*, pages 1–4, September 2011.

[116] R. Emery, A. Yakovlev, and G. Chester. Connection-centric network for spiking neural networks. In *Proceedings of the ACM/IEEE International Symposium on Networks-on-Chip*, pages 144–152, May 2009.

[117] S. Lashkare, S. Chouhan, T. Chavan, A. Bhat, P. Kumbhare, and U. Ganguly. PCMO RRAM for Integrate-and-Fire Neuron in Spiking Neural Networks. *IEEE Electron Device Letters*, 39(4):484–487, April 2018.

[118] A. Jaiswal, S. Roy, G. Srinivasan, and K. Roy. Proposal for a leaky-integrate-fire spiking neuron based on magnetoelectric switching of ferromagnets. *IEEE Transactions on Electron Devices*, 64(4):1818–1824, April 2017.

[119] T. Tuma, A. Pantazi, M. Le Gallo, A. Sebastian, and E. Eleftheriou. Stochastic phase-change neurons. *Nature Nanotechnology*, 11:693, May 2016.

[120] Sangya Dutta, Vinay Kumar, Aditya Shukla, Nihar R. Mohapatra, and Udayan Ganguly. Leaky Integrate and Fire Neuron by Charge-Discharge Dynamics in Floating-Body MOSFET. *Scientific Reports*, 7(1):8257, August 2017.

[121] E. R. Kandel and J. H. Schwartz. Principles of Neural Science. *Elsevier*, 1985.

[122] Sentaurus device user guide. *Version H-2013.03, SYNOPSYS*, 2013.

[123] Samsung. Samsung Launches Highest-capacity Mobile DRAM to Accommodate Next-generation Smartphones. `https://www.samsung.com/semiconductor/dram/lpddr4x/`, March 2019.

[124] Roman Pletka, Ioannis Koltsidas, Nikolas Ioannou, Saša Tomić, Nikolaos Papandreou, Thomas Parnell, Haralampos Pozidis, Aaron Fry, and Tim Fisher. Management of Next-Generation NAND Flash to Achieve Enterprise-Level Endurance and Latency Targets. *ACM Transactions on Storage*, 14(4):33:1–33:25, December 2018.

[125] S. Natarajan, M. Agostinelli, S. Akbar, M. Bost, A. Bowonder, V. Chikarmane, S. Chouksey, A. Dasgupta, K. Fischer, Q. Fu, T. Ghani, M. Giles, S. Govindaraju, R. Grover, W. Han, D. Hanken, E. Haralson, M. Haran, M. Heckscher, R. Heussner, P. Jain, R. James, R. Jhaveri, I. Jin, H. Kam, E. Karl, C. Kenyon, M. Liu, Y. Luo, R. Mehandru, S. Morarka, L. Neiberg, P. Packan, A. Paliwal, C. Parker, P. Patel, R. Patel, C. Pelto, L. Pipes, P. Plekhanov, M. Prince, S. Rajamani, J. Sandford, B. Sell, S. Sivakumar, P. Smith, B. Song, K. Tone, T. Troeger, J. Wiedemer, M. Yang, and K. Zhang. A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size. In *Technical Digest of the IEEE International Electron Device Meeting*, pages 3.7.1–3.7.3, December 2014.

[126] Sushobhan Avasthi, William E. McClain, Gabriel Man, Antoine Kahn, Jeffrey Schwartz, and James C. Sturm. Hole-blocking titanium-oxide/silicon heterojunction and its application to photovoltaics. *Applied Physics Letters*, 102(20):203901, May 2013.

[127] Qifeng Zhang, Christopher S. Dandeneau, Xiaoyuan Zhou, and Guozhong Cao. ZnO nanostructures for dye-sensitized solar cells. *Advanced Materials*, 21(41):4087–4108, October 2009.

[128] Benjamin J. Morgan and Graeme W. Watson. Intrinsic n-type defect formation in TiO_2: A comparison of rutile and anatase from gga+u calculations. *The Journal of Physical Chemistry C*, 114(5):2321–2328, January 2010.

[129] Yun Jeong Hwang, Akram Boukai, and Peidong Yang. High density n-Si/n-TiO_2 core/shell nanowire arrays with enhanced photoactivity. *Nano Letters*, 9(1):410–415, December 2009.

[130] Jun Akikusa and Shahed U.M. Khan. Photoresponse and AC impedance characterization of n-TiO_2 films during hydrogen and oxygen evolution reactions in an electrochemical cell. *International Journal of Hydrogen Energy*, 22(9):875–882, December 1997.

[131] Baochen Liao, Bram Hoex, Armin G. Aberle, Dongzhi Chi, and Charanjit S. Bhatia. Excellent c-Si surface passivation by low-temperature atomic layer deposited titanium oxide. *Applied Physics Letters*, 104(25):253903, June 2014.

[132] Ziv Hameiri Xinbo Yang, Klaus Weber and Stefaan De Wolf. Industrially feasible, dopant-free, carrier-selective contacts for high-efficiency silicon solar cells. *Progress in Photovoltaics: Research and Applications*, May 2017.

[133] R. Chau, R. Arghavani, M. Alavi, D. Douglas, R. Green, S. Tyagi, J. Xu, P. Packan, S. Yu, and Chunlin Liang. Scalability of partially depleted SOI technology for sub-0.25 μm logic applications. In *Technical Digest of the IEEE International Electron Device Meeting*, pages 591–594, December 1997.

[134] Masood Mehrabian, Sina Dalir, and Hossein Shokrvash. Numerical simulation of CdS quantum dot sensitized solar cell using the Silvaco-Atlas software. *Optik - International Journal for Light and Electron Optics*, 127(20):10096 – 10101, October 2016.

[135] G. Rawat, D. Somvanshi, H. Kumar, Y. Kumar, C. Kumar, and S. Jit. Ultraviolet detection properties of p-Si/n-TiO_2 heterojunction photodiodes grown by electron-beam evaporation and sol-gel methods: A comparative study. *IEEE Transactions on Nanotechnology*, 15(2):193–200, March 2016.

[136] H. G. Virani, R. B. R. Adari, and A. Kottantharayil. Dual-k spacer device architecture for the improvement of performance of silicon n-channel tunnel FETs. *IEEE Transactions on Electron Devices*, 57(10):2410–2417, October 2010.

[137] T. Shino, T. Ohsawa, T. Higashi, K. Fujita, N. Kusunoki, Y. Minami, M. Morikado, H. Nakajima, K. Inoh, T. Hamamoto, and A. Nitayama. Operation voltage dependence of memory cell characteristics in fully depleted floating-body cell. *IEEE Transactions on Electron Devices*, 52(10):2220–2226, 2005.

[138] L. De Michielis, L. Lattanzio, K. E. Moselund, H. Riel, and A. M. Ionescu. Tunneling and occupancy probabilities: How do they affect tunnel-fet behavior? *IEEE Electron Device Letters*, 34(6):726–728, June 2013.

[139] K. G. Anil, S. Mahapatra, and I. Eisele. Role of inversion layer quantization on sub-bandgap impact ionization in deep-sub-micron n-channel MOSFETs. In *Technical Digest of the IEEE International Electron Device Meeting*, pages 675–678, December 2000.

[140] K. G. Anil, S. Mahapatra, and I. Eisele. Electron-electron interaction signature peak in the substrate current versus gate voltage characteristics of n-channel silicon MOSFETs. *IEEE Transactions on Electron Devices*, 49(7):1283–1288, July 2002.

[141] B. Das, S. Sushama, J. Schulze, and U. Ganguly. Sub-0.2 V impact ionization in Si n-i-p-i-n diode. *IEEE Transactions on Electron Devices*, 63(12):4668–4673, December 2016.

[142] T. Saraya, M. Takamiya, T. N. Duyet, T. Tanaka, H. Ishikuro, T. Hiramoto, and T. Ikoma. Floating body effects in 0.15 μm partially depleted SOI MOSFETs below 1 V. In *Proceedings of the IEEE International SOI Conference*, pages 70–71, September 1996.

[143] S. Okhonin, P. Fazan, and M. E. Jones. Zero capacitor embedded memory technology for system on chip. In *Proceedings of the IEEE International*

Workshop on Memory Technology, Design, and Testing, pages 21–25, August 2005.

[144] N. Z. Butt and M. A. Alam. Scaling limits of double-gate and surround-gate Z-RAM cells. *IEEE Transactions on Electron Devices*, 54(9):2255–2262, September 2007.

[145] Asen Asenov. Random dopant induced threshold voltage lowering and fluctuations in sub 50 nm MOSFETS: a statistical 3D atomistic simulation study. *Nanotechnology*, 10(2):153–158, January 1999.

[146] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M. Buehler, A. Cappellani, R. Chau, C. . Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Huessner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. McIntyre, P. Moon, J. Neirynck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams, and K. Zawadzki. A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging. In *Technical Digest of the IEEE International Electron Device Meeting*, pages 247–250, December 2007.

[147] S. Herculano-Houzel. The Human Brain in Numbers: a Linearly Scaled-up Primate Brain. *Frontiers in Human Neuroscience*, 3:31:246–253, November 2009.

[148] SciencePhotoLibrary. `https://in.pinterest.com/pin/567523990527614888/?autologin=true`.

[149] Jeff Hawkins and Subutai Ahmad. Why neurons have thousands of synapses, a theory of sequence memory in neocortex. *Frontiers in Neural Circuits*, 10:23, March 2016.

[150] W. Maass and C. M. Bishop. Pulsed Neural Networks. *MIT Press, Massachusetts*, 275, 1999.

[151] N. Loubet, T. Hook, P. Montanini, C. . Yeung, S. Kanakasabapathy, M. Guillom, T. Yamashita, J. Zhang, X. Miao, J. Wang, A. Young, R. Chao, M. Kang, Z. Liu, S. Fan, B. Hamieh, S. Sieg, Y. Mignot, W. Xu, S. . Seo, J. Yoo,

S. Mochizuki, M. Sankarapandian, O. Kwon, A. Carr, A. Greene, Y. Park, J. Frougier, R. Galatage, R. Bao, J. Shearer, R. Conti, H. Song, D. Lee, D. Kong, Y. Xu, A. Arceo, Z. Bi, P. Xu, R. Muthinti, J. Li, R. Wong, D. Brown, P. Oldiges, R. Robison, J. Arnold, N. Felix, S. Skordas, J. Gaudiello, T. Standaert, H. Jagannathan, D. Corliss, M. . Na, A. Knorr, T. Wu, D. Gupta, S. Lian, R. Divakaruni, T. Gow, C. Labelle, S. Lee, V. Paruchuri, H. Bu, and M. Khare. Stacked nanosheet gate-all-around transistor to enable scaling beyond FinFET. In *Proceedings of the Symposium on VLSI Technology*, pages T230–T231, June 2017.

[152] J. Han, S. Ryu, S. Kim, C. Kim, J. Ahn, S. Choi, J. S. Kim, K. H. Kim, G. S. Lee, J. S. Oh, M. H. Song, Y. C. Park, J. W. Kim, and Y. Choi. A Bulk FinFET Unified-RAM (URAM) Cell for Multifunctioning NVM and Capacitorless 1T-DRAM. *IEEE Electron Device Letters*, 29(6):632–634, June 2008.

[153] B. Rajendran, Y. Liu, J. Seo, K. Gopalakrishnan, L. Chang, D. J. Friedman, and M. B. Ritter. Specifications of Nanoscale Devices and Circuits for Neuromorphic Computational Systems. *IEEE Transactions on Electron Devices*, 60(1):246–253, January 2013.

[154] Young Jun Lee, Jihyun Lee, Y. B. Kim, J. Ayers, A. Volkovskii, A. Selverston, H. Abarbanel, and M. Rabinovich. Low power real time electronic neuron VLSI design using subthreshold technique. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, volume 4, pages IV–744, May 2004.

[155] Subramanian S. Iyer. The Evolution of embedded Memory in High Performance Systems - A case of Orthogonal Scaling. In *Proceedings of the IEEE SOI-3D-Subthreshold conference*, 2013.

[156] Roland Kircher and Wolfgang Bergner. *Examples: Leakage in DRAM Cell Structures (chapter-5) in Three-Dimensional Simulation of Semiconductor Devices*, pages 87–115. Birkh user Basel, 1991.

[157] A. Sengupta, P. Panda, P. Wijesinghe, Y. Kim, and K. Roy. Magnetic tunnel junction mimics stochastic cortical spiking neurons. *Scientific Reports*, pages 1–8, July 2016.

[158] Kinam Kim, Chang-Gyu Hwang, and Jong Gil Lee. DRAM technology perspective for gigabit era. *IEEE Transactions on Electron Devices*, 45(3):598–608, March 1998.

[159] H. Tang, K. Prasad, R. Sanjinès, P. E. Schmid, and F. Lévy. Electrical and optical properties of TiO_2 anatase thin films. *Journal of Applied Physics*, 75(4):2042–2047, February 1994.

Publications

List of Publications

- **D. Chatterjee** and A. Kottantharayil, *"A TiO_2 S/D n-Channel FD-SOI MOSFET Based Zero Capacitor Random Access Memory Device*", **Journal of Computational Electronics**, Oct. 2020, DOI: 10.1007/s10825-020-01594-3.

- **D. Chatterjee** and A. Kottantharayil, *"A CMOS Compatible Bulk FinFET Based Ultra Low Energy Leaky Integrate and Fire Neuron For Spiking Neural Networks*", **IEEE Electron Device Letters (EDL)**, vol. 40, no. 8, pp. 1301-1304, Aug. 2019, DOI: 10.1109/LED.2019.2924259.

- **D. Chatterjee** and A. Kottantharayil, *"An Improved 1T-DRAM Cell Using TiO_2 as the Source and Drain of an n-Channel PD-SOI MOSFET* ", **Proceedings in** 76^{th} **IEEE Device Research Conference (DRC)**, June, 2018 (University of California Santa Barbara, CA, USA), DOI: 10.1109/DRC.2018.8442180.