PERFORMANCE ASSESSMENT OF TUNNELING BASED TRANSISTORS FOR CAPACITORLESS DYNAMIC MEMORY APPLICATIONS

Ph.D. Thesis

By

NUPUR NAVLAKHA

DISCIPLINE OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

JANUARY 2018

PERFORMANCE ASSESSMENT OF TUNNELING BASED TRANSISTORS FOR CAPACITORLESS DYNAMIC MEMORY APPLICATIONS

A THESIS

Submitted in partial fulfillment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

by

NUPUR NAVLAKHA

DISCIPLINE OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY INDORE
JANUARY 2018

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **PERFORMANCE ASSESSMENT OF TUNNELING BASED TRANSISTORS FOR CAPACITORLESS DYNAMIC MEMORY APPLICATIONS** in the partial fulfillment of the requirements for the award of the degree of **DOCTOR OF PHILOSOPHY** and submitted in the **DISCIPLINE OF ELECTRICAL ENGINEERING, Indian Institute of Technology Indore**, is an authentic record of my own work carried out during the time period from July 2014 to January 2018 under the supervision of Dr. Abhinav Kranti, Professor, Discipline of Electrical Engineering, Indian Institute of Technology Indore.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institute.

Signature of the student with date
(NUPUR NAVLAKHA)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Signature of Thesis Supervisor
(Prof. ABHINAV KRANTI)

NUPUR NAVLAKHA has successfully given her Ph.D. Oral Examination held on **July 24, 2018.**

Signature of Chairperson	Signature of External Examiner	Signature of Thesis Supervisor
Date:	Date:	Date:
Signature of PSPC Member	Signature of PSPC Member	Signature of Convener, DPGC
Date:	Date:	Date:

Signature of Head of Discipline
Date:

ACKNOWLEDGEMENTS

Foremost, I would like to express my deep sense of gratitude to my Ph.D. supervisor Prof. Abhinav Kranti for his continuous guidance, support, encouragement, and patience. It was his invaluable motivation and immense knowledge, which enabled me to successfully complete my doctoral research and Ph.D. dissertation. It has been a memorable experience for me throughout my doctoral work at IIT Indore.

I am extremely grateful to my doctoral committee members, Dr. Shaibal Mukherjee and Dr. Kazi Sabiruddin, for their encouragement, and valuable comments, which helped in improving my research work.

My sincere gratitude to Prof. Jyi-Tsong Lin for his encouragement and support throughout the research work. I am grateful to him for the valuable suggestions and the discussions that have helped me in better understanding and improving the research work.

I would like to thank IIT Indore for providing the essential research facilities for my thesis work. I gratefully acknowledge the contribution of all the faculty members of Electrical Engineering for their help and support. I would also like to thank all the staff members of IIT Indore for their kind help with all the academic work.

I would like to express my sincere gratitude to Ministry of Human Resource Development (MHRD), Government of India, for providing the Teaching Assistantship during the course of the Ph.D. I would also like to thank Department of Science and Technology (DST), and IIT Indore for the travel funding for attending and presenting my work in international conferences. I would also like to acknowledge the on-going Indo-Taiwan bilateral research project funded by DST through Global Innovation Technology Alliance (GITA).

I would like to express my wholehearted thanks to my friends, Hasan and Manish for their trustworthy help, constant encouragement and support during the research work. I would

also like to thank Mukta, Bhuvaneshwari, Praveen, Nivedita, Pramila and Pooja for their valuable support and co-operation during my work at IIT Indore.

I express my deepest gratitude to my grandparents, Shri Girnar Singh Navlakha, and Dr. Nirmala Navlakha for encouraging me to pursue Ph.D. I am wholeheartedly thankful to my parents, Mrs. Sarita Navlakha and Mr. Atul Navlakha, my brother, Mr. Sarthak Navlakha and other family members who have always motivated me and nurtured me through their love, affection, and care. I cannot forget to acknowledge my few friends, Apeksha, Praneeta, Nitisha, Yash, Tripti, Priyanka, Sapna, Pragya, Rahul, Nikita, Bhawna, Aditya, and Mallikarjuna. I thank them all for the memorable moments, and much needed moral support.

Finally, I want to express my heartiest gratefulness to almighty, who gave me life, faith with capability for this very crucial period of my life.

Nupur Navlakha

Dedicated to my parents

ABSTRACT OF THE DISSERTATION

Performance Assessment of Tunneling based Transistors for Capacitorless Dynamic Memory Applications

With the advent of Moore's law and Dennard's scaling theory, the performance of processor units in computers improved. However, the overall system performance which was based on the interaction between the processor and memory units, didn't improve with the same pace due to the lack of focus on memory. Thus, in the last decade, the emphasis has been directed to enhance the speed and density of memory with operation at low power. This necessitated reduction in the size of the capacitor in the conventional Dynamic Random Access Memory (DRAM) based on one transistor and capacitor (1T-1C). However, capacitor scaling adversely affects the charge retention, requires more refresh cycles, and consequently, dissipates more power. The problem was circumvented with the introduction of the single transistor (1T) as DRAM cell. While conventional Metal Oxide Semiconductor Field Effect Transistor (MOSFET) based 1T-DRAMs have shown promising results, the focus has shifted towards use of devices with potential to operate at low power. Thus, the thesis work focuses on an energy efficient device, Tunnel FET (TFET) as DRAM, improving its Retention Time (*RT*), Sense Margin (*SM*), speed, and scalability at low power.

Through the TFET design presented in previous works, the *RT* achieved was lower than the target of 64 ms, specified by International Technology Roadmap for Semiconductors (ITRS). Therefore, a careful reinvestigation is required to enhance the *RT* of TFET based DRAM. The work in the thesis provides insights into the understanding of the performance and behaviour of TFET devices for memory applications by means of comprehensive physical device simulations. The key contribution of this research is to provide insights into the physical phenomenon occurring in the device, which influences the operation of TFET as dynamic memory. The thesis work demonstrates device perspective, where various metrics of DRAM are regulated by device architecture (misaligned, twin, and planar tri-gate TFET), geometry (gate lengths, film thickness), parameters (oxide thickness, gate workfunctions), biases and temperature. These govern hole generation and recombination in the storage region that defines distinct operations (write, read and hold) of DRAM.

The functionality of TFET based dynamic memory is based on the distinct roles of the gates, where the first gate (G1), aligned to source at top surface, is utilized for read mechanism based on band-to-band tunneling, while the second gate (G2) is responsible for creation of a dedicated volume for charge storage. G2 is aligned to drain, positioned at the bottom, at the front un-gated region in misaligned DG, and at the front adjacent to the first gate in twin gate and planar tri-gate topologies. The non-overlapping of G1 and G2 result into a profound well formation that enhances the charge retention. The p^+ poly G2 creates a deep potential well that sustains charges for longer duration. Results demonstrate *RT* in few seconds for all proposed architectures with length of G1 (L_{g1}) as 400 nm and length of G2 (L_{g2}) as 200 nm at 85 °C, which shows a remarkable improvement over previously proposed structures.

The drawback of low *SM* has been overcome through incorporation of a symmetric G1 as in planar tri-gate TFET. The improved electrostatic control of G1 over channel, improves the *SM*, and also, scalability of G1 (down to 25 nm) and drain voltage (0.8 V) with acceptable *SM* and *RT* in planar tri-gate topology. Exploring TFET for functionality at shorter gate lengths, the design optimization through incorporation of a lateral spacing between the gates (L_{gap}) and an underlap (L_{un}) between the drain and G2, can further enhance retention and scalability. While scaling G2 reduces *RT*, scaling G1 degrades *SM*. Therefore, an analysis of each individual length ($L_{g1}, L_{g2}, L_{gap}, L_{un}$) estimates the minimum length required to attain *RT* > 64 ms. While the total length ($L_{total} = L_{g1} + L_{g2} + L_{gap} + L_{un}$) for misaligned DG and twin gate TFET can be scaled down to ~160 nm, planar tri-gate show an improved scalability down to ~115 nm. Other than improved *SM* and *RT* at reduced size and high temperature, functionality at lower drain voltage with low write time, nominates TFET for embedded applications.

The work presented in the thesis showcases new viewpoints for TFET to function efficiently as dynamic memory. The physical insights and analysis of different attributes with optimal utilization of each lead to improved metrics as well as suppressed trade-offs. The systematic analysis through innovative approaches leads to capacity, retention, at low energy with operation at reduced size. The use of an energy optimized DRAM memory with *RT* > 64 ms is well-suited for standalone applications, and as well as, for integrated circuits embedded with logic devices.

LIST OF PUBLICATIONS

A. Patent:

1. Abhinav Kranti, and **Nupur Navlakha**, "Multiple gate tunneling field effect transistor device for capacitorless dynamic memory," India, 201621007078 (2016), Status: Pending.

B. Peer-reviewed Journals:

1. **Nupur Navlakha**, and Abhinav Kranti, "Overcoming the drawback of lower sense margin in tunnel FET based dynamic memory along with enhanced charge retention and scalability," *Nanotechnology*, 28, no. 44, article 445203 (2017).

2. **Nupur Navlakha**, and Abhinav Kranti, "Insights into operation of planar tri-gate tunnel field effect transistor for dynamic memory application," *Journal of Applied Physics*, 122, no. 4, article 044502 (2017).

3. **Nupur Navlakha**, Jyi-Tsong Lin, and Abhinav Kranti, "Retention and Scalability Perspective of Sub 100 nm Double Gate Tunnel FET DRAM," *IEEE Transactions on Electron Device*, 64, no. 4, pp. 1561 - 1567 (2017).

4. **Nupur Navlakha**, Jyi-Tsong Lin, and Abhinav Kranti, "Improved retention time in twin gate 1T DRAM with tunneling based read mechanism", *IEEE Electron Device Letters*, 37, no. 9, 1127-1130 (2016).

5. **Nupur Navlakha**, Jyi-Tsong Lin, and Abhinav Kranti, "Improving retention time in tunnel field effect transistor based dynamic memory by back gate engineering", *Journal of Applied Physics*, 119, no. 2, article 214501 (2016).

C. Proceedings in International Conferences:

1. **Nupur Navlakha**, Jyi-Tsong Lin, and Abhinav Kranti, "Tunnel FET Capacitorless DRAM," in proceedings of *19th International Workshop on Physics of Semiconductor Devices, IWPSD 2017*, New Delhi, India, December (2017).

2. **Nupur Navlakha**, and Abhinav Kranti, "Performance Assessment of Tunnel Field Effect Transistors based Capacitorless Dynamic Memory," in abstracts of *6th International Symposium on Integrated Functionalities, ISIF 2017*, New Delhi, India, pp. 18-19, December (2017).

3. **Nupur Navlakha**, Jyi-Tsong Lin, and Abhinav Kranti, "Design Optimization of Tunnel FET for Dynamic Memory Applications," *13th IEEE Electron Devices and Solid-State Circuits, EDSSC 2017*, Hsinchu, Taiwan, October (2017), doi: 10.1109/EDSSC.2017.8126454.

4. **Nupur Navlakha**, Jyi-Tsong Lin, and Abhinav Kranti, "Twin gate Tunnel FET based capacitorless dynamic memory, in *IEEE International Symposium on VLSI Technology, Systems And Application, VLSI-TSA 2017*, Hsinchu, Taiwan, April (2017), doi: 10.1109/VLSI-TSA.2017.7942458

5. **Nupur Navlakha**, Jyi-Tsong Lin, and Abhinav Kranti, "Optimization of Back Gate Workfunction, Alignment and Bias for Charge Retention in TFET based DRAM," in *3rd IEEE International Conference on Emerging Electronics, ICEE 2016*, Mumbai, India, December (2016), doi: 10.1109/ICEmElec.2016.8074567.

6. **Nupur Navlakha**, Jyi-Tsong Lin, and Abhinav Kranti, "Enhanced Retention Characteristics in Double Gate Tunnel FET based DRAM," in extended abstracts of the *48th International Conference on Solid State Devices and Materials, SSDM 2016*, Tsukuba, Japan, pp. 735-736 (2016).

7. **Nupur Navlakha**, and Abhinav Kranti, "Impact of Back Gate Workfunction on Performance of TFET based DRAM," in abstracts of *18th International Workshop on Physics of Semiconductor Devices, IWPSD 2015*, Bangalore, India, pp. 402, December (2015).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

1T	*Single Transistor*
3D	*Three-dimensional*
3T-1C	*Three Transistor-One Capacitor*
BL	*Bitline*
BOX	*Buried Oxide*
BTBT	*Band-to-Band Tunneling*
CMOS	*Complementary Metal Oxide Semiconductor*
CPU	*Central Processing Unit*
CR	*Current Ratio*
DDR	*Double Data Rate*
DG	*Double Gate*
DRAM	*Dynamic Random Access Memory*
eDRAM	*Embedded Dynamic Random Access Memory*
EEPROM	*Electrically Erasable Programmable Read Only Memory*
EHPs	*Electron-Hole Pairs*
EPROM	*Erasable Programmable Read Only Memory*
ESD	*Electrostatic Discharge*
eV	*Electron Volt*
FBCs	*Floating Body Cells*
FBEs	*Floating Body Effects*
FB-FET	*Feedback Field Effect Transistor*
FDSOI	*Fully Depleted Silicon on Insulator*
FED	*Field Effect Diode*
FET	*Field Effect Transistor*
FS	*Forward Sweep*
GEN-REC	*Generation Recombination*

GIDL	*Gate Induced Drain Leakage*
H-FED	*Heterostructure-FED*
IMOS	*Impact Ionization MOS*
IoT	*Internet of Things*
IT	*Information Technology*
ITRS	*International Technology Roadmap for Semiconductors*
LPDDR4	*Low Power Double Data Rate-4*
M-FED	*Modified-FED*
MOSFET	*Metal Oxide Semiconductor Field Effect Transistor*
NC	*Negative Capacitance*
NVRAM	*Non-volatile Random Access Memory*
OTAs	*Operational Transconductance Amplifiers*
PDSOI	*Partially Depleted Silicon on Insulator*
PRE	*Precharge*
R&D	*Research and Development*
RDF	*Random Dopant Fluctuation*
RS	*Reverse Sweep*
RT	*Retention Time*
SA	*Sense Amplifier*
SCE	*Short Channel Effect*
SDRAM	*Synchronous Dynamic Random Access Memory*
SFED	*Side Contacted-FED*
SILVACO	*Device Simulation Software*
SM	*Sense Margin*
SOI	*Silicon on Insulator*
SRAM	*Static Random Access Memory*
SRH	*Shockley Read Hall*
STI	*Shallow Trench Isolation*

TCCT	*Thin Capacitively Coupled Thyristor*
TFET	*Tunnel Field Effect Transistor*
UTBOX	*Ultra-Thin Buried Oxide*
WKB	*Wentzel-Kramer-Brillouin*
WL	*Wordline*
WSTS	*World Semiconductor Trade Statistics*
Z^2-FET	*Zero sub-threshold swing and Zero impact ionization FET*
Z-RAM	*Zero capacitor RAM*

NOMENCLATURE

κ	Gate dielectric
Φ	Potential at the surface
α	Unitless scaling constant
ΔI_{ds}	Shift in current levels
ΔV	Potential depth
ΔV_{th}	Shift in threshold voltage,
C_B	Bitline capacitance
C_d	Depletion capacitances
C_{g1}	Gate capacitances at G1
C_{g2}	Gate capacitances at G2
C_{ox}	Oxide capacitances
C_s	Storage capacitance
D	Statistical factor in BTBT
E	Magnitude of the electric field
E_c	Conduction band energy
E_v	Valence band energy
E_f	Fermi-level
E_g	Band gap
g_m	Transconductance
H1	Hold ‘1’
H0	Hold ‘0’
I	Current
I_1	Read current for state ‘1’
I_0	Read currents for state ‘0’
I_d	Drain current
I_{off}	Off-current
I_{on}	On-current
I_{on}/I_{off}	On-to-off current ratio, Switching speed

k	Boltzmann constant
kT/q	Thermal voltage
L_{g}	Gate length
L_{gap}	Lateral spacing between the gates
L_{in}	Front ungated region
L_{total}	Total length between source and drain
L_{un}	Underlap between drain and gate
m^*	Effective mass
N_a	Doping concentration
q	Electronic charge
Q_s	Stored hole charge
R0	Read ‘0’
R1	Read ‘1’
S	Subthreshold swing
S/D	Source/Drain
T	Temperature
T_{BOX}	Buried oxide thickness
T_{ox}	Oxide thickness
T_{si}	Silicon film thicknes
V	Voltage
V_{A}	Array voltage in DRAM cell
V_{B}	Bitline voltage in DRAM cell
V_{bg}	Back gate voltage
V_{d}	Drain voltage
V_{dd}	Supply voltage
V_{dd} - V_{th}	Overdrive voltage
V_{fg}	Front gate voltage
V_{g}	Gate voltage
V_{n}	Electron injection barrier
V_{p}	Hole injection barrier
V_{th}	Threshold voltage of MOSFET

W0	Write ‘0’
W1	Write ‘1’
W_g	Gate width
$\Delta\Phi$	Energy window for tunneling
ε	Dielectric permittivity
λ	Screening tunnelling length
$\tau_{n0/p0}$	Electron and hole lifetime at 300 K
φ_m	Gate workfunction
Ψ_s	Surface potential

Chapter 1

Introduction

1.1 Motivation for memory

Since last 50 years, the trend in the semiconductor industry is directed towards device scaling for increasing both, speed and density along with reduction in power dissipation, intended to increase device functionality [1-13]. The concept of miniaturization was guided by Moore's law [2,3] in conjugation with scaling theory proposed by Dennard *et al.* [4]. In 1965, Gordon Moore postulated that the number of transistors would double annually, which was subsequently revised to 18 months [3]. The vision of Moore continued to increase the transistor count [5,6], which on integration with Dennard's scaling [7] resulted in improvement in switching speed as well as power dissipation [8-13], indicative in Table 1.1.

Table 1.1 Dennard's scaling results for device performance [3].

Device and Circuit Parameter	***Scaling Factor***
Device dimensions, W_g, L_g, T_{ox}	$1/\alpha$
Doping Concentration, N_a	α
Voltage, V	$1/\alpha$
Current, I	$1/\alpha$
Capacitance, $\varepsilon A/T_{ox}$	$1/\alpha$
Delay time, CV/I	$1/\alpha$
Power dissipation, VI	$1/\alpha^2$

α is the unitless scaling constant.
W_g, L_g, and T_{ox}, represent the gate width, length, and oxide thickness, respectively.
ε is the dielectric permittivity.
A is the area of the associated capacitance.

The advent of Moore's law and Dennard's scaling enhanced the speed and scalability of Central Processing Unit (CPU) in computers [13-22]. While the

microprocessors have been emphasized for high performance, memories were aimed at being cost-effective [18,19]. The processors have benefited from the use of advanced transistors, metal and packaging technology, in the quest for high speed and processing. Additionally, computer architecture included larger caches, multithreading, deeper pipelines, innovations that transited finally to multicore processors for improving the performance [20-24]. However, the improved speed through multicore increased the burden on memory as the memory controllers per core decreased [14, 19]. Thus, the overall system performance which is based on the interaction between the processor and memory units, didn't improve with the same pace due to the lack of focus on memory system. Thus, in the last decade the focus has been directed to enhance the memory performance [25-28].

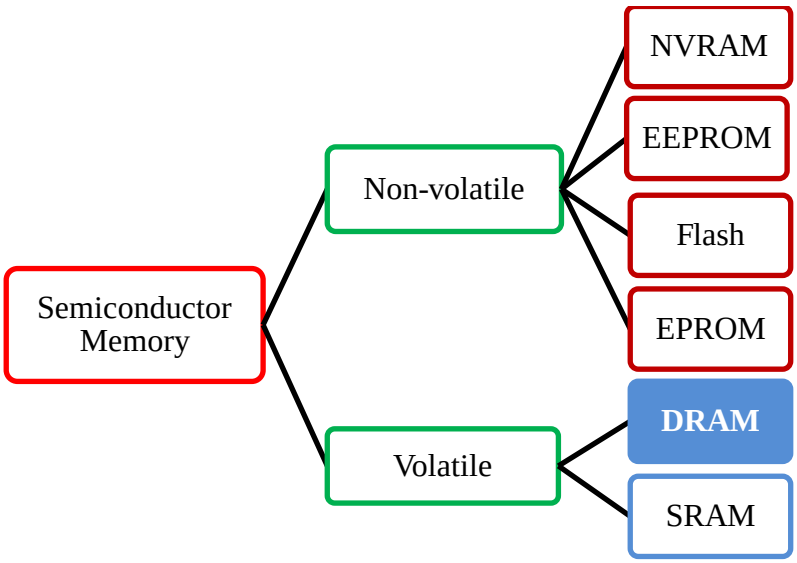

Fig. 1.1 Classification of various solid-state memories [25].

1.2 Memory technology

1.2.1 Types of memory

The quest for improved memory performance resulted into invention of various memory technologies over the past decades [25-34]. Broadly classified as volatile and non-volatile memory (Fig. 1.1), which includes, Static Random Access Memory (SRAM), Dynamic RAM (DRAM), Non-volatile Random Access Memory (NVRAM), NOR flash, Erasable Programmable Read Only Memory (EPROM), Electrically Erasable Programmable Read Only Memory (EEPROM), and NAND flash [32-34]. A volatile memory does not retain data when power is

turned off, while, a non-volatile memory retains data even after the power is turned off. Their utility in real-time applications is based on various parameters, a few of them are illustrated in Table 1.2.

Table 1.2 Comparison of various metrics of memory technologies [25,31,32].

Memory	***Volatility***	***Density***	***Write Speed***	***Read Speed***	***Cost/bit***
Flash	No	High	Slow	Fast	Moderate
EPROM	No	High	Slow	Fast	Moderate
EEPROM	No	Moderate	Slow	Fast	Expensive
SRAM	Yes	Low	Fast	Very Fast	Expensive
DRAM	Yes	Very High	Moderate	Moderate	Moderate

1.2.2 Evolution of memory

The semiconductor industry has evolved remarkably in the past few years with memory technology as an essential and crucial component of electronic system. A tremendous change has been observed in the usage and applicability of the solid-state memories [35]. The paradigm shift is attributed to invention of various mobile multimedia applications and advanced cell phones along with computing segment [29,30-36]. Current smartphones and tablets require much more memory than previous mobile devices [38]. The need is also driven by the continuous growth in the digital information and internet applications that generates immense amount of data which are stored in a cloud [39-41]. Hence, the focus of the devices is shifting from computational to data intensive applications that further demands innovation [18,20,42,43].

Another requirement is due to shift in the trend of memory devices from stand-alone applications to embedded applications, where the memory system is incorporated with logic devices on an integrated circuit [17,44-46]. The technology metrics for memory are stability, reliability, data retention, on–off ratio, power and endurance [32]. Thus, the key contributors necessitating the need for innovative technology includes, Internet of Things (IoT), mobility, networking, cloud computing and big data application [18,29,30,36-43] that drives the need for ease of system integration, lower power consumption,

increased storage and memory density, faster storage and retrieval, real time analytics [47-49]. The major barrier is trade-off between these metrics, and hence, optimization is crucial to design application specific memory devices.

1.2.2.1 Memory technology in market

The dominating memory technologies in the semiconductor market today are SRAM and DRAM (volatile) [14,19,29], and NAND flash (non-volatile) [29,46]. Amongst non-volatile memories, flash memory is the leading memory as along with capability to electrically erase and programme like EPROM and EEPROM, it features lower cost, better scaling perspective and faster read [13,18,32]. Amongst the volatile memory, SRAM is exploited for high speed applications, and is usually utilized as cache memory [19]. Unlike DRAM, SRAM doesn't require refresh cycles and thus, retains the information with power-on [19,29]. DRAM is usually used as main memory, and sometimes, as embedded cache memory [19,35,45], but requires a controller to refresh its data after a fixed time interval. However, DRAM is a dense memory, and thus, high integration capability along with low power and cost, nominates it as a memory to be explored further [18].

1.3 Dynamic Memory

1.3.1 Motivation for DRAM

DRAM is a fast and robust memory with high density, low power and cost [50-56], that could be used for the data management [39-43], but need further innovations to compete with cheap and low power NAND memory [46]. The various applications of DRAM in the past and prospects in the future have motivated continuous and evolutionary, as well as revolutionary changes in the DRAM technology [18,29,55,56]. Despite the introduction of new memory devices, the need for main memory and scratch pad memory would continue the demand for DRAM in future [47].

1.3.1.1 Evolution of dynamic memory in semiconductor industry

Apple II was the first commercially successful desktop computer built with DRAM [19]. In 1990s, DRAM used a data paging system for faster memory access [14,17]. Further, need for enhanced capacity and speed, invented a new form of DRAM which featured multitasking, where the memory device would read data while fetching new

information from hard drive [19]. The next significant upgrade was the Synchronous Dynamic Random Access Memory (SDRAM), which was synchronized to a computer's internal clock, resulting in faster processing than its asynchronous predecessors [19]. The next generation utilized series of Double Data Rate (DDR) SDRAM with increased speed [50] that led to development of laptop and notebook [38,47]. Further notable advancement was observed with the introduction of smartphone. Since 2012, most of the memory sector's strength can be attributed to consolidation within the crowded DRAM segment [38,48]. DRAM has established as a leading memory device in the semiconductor industry. World Semiconductor Trade Statistics (WSTS) [48] reports dynamic memory as 13% contributor for the revenue generated in the semiconductor market, and maximum amongst the memories (Fig. 1.2). The sources [37,38,47], also indicate increase in the demand for DRAM due to major growth in mobile market [47]. Thus, DRAM application is not only limited to various applications for computers, including email, video streaming, but also, wide range of smart products, including mobile devices, high definition television and MP3 players [18].

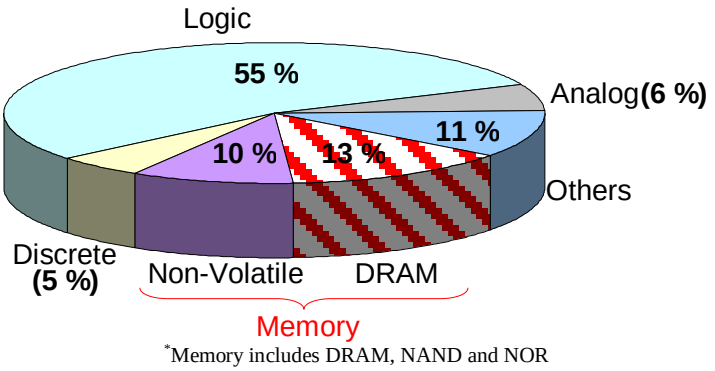

Fig. 1.2 Pie-chart demonstrating the income earned in 2015 World semiconductor market (Total: 334 billion$) through various segments of industry [48].

The popularity of IoT and cloud computing is shifting manufacture's Research and Development (R&D) to develop higher-density DRAM [37,39,40,42]. In fact, according to a survey by Forbes [57], amongst top new areas of Information Technology (IT) in 2015, 32% has been invested for IoT application, while the rest is spend on high performance computing (22%), and energy-saving and carbon-reducing technologies (16%). Thus, IoT application is consistently evolving within the industry. Therefore, the advent of mobile, cloud computing, virtual reality, and IoT applications, thus, require the exploitation of devices to

boost characteristics of DRAM [18,38,47]. An energy-optimized DRAM array provides efficient access to memory bits, providing an "intelligent" memory device [18,29,30]. The use of intelligent memory in the integrated circuits, embedded with logic device would definitely enhance performance. Thus, demand is of devices that are compatible with Complementary Metal Oxide Semiconductor (CMOS) and possess capability to deliver capacity, high speed, high functional bit density, and low energy at reduced size [47].

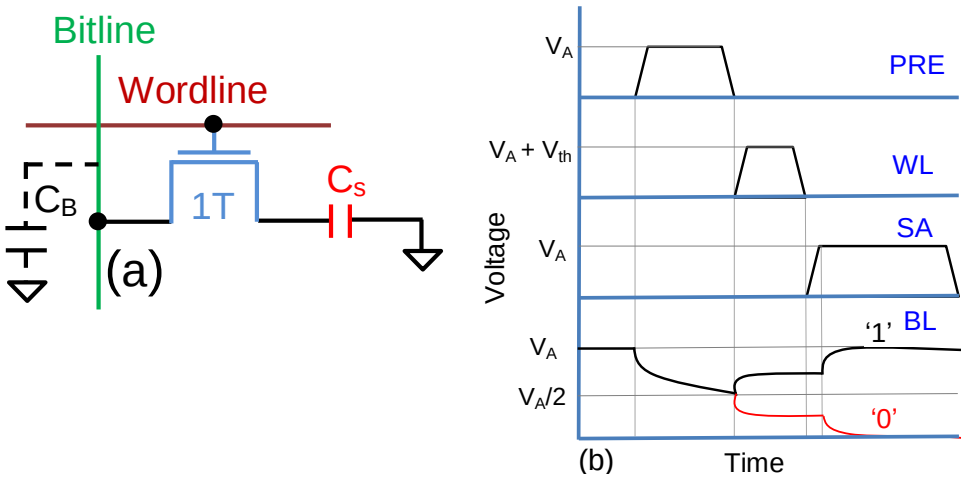

Fig. 1.3 (a) Cell design of 1T-1C DRAM [60]. (b) Qualitative demonstration of timing diagram for read operation for 1T-1C DRAM. SA indicates the signal at the sense amplifier.

1.3.2 History of conventional dynamic memory

In 1966, Robert Dennard perceived an idea of storing data in the form of charge in the capacitor along with a transistor to access the stored data [30]. In 1967, a patent application was filed for the single-transistor (1T) and capacitor (1C) based DRAM (1T-1C) [58-61], and the same was issued in 1968 [58]. The invention by Dennard remarkably improved the mainframe computers and led to development of personal computers [14,19,58]. In early 70's, the commercial DRAM was released by Intel for 1 Kilobyte of storage, but with 3 transistors and a capacitor (3T-1C) [58]. However, in late 70's 1T-1C based DRAM was commercially introduced and till now, there have been various modifications in the design of 1T-1C [19].

1.3.2.1 Operation of conventional dynamic memory

The operation of 1T-1C dynamic memory (Fig. 1.3(a)) is based on charging and discharging of the storage capacitor (C_s) [58-61]. Bit "1" is the state with high

voltage across the capacitor, while bit "0" is the state with storage of zero volts. The bitlines (BLs) are precharged (PRE) to 0.5 × V_A, where V_A is an array voltage which is the maximum positive voltage applied to BLs. The transistor is accessed during write and read operation by applying a high voltage ($\geq V_A + V_{th}$) at the wordline (WL), where V_{th} is the threshold voltage of Metal Oxide Semiconductor Field Effect Transistor (MOSFET) [61]. During write, 'V_A' ('0') is applied at bitline that stores state '1' ('0') at C_s, respectively. After write operation, WL is reset at 0 V and BL to 0.5 × V_A which cuts off the MOSFET, and thus, DRAM is in its hold state.

During read, the transistor is accessed again and based on the data stored in the storage capacitor; the charge is distributed among storage capacitance (C_S) and bitline capacitance (C_B). As shown in Fig. 1.3(b), the wordline, is activated to a boosted voltage, $V_A + V_{th}$. This turns on the access device, and charge sharing occurs between the cell capacitor, C_S, and the bitline-parasitic capacitance, C_B. If the cell to be read stores '1' ('0'), then the cell capacitor, C_S, will be initially charged (discharged) to 'V_A' ('0') V and thus the bitline voltage, V_B, rises (decreases) from '$V_A/2$' to '$V_A/2 + \Delta V_{B1}$' ('$V_A/2 - \Delta V_{B0}$'). Upon activating the sense amplifier by the *SA* signal, the small differential voltage, ΔV_{B1} (ΔV_{B0}), between the two bitlines will be amplified until the *BL* voltage rises (decreases) to 'V_A' ('0') V. The final bitline voltage, 'V_A' ('0') V, is used to write back the stored data on the memory cell as the access transistor is still activated. Thus, the stored data is sensed based on the voltage at the bitline after read operation [60]. An efficient memory should retain the stored data for longer duration. The pursuit to attain higher charge sustenance along with high density led to adaption of various designs of 1T-1C cell.

1.3.2.2 Evolution of 1T-1C DRAM

Scaling the physical capacitor reduces the charge storage capability and hence, applicability as dynamic memory. Thus, the capacitor has evolved from planar to complex three dimensional (3D) structures [14,16,17]. The planar capacitor used in early DRAMs (1970-80) was located adjacent to the transistor and consumed 30% of the cell area and thus, scalability was an issue [50]. The design was improved by use of a trench capacitor, invented by Hideo Sunani in 1974 [62].

The storage capacitor in such topology was fabricated in trench etched into substrate. Although, the area was reduced, there were issues related to fabrication, degraded oxide uniformity on the trench wall, trench to trench leakage associated with parasitic MOS of two memory cells and soft errors that includes impact of high energy particles such as alpha particles [14]. Trench capacitors are also less suitable for integration with high-permittivity (κ) dielectrics, because they are formed before the transistors with their associated high temperature anneals [14,18]. Along with trench capacitors, stacked capacitors were also introduced, with the concept proposed by Mistu Koyangi *et al.* [63].

The stacked capacitor has been fabricated using a thin insulator rounded by two polysilicon layers to reduce the area, and thus, improves the device integrity. However, fabrication and the metal layers used as interconnects increase the complexity and cost [14]. Recently, the focus has diverted towards complex 3D structures [42]. The stacked 3D structures increase the chip density, but with fabrication cost as an overhead [51]. The other requirements of the capacitor for efficient memory operation include, low series resistance for fast memory, and low inter-cell coupling [14,64,65]. Other limitations of 1T-1C DRAM include increase in leakage current associated with transistor scaling that affects the charge retention, and also, the increase in off-current reduces the operating speed [14,19,52-55]. The drawbacks of the 1T-1C, motivated invention of innovative technology to overcome the problem associated with capacitor scaling, and thus, the concept of capacitorless dynamic memory was proposed in 1990's [68].

1.4 Evolution of MOSFET Architectures

1.4.1 Bulk MOSFETs

The conventional silicon wafers, called bulk wafers were ~ 800 micrometer (μm) thick with a thin top region (0.1% of wafer) utilized for conduction, while the rest inactive bottom region of silicon provides strength [69]. However, this unused substrate of a bulk MOSFET (Fig. 1.4) resulted into parasitic effects [69-73] that included,

(1) The latch up, which is superposition of *p-n-p* and *n-p-n* transistors in the same diffused regions. It serves as thyristor causing uncontrollable high current and hence, failure of circuit [71].

(2) Interconnection with other device that require lateral isolation technique, Shallow Trench Isolation (STI) or channel stop implantation that increase fabrication steps and also, consumes larger area [71].

(3) Capacitance between source/drain and the substrate. The device scaling reduces the gate controllability and hence, enhances the Short Channel Effects (SCEs) [74]. The problem can be suppressed with heavier channel doping. However, higher doping further increase the parasitic junction capacitances and has an unfavourable impact on mobility, threshold voltage, transconductance and device variability [69,71].

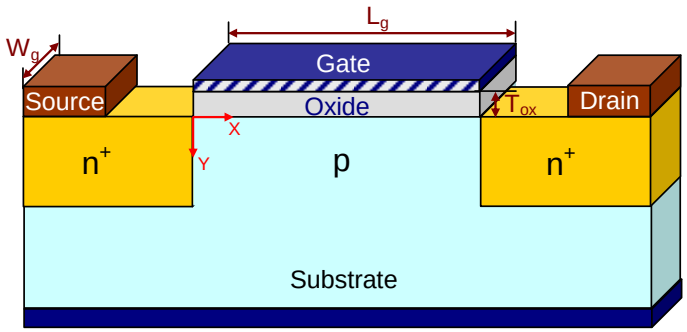

Fig. 1.4 Schematic diagram of an *n*-type bulk MOSFET, with a gate length of L_g, gate oxide thickness of T_{ox} and width of W_g. The doping of substrate is moderate *p*-type and source and drain are heavily doped with *n*-type of impurities (n^+) [71].

1.4.2 Silicon-on-Insulator (SOI) MOSFETs

The issues related to parasitic effects in bulk MOSFETs was addressed by the isolation of active region from substrate through a dielectric film. The technology introduced, known as Silicon-on-Insulator (SOI) [75-78] with active region resting on a Buried Oxide (BOX), as shown in Fig. 1.5. Various advantages [69,71] obtained from SOI technology included,

[1] Dielectric isolation: The isolation between substrate and active region offers following benefits over bulk MOSFETs [79],

(i) No latch up,

(ii) Lower power consumption,

(iii) Reduced leakage current,

(iv) Higher transconductance, and

(v) Reduction in parasitic capacitances

[2] Random Dopant Fluctuations (RDF): Undoped/lower body doping reduces device variability, and thus, is more immune to RDFs [69].

[3] Improved integrity: The capability of integrating different devices (CMOS, power devices, etc) on same wafer. The device structure also shows possibility of stacking more than one layer of devices [73].

[4] Reliability: SOI MOSFETs offer excellent tolerance to radiation effects [71,79]. The incoming particles generate electron-hole pairs in proportion to the device volume exposed to carrier generation, which is 2-3 orders of magnitude smaller than in bulk Si [69].

[5] Short channel effects: The extension of source/drain depletion regions is restricted by the junction size (vertical) and by the dual-gate control (via front-gate and silicon substrate), and thus, SOI devices are more immune to short channel effects and avail better subthreshold slope, when compared with bulk devices [71,80].

[6] Circuit design and processing: Fabrication of CMOS circuits on SOI is much easier than in bulk silicon [73]. The number of fabrication process steps is reduced due to the absence of wells and inter device trenches which were present in bulk design [72,73].

The benefits offered by SOI technology are numerous, but the device suffers from interconnection issues and problems related to self-heating [69]. The other constraints include (i) large series resistance which can be reduced by use of raised source/drain [73] and (ii) the Floating Body Effects (FBEs) [78,81-87] which can be eliminated by use of fully depleted device.

1.4.2.1 Partially and fully depleted SOI MOSFETs

Depending upon the thickness and doping concentration of silicon film, the SOI can be classified [71] as:

(1) Partially Depleted Silicon-on-Insulator (PDSOI): The top film is depleted of carriers while the bottom region remains neutral at zero bias.

(2) Fully Depleted Silicon-on-Insulator (FDSOI): Complete silicon film is depleted of charge carriers at zero bias.

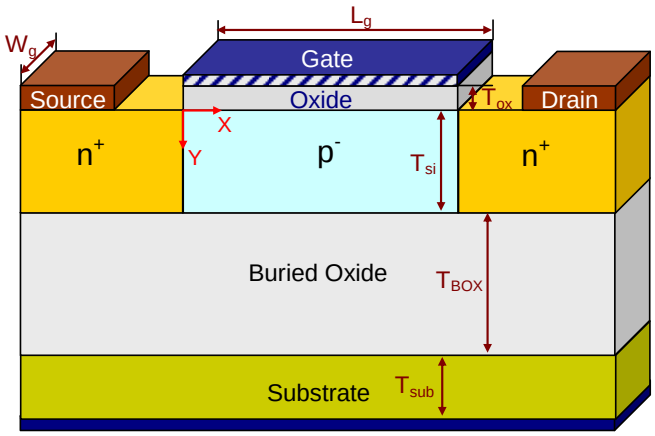

Fig. 1.5 Schematic diagram of an *n*-type Silicon-on-Insulator MOSFET, with a gate length of L_g, gate oxide thickness of T_{ox}, silicon film thickness of T_{si}, buried oxide thickness of T_{BOX}, width of W_g, and substrate below buried oxide with thickness of T_{sub}. The doping of channel region is moderate *p*-type (p^-) and source and drain are heavily doped with *n*-type of impurities (n^+) [69].

Table 1.3 Comparison between Bulk, PDSOI, and FDSOI technologies [69,71].

Parameters	***Bulk***	***PDSOI***	***FDSOI***
S/D resistance	Low	Moderate	High
On-current, I_{on}	High	Moderate	High
Off-current, I_{off}	Moderate	Very Low	Low
Drain induced barrier lowering	Moderate	Low	Very Low
Subthreshold Swing, *S*	High	Moderate	Very Low
FBE, Kink effect, history effect	No	Yes	Conditional
Coupling channel	No	No	Yes

Table 1.3 compares bulk, PDSOI and FDSOI technologies. FDSOI, being completely depleted of carriers exhibit the most interesting properties, such as low electric fields, high transconductance (g_m), excellent immunity to short channel effects, and better subthreshold swing characteristics that nominates it better than other comparable technology [69]. FDSOI can also be utilized for circuit operation with high performance and at relatively lower bias [71]. However, fully depleted transistors are very sensitive to silicon thickness compared to partially depleted MOSFETs, and thus, require efficient control during fabrication to prevent process variations [71].

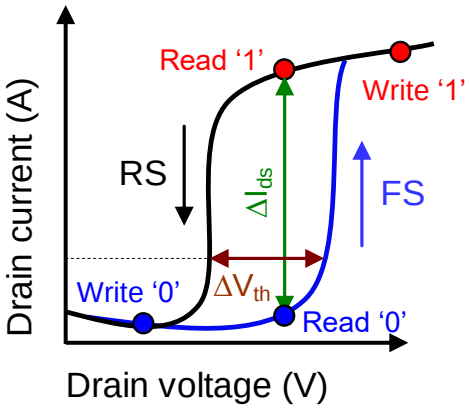

Fig. 1.6 Schematic representation of drain current-voltage characteristics, showing hysteric behaviour with memory window (ΔI_{ds}-ΔV_{th}), depicting write and read mode of operation. ΔV_{th} and ΔI_{ds} demonstrates the shift in threshold voltage, and current levels for same set of bias applied.

In PDSOI devices, a neutral region exists, called as body, and if connected to ground through a body contact, the characteristics of the device is similar to that of a bulk device [71,78]. However, if the body is left electrically floating [80], the charges will be accumulated in the neutral region of semiconductor film. There are various consequences of this charge confinement in the body of PDSOI devices, which are generally attributed to Floating Body Effects (FBEs) [81-87], such as kink effect, hysteresis, negative conductance and transconductance, single transistor latch, bipolar transistor action, premature breakdown. Besides, FDSOI devices are virtually free of FBEs, provided their back interface is not in accumulation [85]. The hysteric behaviour observed in Fig. 1.6 due to FBE shows two threshold voltages for same device and bias parameters. The Reverse Sweep (RS) triggers impact ionization or tunneling that accumulates holes in the silicon film at the floating body contact [78]. The hole accumulation develops a positive body potential that lowers the threshold voltage in comparison the threshold voltage observed in Forward Sweep (FS). Thus, distinct current jumps are observed for the two directions of the voltage sweep, and their separation is governed by the drain-voltage magnitude [69]. The two different thresholds reflect state '1' and '0' of the memory [58]. The other prominent FBE is the kink effect that is symbolized as an abrupt increase in the saturation current of PDSOI when operated in strong inversion [71]. Although FBEs are undesirable for device operation, they are advantageous for memory applications [84].

1.5 Floating Body (FB) for dynamic memory applications

1.5.1 Evolution of FB-DRAM

The operation of FB-DRAM is based on charging and discharging the capacitor associated with floating body [88-127]. The concept of FB-DRAM was implicated by a start-up company, Innovative Silicon in 2003 [88]. The idea was introduced with the name as Zero capacitor RAM (Z-RAM) cell [88-90]. Over the time, the use of PDSOI as Z-RAM cell [89-91] was minimized as the industry was progressing towards thin or ultra-thin FDSOI [92-100]. FDSOI doesn't virtually show FBE due to low barrier between source and channel that lead to rapid recombination of the holes, hence inhibiting hole confinement in the body [71]. However, FDSOI with Ultra-Thin Buried Oxide (UTBOX < 20 nm) along with back bias facilitates the creation of potential well for charge storage [101]. Thus, FDSOI technology has been evolving as capacitorless DRAM, thereby replacing PDSOI based memory devices. The advantage of using FDSOI is the improved SCE due to thinner silicon film, and reduced RDF due to intrinsic channel [69]. The first generation Floating Body Cells (FBCs) were based on impact ionization [89-102] or Gate Induced Drain Leakage (GIDL) [101-111], while the second generation FBCs utilizes bipolar effect [111-122] that showed a higher retention time as compared to the previous methods and also, a low current condition during the erase operation that minimizes the power consumption [58]. Body charge dynamics were also improved due to dynamic coupling between the front and back interfaces in FBCs [123,124]. The devices with trenched body [125] are also utilized to achieve desirable performance in terms of a larger programming window and longer *RT*.

1.5.2 Operating principle as DRAM

1.5.2.1 Write Operation

The hole storage is defined as write '1' or programming operation and is performed using following approaches:

1 **Write through impact ionization** [88-102]: The principle operation of impact ionization is based on the avalanche breakdown, where a high electric field at the channel/drain transition region creates Electron-Hole Pairs (EHPs) on collision with bound electrons [70]. The generated electrons drift towards

positively biased drain while holes are stored at lower potential region, which is the neutral body of the film [78]. Write '1' based on impact ionization is demonstrated in Fig. 1.7. This is a simple and fast carrier injection method. However, the need of high drain bias is a concern [71]. The high electric field may give rise to hot electrons that can inject into gate oxide, thereby damaging the oxide-silicon interface [71]. Thus, reliability of device is an issue associated with write using impact ionization.

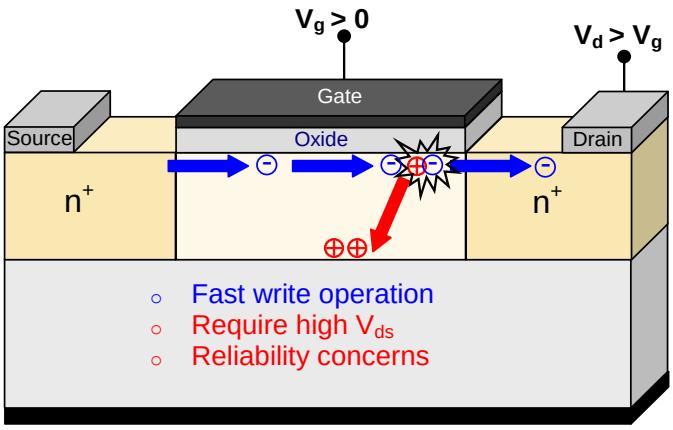

Fig. 1.7 Schematic representing write '1' operation through impact ionization [90]. The electrons accelerated towards drain strike the bonded electrons to generate EHPs. The holes are accumulated at the lower potential region, representing state '1' [88].

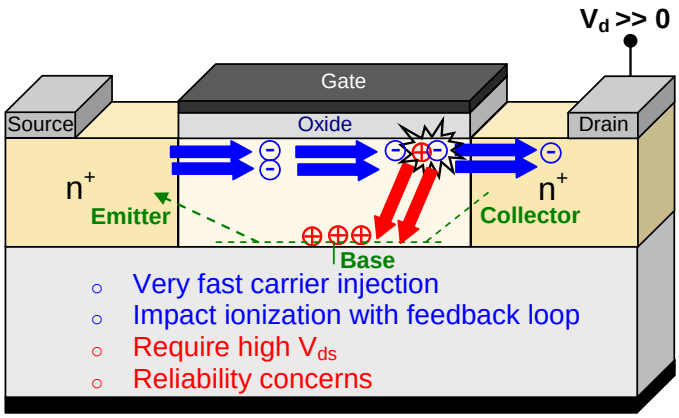

Fig. 1.8 Schematic diagram illustrating program through bipolar action, where impact ionization is enhanced with a feedback mechanism triggered through parasitic bipolar transistor [88].

2 **Write through bipolar action** [112-122]: This method is based on impact ionization along with a feedback loop that generates more electron-hole pairs.

The feedback action is triggered through forward biasing the intrinsic bipolar transistor with *p*-type body, n^+ source and n^+ drain as base, emitter and collector, respectively. The accumulated holes forward bias base-emitter junction, that enhances the channel current ($I_{Bipolar}$ + I_{MOS}). Therefore, more electrons are accelerated towards drain that further trigger generation of EHPs and thus, speed up the write mechanism (Fig. 1.8). Impact ionization enhanced by parasitic bipolar transistor is a very fast write mechanism. However, the use of high drain bias and reliability can be a concern [71].

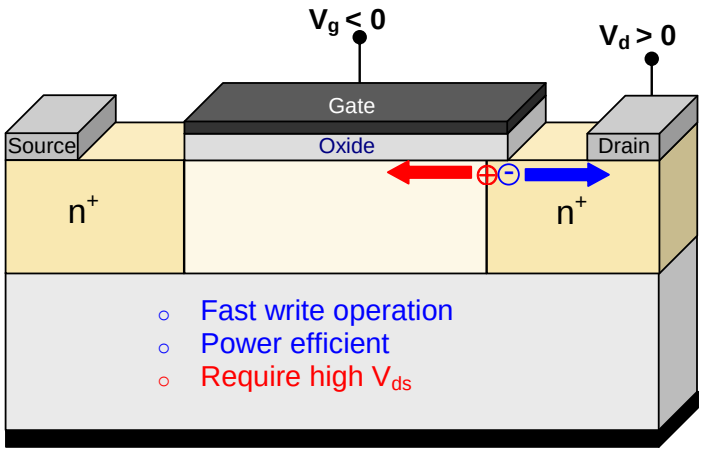

Fig. 1.9 Schematic representation showing write '1' through band-to-band tunneling of electrons from channel region towards drain that generates holes, being confined at the channel region [88].

3 **Write through Band-to-Band Tunneling (BTBT)** [101-111]: Tunneling operation is based on the concept of Zener tunneling, where a negative bias at the gate and a positive bias at drain reverse biases the channel/drain junction [58]. The negative bias at gate with a positive bias at drain creates a large electrostatic potential barrier between the regions that cease the charge transfer through thermionic emission [72] and hence, the transport of charge carriers is possible through tunneling. The electrons tunnel towards the drain, thereby, accumulating holes into the potential well [101]. This tunneling is usually categorized as GIDL [105], and is utilized with gate/drain overlap as illustrated in Fig. 1.9. Although this write mechanism require high drain bias, the write current is smaller compared to previous methods (write through impact ionization and bipolar action) and thus, is a power efficient mechanism with minimized reliability concerns [58].

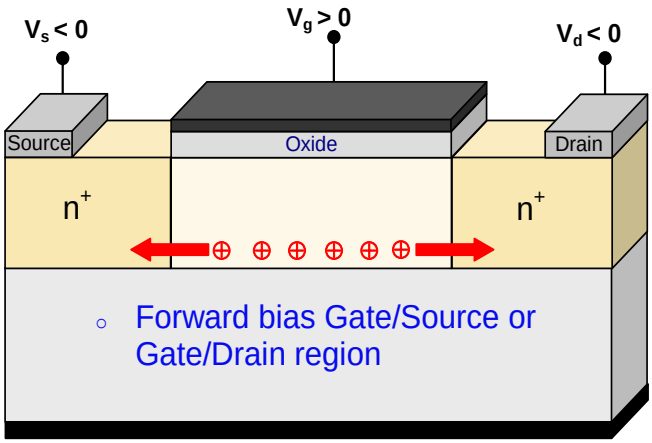

Fig. 1.10 Schematic representation showing write '0' through forward biasing drain/source and channel region [58].

While hole accumulation defines write '1', their depletion constitutes write '0' or erase operation, which requires hole removal from the storage region [58]. The holes are evacuated from the body through forward bias as illustrated in Fig. 1.10. For an *n*MOS, the positive potential at gate removes the holes from the body through capacitive/dynamic coupling [14]. Dynamic coupling between the front and back gate can be understood as the impact of gate bias, where increase in front gate bias increases the body potential above equilibrium and naturally, turns on the junctions. Another way of forward biasing the junction is through application of negative bias at drain/source [58]. The forward biased body and source/drain junction results in the recombination of holes at the heavily doped n^+ region, and thus, holes are depleted from the storage region.

1.5.2.2 Read Operation

The presence of excess holes develops a positive potential at the back. This lowers the threshold voltage of the front gate, and thus, enhance the Read current for state '1' (Fig. 1.11(a)) as compared to state '0' (Fig. 1.11(b)). Thus, two different front gate thresholds are observed based on the presence and absence of the excess charge that distinguishes the two states [101]. Fig. 1.11(c) shows the threshold shift (ΔV_{th}), between the two states that lead to two different current levels (ΔI_{ds}) for same set of bias.

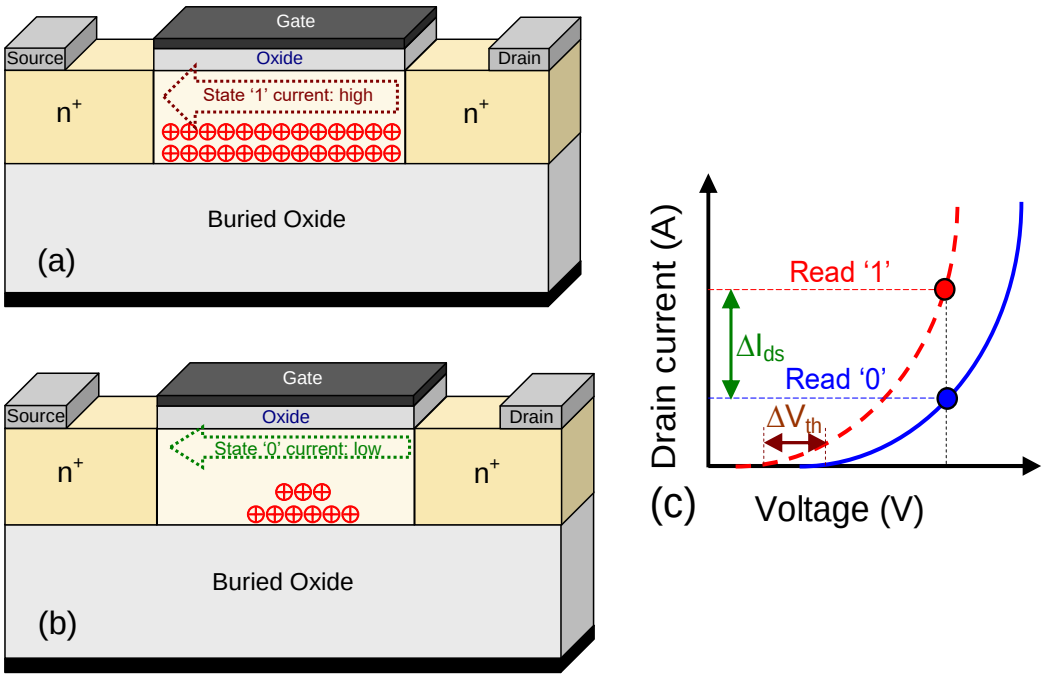

Fig. 1.11 Schematic diagram of a SOI device, illustrating the read mechanism, where presence of excess holes in the neutral body region lead to higher current for (a) state '1', compared to (b) state '0' and the variation is observable in drain current – gate voltage characteristics through (c) shift in threshold voltage (ΔV_{th}) and difference in the current levels (ΔI_{ds}) [58].

1.5.2.3 Hold operation

As mentioned before, the two states are stored through write operation and are distinguished based on the read operation. However, hold operation is performed between write and read operation that determines the retention of a state [101]. The charge sustenance is regulated through hole generation and recombination that is controlled through device architecture, geometry, bias and temperature [88-127]. State '0' is degraded due to thermal generation and BTBT of electrons towards drain/source that generates holes in the potential well [88]. State '1' is disturbed due to decrement in hole concentration in the storage region due to thermal recombination and hole diffusion [88]. Thus, to attain high retention time, which is one of the key metrics defining DRAM performance, the regulation of hole recombination and generation is essential.

MOSFETs have proven to be a promising architecture as a 1T dynamic memory [88-127], achieving retention time that satisfies the 64 ms specification for

standalone DRAM [13] with acceptable read sensitivity. However, the rapidly growing semiconductor industry demands evolution and innovative methodologies to enhance performance with improvement in speed, power and density [18]. Thus, the focus has shifted towards steep switching devices for various applications to facilitate a high speed operation, while maintaining the need for low power and high density.

1.6 Steep Switching Devices

1.6.1 Motivation for steep switching devices

Silicon CMOS technology is the most promising technology that has been continuously evolving with downscaling. However, increase in power dissipation with the scaling of CMOS transistors is one of the basic concerns in semiconductor industry [1,128]. The downscaling of transistor size improves the device functionality, speed, density and cost [1-13,128,129]. It necessitates reduction in the threshold voltage (V_{th}) along with supply voltage (V_{dd}), to maintain a high overdrive voltage (V_{dd} - V_{th}), so as to achieve a high on-current (I_{on}) [130]. However, the reduction in V_{th} increase off-current (I_{off}), and thus, the static power dissipation. The increase in I_{off}, also reduces the switching speed (I_{on}/I_{off}), and hence, degrades the device performance. Therefore, the requirement is to maintain the performance of device in terms of speed, while accomplishing the need of low power and high density [131,132].

The dynamic and static power of the device is proportional to V_{dd} [133-135]. In advanced technology, the static power (~$I_{off} \times V_{dd}$) can be a significant contributor to total power consumption, and even dominant over dynamic power [128,132]. Therefore, computing systems should address both issues related to dynamic, as well as, static power. This can be achieved by reducing the supply voltage and also, increasing the turn-on steepness, being quantified through a parameter, Subthreshold Swing (S), which is defined as the gate voltage required to change the drain current by an order with transistor operation in sub-threshold region, and is evaluated [70,133] as,

$$S = \frac{dV_g}{d\psi_s} \frac{d\psi_s}{d(log_{10} I_D)} \cong \left(1 + \frac{C_d}{C_{ox}}\right) \ln 10 \frac{kT}{q} \quad (1.1)$$

where V_g is the gate voltage, I_d is the drain current, Ψ_s is the surface potential, kT/q is the thermal voltage, which is ~ 60 mV/decade for MOSFET at 300 K and C_d and C_{ox} are the depletion and the oxide capacitances, respectively. $\frac{dV_g}{d\psi_s}$ is the transistor body factor, defined as *m*, and $\frac{d\psi_s}{d(log_{10} I_D)}$ is denoted as *n*, a factor that characterizes the change of the drain current with the surface potential, reflecting the conduction mechanism in the channel. The significance of *S* can be understood through equation [136],

$$S_{avg} \cong \frac{V_{dd}}{\log(I_{on}/I_{off})} \quad (1.2)$$

Equation 1.2 suggests that reduction in supply voltage is possible while maintaining the switching characteristics through reduction in subthreshold slope. However, the switching capability of conventional MOS is restricted by Boltzmann limit (~*kT*/*q*), with subthreshold swing limited to 60 mV/decade at room temperature [133]. However, the steepness can be improved through emerging energy efficient devices [136] with a transport mechanism different from conventional MOS devices [133].

1.6.2 Various steep switching devices

Equation 1.1 demonstrates that the sub-thermal swing (< 60 mV/decade) can be obtained by modifying the factor *m* or *n*. The factor *m* can be altered to a value, lower than 1. This is possible through modification in the electrostatics of gate through Negative Capacitance (NC) effect [137-143], showing *S* of ~13 mV/decade [152] or using electromechanical gates [144-148] with *S* of ~2 mV/decade) [144]. The operation of NCFETs depends on polarization, and therefore, material optimization and its integration are essential [137]. The voltage amplification in NCFETs depends on the viscosity coefficient of ferroelectric. However, a low viscosity coefficient is required for a steep switch, which makes it unsuitable for high speed applications [141]. Also, their practical use in hysteretic [140] or non-hysteretic [143] applications is still under investigation. The electromechanical gates use movable electrodes, where instability points between electrical and mechanical points define transition. Although, devices have experimentally demonstrated a swing < 2 mV/decade, with less temperature

dependency, the issues related to reliability, voltage scaling and the requirement of a controlled environment during packaging, limits its use [128,129].

Another way of reducing *S* is by modulating the factor *n* that can be reduced by utilizing a different carrier injection method. It includes the transport mechanism based on (i) impact ionization, as in Impact Ionization Metal Oxide Semiconductor (IMOS) [149-151] and (ii) quantum mechanical Band-to-Band tunneling (BTBT), as in Tunnel Field Effect Transistor (TFET) [134,135]. Both these mechanisms occur at high electric field. IMOS devices have a very steep transition with experimentally reported value of swing, ~6 mV/decade [150]. However, scaling the operating voltage in IMOS is challenging and also, the device suffer from hot carrier injection that can result into reliability issues [128].

Compared to all other steep switching devices, TFET has larger *S* of ~30 mV/decade [152,153], and one of the main limitations of the device is low on-current (I_{on}) [134]. However, various techniques that include modification in architecture and material have been proposed to improve its driving capability and are still under investigation by various research groups for further improvement [128,134]. The advantage of TFET device is low off-current I_{off} (except electromechanical switches) and highest potential for voltage scaling amongst CMOS compatible devices [13,129]. Before discussing the applicability of TFET as memory, understanding its operating principle, limitations and benefits is essential.

1.6.3 Tunnel Field Effect Transistor (TFET)

TFET is a gated *p-i-n* diode with heavily doped *p* and *n* on either side of the intrinsic region (Fig. 1.12). The evolution of tunneling and the associated devices is summarized in Table 1.4 [154-172]. The concept of gated *p-i-n* structure was proposed by Quinn *et al.* [173], and its study was further extended by Banerjee *et al.* [174] with three terminal silicon based TFET. The prospect of device scalability was studied first by Takeda *et al.* [175]. Further, surface tunneling transistors were fabricated with III-V material [176]. In 1996, a forward biased silicon tunneling device was proposed by Koga and Torumi [177]. In 2000, reverse biased based tunneling device were proposed in vertical structure [178],

and in 2006, Aydin *et al.* [179] proposed a lateral tunneling based device with operating principle similar to TFET, but without intrinsic region. Recently, TFET has emerged as a promising candidate with different geometry and materials [152, 180-184].

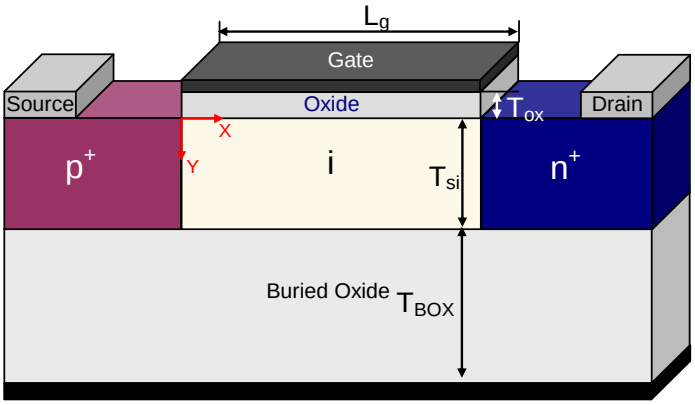

Fig. 1.12 Schematic diagram of a TFET, with a gate length of L_g, gate oxide thickness of T_{ox}, silicon film thickness of T_{si}, buried oxide thickness of T_{BOX}. The device has an intrinsic channel doping, while source and drain are heavily doped with *p* and *n*-type of impurities, respectively [134].

1.6.3.1 Operation of TFET

The operation of TFET is based on interband tunneling with the concept based on Zener tunneling, where carriers are transferred from a heavily doped p^+/n^+ regions [167]. TFET is an ambipolar device where electron dominant conduction signifies *n*-type TFET, while that hole dominant, as the *p*-type TFET. Fig. 1.13(a) shows a qualitative comparison of a MOSFET and a Tunnel FET. At moderate performance requirements (@ V_{g1}), TFETs offer improved I_{on}/I_{off}, and also superior performance (higher I_{on} at the same voltage) or power savings at the same performance (lower voltage for the same I_{on}) over MOSFETs [14]. However, when a much higher performance (@ V_{g2}) is required, a MOSFET is the better solution. Other than low I_{on}, the performance of TFET is also affected by the traps that result into threshold shift and subthreshold slope degradation [128,135]. The low I_{off} offers better energy efficiency at lower or moderate performance level. The steeper swing for TFET demonstrates operation at lower supply voltage for same I_{on}/I_{off}, as illustrated in Fig. 1.13(b) that nominates TFET as an energy efficient device that could possibly replace MOSFETs without performance loss [134].

Table 1.4 Selected history of tunneling and devices [154].

Advances in tunneling and devices	***Investigators***	***Year***
• Observation of field-emission from metals	Lilienfield [155]	1922
• Explaination of field-emission	Fowler and Nordheim [156]	1928
• Theory of interband tunneling in solids	Zener [157]	1934
• Field emission microscope	Muller [158]	1937
• Observation of Zener breakdown	Chynoweth and Mckay [159]	1957
• Tunneling in degenerate *pn* junctions	Esaki [160]	1958
• Extension of Zener's theory to tunnel diodes	Keldysh, Kane, *et al.*[161-162]	1958-1961
• Measurement of energy gap of superconductors	Giaver [163]	1960
• Pertubation treatment of tunneling	Bardeen [164]	1961
• Tunneling of Cooper particles	Josephson [165]	1962
• Floating gate memory	Kahng and Sze [166]	1967
• Double barrier Resonant Tunneling Diode	Chang, Esaki, and Tsu [167]	1974
• Scanning Tunneling Microscope	Binning [168]	1981
• Resonant Tunneling Transistor	Capasso, Frensley, Reed [169]	1986
• RTD SRAM	Van der Wagt [170]	1998
• Tunnel FET	Bhuwalka, *et al*. [171]	2004
• Graphene-Insulator-Graphene tunnel junctions	Feenstra, Jena [172]	2012

Fig 1.14 indicates the operating principle of an *n*-type TFET, where a positive bias at the gate modifies band energy such that the valence band of source region lie adjacent to conduction band of channel region [134]. The narrow barrier between the two regions lead to tunneling of electron, as demonstrated in the figure. The energy bands are controlled by the applied gate bias and depend on the transmission probability. Assuming a triangular potential barrier and approximated using Wentzel–Kramer–Brillouin (WKB) equation [133], transmission probability (T_{WKB}) can be expressed as,

$$T_{WKB} \approx exp\ \left(-\frac{4\lambda\sqrt{2m^{*}E_{g}^{3}}}{3q\hbar\left(E_{g}+\Delta\Phi\right)}\right) \tag{1.3}$$

where m^* is the effective mass of the carrier and E_{g} is the bandgap. λ is the screening tunnelling length that describes the spatial extent of the transition region at the source–channel interface. $\Delta\Phi$ is the energy difference between the valence band of source and the conduction band of channel (Fig. 1.13).

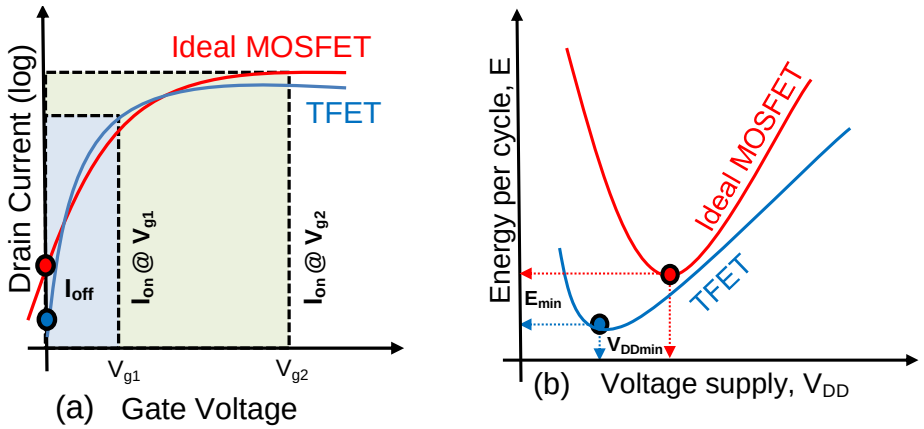

Fig. 1.13 (a) Drain current-gate voltage characteristics of an ideal MOSFET and TFET with a qualitative comparison. I_{off} of TFET < MOS, I_{on} of TFET > MOS at lower gate voltage (V_{g1}), while I_{on} of TFET < MOS at higher gate voltage (V_{g2}) [147]. (b) Qualitative comparison of the minimum switching energy, E_{min}, and the corresponding voltage supply, V_{DDmin}, for a TFET and the ideal MOSFET at the same I_{on}/I_{off} [134].

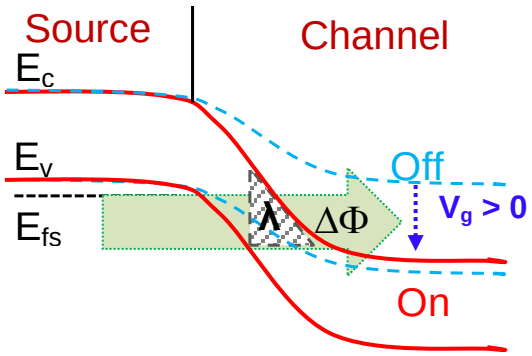

Fig. 1.14 Schematic energy band profile for the off (dashed) and on state in an *n*-type TFET [134]. E_c, E_v, E_{fs} represents the conduction band energy, valence band energy, and the fermi-level of the semiconductor, respectively.

1.6.3.2 Improving performance of TFET

At a constant drain voltage, V_d, the increase in V_g reduces λ and increases the energy window ($\Delta\Phi$), thereby increasing the tunneling probability [128]. Equation 1.3 demonstrates that lower values of m^*, λ and E_g will enhance the tunneling probability. While E_g and m^* are material properties, λ depends on the electric field at the tunneling junction that is regulated by device geometry, doping, dimensions and gate capacitance [184,185]. Thus, enhancement in tunneling current is feasible through the following options:

(i) a high-κ gate dielectric with an equivalent oxide thickness as low as possible [128,135],

(ii) a thinner body [136,184],

(iii) abruptness at the tunnel junction [134],

(iv) high source doping with intrinsic channel [134],

(v) a shorter intrinsic region [134],

(vi) alignment of gate oxide with the intrinsic region, and

(vii) multiple gate operation [134, 136].

Other than the above, decreasing m^* and E_s through use of materials with small m^* and different band edge alignments, like strained Silicon (Si), Germanium (Ge), SiGe materials, III-V materials, carbon nanotubes and graphene ribbons, are viable option to boost the driving capability of TFET. Other techniques adopted to further enhance the on-current, includes insertion of tunnel dielectric, spacer engineering, use of dopant pockets, and alternative gate positioning [128,134,186-189]. The low off-current, which is one of the prominent features of TFET, when incorporated with higher I_{on} show steep switching.

1.6.3.4 Applications of TFET

Further studies, reflect the advantage of TFET in terms of less sensitivity to temperature [171,190-193] and random dopant fluctuations [128,133], and also, better potential to scale compared to conventional MOS transistor. Based on the performance metrics, the potential of device was further explored by circuit designers for digital as well as analog applications [186, 194-196]. Although less explored, TFETs can pave a way forward in analog design that requires high gain at low voltage and less dependency on temperature [191]. TFET have shown an improvement in the transconductor efficiency [128], ratio between transconductance (g_m) and current (I_d), i.e. g_m/I_d, and has overcome the fundamental limit of g_m/I_d imposed by conventional MOS devices. g_m/I_d signifies the efficiency of the MOS transistor to translate given current (power) into an equivalent transconductance. The other analog applications that include several building blocks such as Operational Transconductance Amplifiers (OTAs), current mirrors, and track-and-hold circuits have been examined by Sedighi *et al.* [196], which confirms enhanced power and gain than conventional MOS.

Although TFETs have shown promising results as “Beyond-CMOS” options, there are various other applications for which feasibility assessment is required. Recent studies have paved a way towards utility of TFET for dynamic memory applications [197-200]. The topologies suggested either used an asymmetric architecture [197,198] or a precise *p*-type doping [199,200] to create the electrostatically induced potential well. Although, previous works [197-199] provided opportunity for DRAM, the retention time, which is the most crucial metric for memory application is limited. Moreover, the physical insights and impact of various parameters affecting their performance require analysis to be employed in real-time applications. The research work presented in the thesis provides insights into the understanding of the performance and behaviour of TFET for application as memory device, by means of comprehensive physical device simulations [201]. The research shows comparative analysis with other TFET based DRAM [197-200] and also, similar architectures [201-215] that highlights the limitations and the benefits of the proposed architectures.

1.7 Organisation of thesis

The research work in this thesis has focused on evaluation and feasibility of TFET for application as dynamic memory. A systematic attempt has been made to present insights into device physics and operation for functionality as DRAM. The work showcases various attributes to enhance the performance of TFET for dynamic memory applications. The significance of device architecture, geometry, parameters, biases, temperature has been evaluated. The optimal design have shown improved retention, scalability, read sensitivity, reliability at low power, which is a significant improvement over the previous proposed TFET based DRAM [197-199]. The thesis presents advancement in TFET based DRAM while progressing from an UTBOX TFET to dual gate, and finally, to a planar tri-gate topology. The concept, design and operation of various architectures proposed provide valuable viewpoints for tunneling based DRAM. The thesis present a systematic methodology to enhance various DRAM metrics with prime focus on improving retention characteristics, and further progressing towards scalability, speed, power and sense margin.

Following is the chapter wise organization of the thesis:

Chapter 1 introduces dynamic memory as an essential component for semiconductor industry. The investigation with emphasis on its requirement in the past and the prospects in the future, reflects the need for replacing conventional DRAM (1T-1C) with capacitorless DRAM. Further, the quest for low power dynamic memory demonstrates use of steep switching devices as a substitute to conventional MOS devices. Among various steep switching devices, TFET has shown promising results and potential as energy efficient device, and therefore, our thesis work explores tunneling based transistors for capacitorless dynamic memory applications.

Chapter 2 demonstrates various limitations and the possible outcomes for utilizing TFET as DRAM. The previous work on TFET presented opportunity as dynamic memory [197-199], but low retention (< 64 ms, target specified by ITRS [13]) was a critical issue. Therefore, a topology with misaligned architecture has been proposed. The work focuses on creating a profound potential well, achieved through appropriate use of back gate workfunction, position, and bias to improve retention characteristics. Thus, chapter 2 reports on a misaligned structure, traditionally considered detrimental for device design, as an advantage for charge retention in tunneling based transistors.

Chapter 3 extends the idea of a misaligned architecture to further improve retention time as well as scalability of the device, possible through optimization in three steps. This includes (i) misalignment, (ii) lateral spacing (L_{gap}) between the gates and an underlap (L_{un}) between the back gate and drain region shows better scaling and retention perspective. The work presents a systematic analysis of individual lengths (length of front and back gates, $L_{gap,}$ L_{un}) as well as total length that can serve as a guideline for future design.

Chapter 4 demonstrates the significance of device architecture, design, and bias through a twin gate topology. Although, the operating principle is similar to misaligned DG TFET, the device is also examined for embedded applications. The assessment of twin gate focuses on speed and power, while maintaining with

high retention at scaled devices. The work highlights different operating principle compared to other devices with similar topology (p^+-i-n^+ structure), such as Field Effect Diode (FED). The optimal design exhibit enhanced retention, speed, reliability and operation at higher temperature (> 85 °C) with low power consumption. Other than these DRAM metrics, sense margin is also a crucial metric that defines memory read sensitivity. However, retention time and sense margin show trade-off and thus, the device architecture is further modified to preserve both the important DRAM metric.

Chapter 5 explores a planar tri-gate TFET to enhance the sense margin and scalability and thereby, overcome the critical bottleneck faced by TFET based dynamic memories. Insights into device operation demonstrate that the choice of appropriate architecture and biases not only limit the trade-off between sense margin and retention time, but also result in improved scalability of drain voltage and total length. Along with sense margin and retention time, current ratio of device is also an important DRAM metric. The work further demonstrates two composite metrics (*M1* and *M2*), namely, product of (i) Sense Margin (*SM*) and Retention Time (*RT*) i.e. $M1 = SM \times RT$, and (ii) Sense Margin and Current Ratio (CR) i.e. $M2 = SM \times CR$, where their values show optimal performance. The planar tri-gate shows advancement in TFET based dynamic memory with high charge retention and read sensitivity, improved scalability and operation at lower drain bias and thus, as a low power dynamic memory.

Chapter 6 summarizes the key results and contributions of this dissertation and proposes the scope for future work.

1.8 References

[1] Borkar S., (2001) Low power design challenges for the decade, in Proceedings of ASP-DAC, pp. 293-296.

[2] Moore G. E., (1965) Cramming More Components onto Integrated Circuits, Electronics, 38, 114-116.

[3] Moore G. E., (1975) Progress in digital integrated electronics, IEEE Electron Devices Meeting, 21, 11-13.

[4] Dennard R. H., Gaensslen F. H., Rideout V. L., Bassous E., and LeBlanc A. R., (1974) Design of Ion-Implanted Small MOSFET's Dimensions with Very Small Physical Dimensions, IEEE Journal of Solid-State Circuits, 9, 256–268.

[5] Schaller R. R., (1997) Moore's law: past, present and future, IEEE spectrum, 34, 52-59.

[6] Mollick E., (2006) Establishing Moore's law, IEEE Annals of the History of Computing, 28, 62-75.

[7] Bohr M., (2007) A 30 year retrospective on Dennard's MOSFET scaling paper, IEEE Solid-State Circuits Society Newsletter, 12, 11-13.

[8] Wong H.-S. P., (2002) Beyond the conventional transistor, IBM Journal of Research and Development, 46, 133-168.

[9] Chen T. C., (2006) Where CMOS is going: trendy hype vs. real technology, IEEE Solid-State Circuits Conference, pp. 1-18.

[10] Antoniadis D. A., Aberg I., Chleirigh C. N., Nayfeh O. M., Khakifirooz A., and Hoyt J. L., (2006) Continuous MOSFET performance increase with device scaling: The role of strain and channel material innovations, IBM Journal of Research and Development, 50, 363-376.

[11] Waldrop M. M., (2016) More than moore. Nature, 530, 144-148.

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

[13] ITRS International Technology Working Groups. International Technology Roadmap for Semiconductors, http://www.itrs.net (2015).

[14] Joodaki, M., (2012) Selected advances in nanoelectronic devices: logic, memory and RF, Springer Science and Business Media, Heidelberg (ISBN-978-3-642-31349-3).

[15] Balestra, F., Kilchytska, V., and Flandre, D., (2014) Functional nanomaterials and devices for electronics, sensors and energy harvesting, Springer International Publishing, Switzerland (ISBN 978-3-319-08803-7).

[16] Kurinec, S.K., and Iniewski, K., (2013) Nanoscale semiconductor memories: Technology and applications, CRC Press, Taylor & Francis Group, Florida (ISBN-978-1-4665-6061-1).

[17] Zhang, K., (2009) Embedded memories for nano-scale VLSIs, Springer Science and Business Media, Oregon (ISBN 978-0-387-88496-7).

[18] Lee S. H., (2016) Technology scaling challenges and opportunities of memory devices, IEEE Electron Devices Meeting, pp. 1-8.

[19] Keeth, B., Baker, J. R., Jacob, B., Lin, F., (2007) DRAM circuit design: fundamental and high-speed topics 2nd ed. John Wiley & Sons Inc., New Jersey (ISBN- 978-0-4701-8475-2).

[20] Bhati I., Chang M. T., Chishti Z., Lu S. L., and Jacob B., (2016) DRAM refresh mechanisms, penalties, and trade-offs, IEEE Transactions on Computers, 65, 108-121.

[21] Wulf W. A., and McKee S. A., (1995) Hitting the memory wall: implications of the obvious, ACM SIGARCH computer architecture news, 23, 20-24.

[22] Mutlu O., (2013) Memory scaling: A systems architecture perspective. IEEE International Memory Workshop, pp. 21-25.

[23] Joao J. A., Suleman M. A., Mutlu O., and Patt Y. N., (2012) Bottleneck identification and scheduling in multithreaded applications. ACM SIGPLAN Notices, 47, 223-234.

[24] Malladi K. T., Shaeffer I., Gopalakrishnan L., Lo D., Lee B.C., and Horowitz M., (2012) Rethinking DRAM power modes for energy proportionality, In Proceedings of the 45th Annual IEEE/ACM International Symposium on Microarchitecture, 131-142.

[25] Fujisaki Y., (2013) Review of emerging new solid-state non-volatile memories, Japanese Journal of Applied Physics, 52, 040001.

[26] Freitas R. F., and Wilcke W. W., (2008) Storage-class memory: The next storage system technology, IBM Journal of Research and Development, 52, 439-447.

[27] Burr G. W., Kurdi B. N., Scott J. C., Lam C. H., Gopalakrishnan K., and Shenoy R. S., (2008) Overview of candidate device technologies for storage-class memory, IBM Journal of Research and Development, 52, 449-464.

[28] Brewer J. E., Zhirnov V. V., and Hutchby, J. A., (2005) Memory technology for post CMOS era, IEEE Circuits and Devices Magazine, 21, 13-20.

[29] Coughlin T., (2014) Thanks for the memories: Nonvolatile memory technology enables the age of spin [The art of storage], IEEE Consumer Electronics Magazine, 3, 85-92.

[30] Adee S., (2009) Thanks for the memories. IEEE Spectrum, 46.

[31] Endoh T., Koike H., Ikeda S., Hanyu T. and Ohno H., (2016) An Overview of Nonvolatile emerging memories—Spintronics for working memories, IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 6, 109-119.

[32] Chen A., (2016) A review of emerging non-volatile memory (NVM) technologies and applications, Solid-State Electronics, 125, 25-38.

[33] Sun G., Zhao J., Poremba M., Xu, C., and Xie, Y., (2017) Memory that Never Forgets: Emerging Non-volatile Memory and the implication for architecture design, National Science Review, 1-16.

[34] Kim H., Hwang S., Lee J. H., and Park B. G., (2017) Combination of volatile and non-volatile functions in a single memory cell and its scalability, Japanese Journal of Applied Physics, 56, 04CE06.

[35] Kim K., and Lee S. Y., (2007) Memory technology in the future, Microelectronic Engineering, 84, 1976-1981.

[36] Vuong H., and Q. Technologies, (2013) Mobile memory technology roadmap for mobile devices, Today & Tomorrow.

[37] Al-Fuqaha A., Guizani M., Mohammadi M., Aledhari M., and Ayyash M., (2015) Internet of things: A survey on enabling technologies, protocols, and applications, IEEE Communications Surveys & Tutorials, 17, 2347-2376.

[38] E. Hannigan, McKinsey on Semiconductors, fifth edition, McKinsey & Company, (2015).

[39] Ousterhout J., Agrawal P., Erickson D., Kozyrakis C., Leverich J., Mazières D., Mitra S., Narayanan A., Parulkar G., Rosenblum M., and Rumble S. M., (2010) The case for RAMClouds: scalable high-performance storage entirely in DRAM, ACM SIGOPS Operating Systems Review, 43, 92-105.

[40] Zhang H., Chen G., Ooi B. C., Tan K. L. and Zhang M., (2015) In-memory big data management and processing: A survey, IEEE Transactions on Knowledge and Data Engineering, 27, 1920-1948.

[41] Greenberg A., Hamilton J., Maltz D. A., and Patel P., (2008) The cost of a cloud: research problems in data center networks, ACM SIGCOMM computer communication review, 39, 68-73.

[42] Qian C., Huang L., Xie P., Xiao N., and Wang Z., (2015) Efficient data management on 3D stacked memory for big data applications, IEEE 10th International Design & Test Symposium, pp. 84-89.

[43] Chang J., Chang L., Dupret A., Kim C., Hamzaoglu F., and Yoshikawa T., (2015) F2: Memory trends: From big data to wearable devices, IEEE Solid-State Circuits Conference, pp. 498-499.

[44] Ghosh S., (2014) Tutorial t6b: Embedded memory design for future technologies: Challenges and solutions, 27th International Conference on VLSI Design and 13th International Conference on Embedded Systems, pp. 14-15.

[45] Itoh K., (2011) Embedded memories: Progress and a look into the future, IEEE Design & Test of Computers, 28, 10-13.

[46] Kim K., (2005) Technology for sub-50nm DRAM and NAND flash manufacturing, IEEE Electron Devices Meeting, pp. 323-326.

[47] Micron Technology, Inc. 2015 Winter Analyst Conference

[48] World Semiconductor Trade Statistics organization, https://www.wsts.org/

[49] Hassan H., Vijaykumar N., Khan S., Ghose S., Chang K., Pekhimenko G., Lee D., Ergin O. and Mutlu O., (2017) SoftMC: A flexible and practical open-source infrastructure for enabling experimental DRAM studies, IEEE International Symposium on High Performance Computer Architecture, pp. 241-252.

[50] Itoh K., (2008) The history of DRAM circuit designs–at the forefront of DRAM development, IEEE Solid-State Circuits Society Newsletter, 13, 27-31.

[51] Kim K., (2000) Perspectives on giga-bit scaled DRAM technology generation, Microelectronics reliability, 40, 191-206.

[52] Itoh K., Nakagome Y., Kimura S. I., and Watanabe T., (1997) Limitations and challenges of multigigabit DRAM chip design, IEEE Journal of Solid-State Circuits, 32, 624-634.

[53] Kim K., Hwang C. G., and Lee J. G., (1998) DRAM technology perspective for gigabit era, IEEE Transactions on Electron Devices, 45, 598-608.

[54] Itoh K., Sasaki K., and Nakagome Y., (1995) Trends in low-power RAM circuit technologies, Proceedings of the IEEE, 83, 524-543.

[55] Itoh K., Watanabe T., Kimura S. I., and Sakata, T., (2000) Reviews and prospects of high-density RAM technology, International Semiconductor Conference, pp. 13–22.

[56] Comfort J. H., (1999) DRAM Technology: outlook and challenges, IEEE International Conference on VLSI and CAD, 99, pp. 182–186.

[57] Dennard R. H., (1968) Field-effect transistor memory, U.S. Patent 3387286.

[58] Ohsawa, T., Hamamoto, T., (2012) A Novel Capacitor-less DRAM: Floating Body Cell, CRC Press, Taylor & Francis Group, Florida, (ISBN: 978-9-81430-308-8).

[59] Sharroush S. M., (2016) Performance optimization of 1T-1C DRAMs: A quantitative study, Microelectronics Journal, 52, 147-164.

[60] Faul J. W., and Henke, D., (2005) Transistor challenges–A DRAM perspective. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 237, 228-234.

[61] Sunami H., (2008) The role of the trench capacitor in DRAM innovation, IEEE Solid-State Circuits Society Newsletter, 13, 42-44.

[62] Koyanagi M., and Hashimoto N., (1979) Novel high density, stacked capacitor MOS RAM**,** IEEE International Electron Devices Meeting, 24, pp. 348 - 351.

[63] Nakagome Y., Aoki M., Ikenaga S., Horiguchi M., Kimura S., Kawamoto Y., and Itoh K., (1988) The impact of data-line interference noise on DRAM scaling, IEEE Journal of Solid-State Circuits, 23, 1120-1127.

[64] Redeker M., Cockburn B. F., and Elliott D. G., (2002) An investigation into crosstalk noise in DRAM structures, IEEE International Workshop on Memory Technology, Design and Testing, pp. 123-129.

[65] Schroeder B., Pinheiro E., and Weber W. D., (2009) DRAM errors in the wild: a large-scale field study, ACM SIGMETRICS Performance Evaluation Review, 37, 193-204.

[66] Hiraiwa A., Ogasawara M., Natsuaki N., Itoh Y., and Iwai, H., (1998) Local-field-enhancement model of DRAM retention failure, IEEE Electron Devices Meeting, pp. 157–160.

[67] Hamamoto T., Sugiura S., and Sawada S., (1998) On the retention time distribution of dynamic random access memory (DRAM), IEEE Transaction on Electron Devices, 45, 1300–1309.

[68] Tack M. R., Gao M., Claeys C. L., Declerck G. J., (1990) The multistable charge-controlled memory effect in SOI MOS transistors at low temperatures, IEEE Transaction on Electron Devices, 31, 1373–1382.

[69] Cristoloveanu, S., and Lee, S., (1995) Electrical characterization of silicon-on-insulator materials and devices, The Kluwer International Series in Engineering and Computer Science, (ISBN 978-0-7923-9548-5).

[70] Streetman, B. G., and Banerjee, S. K., (2009) Solid State Electronic Devices, sixth ed. Pearson Education Inc., New Jersey 07458, (ISBN-978-81 -203-3020-7).

[71] Colinge, J.-P., (1997) Silicon-on-Insulator Technology: Materials to VLSI, third ed. Springer Science Business Media, New York, (ISBN-978-1-4613-4795-8).

[72] Taur, Y., and Ning, T. H., (2009) Fundamentals of Modern VLSI Devices, second ed. Cambridge University Press, New York, (ISBN-978-0-521-83294-6).

[73] Kononchuk, O., and Nguyen, B.-Y., (2014) Silicon-on-Insulator Technology: Manufacture and applications, first ed. Woodhead Publishing, Elsevier, Oxford, (ISBN-978-0-857-09526-8).

[74] Yan R. H., Ourmazd A., and Lee K. F., (1992) Scaling the Si MOSFET: From bulk to SOI to bulk, IEEE Transaction on Electron Devices, 39, 1704–1710.

[75] Colinge J.-P., (1989) Thin-film SOI technology: the solution to many submicron CMOS problems, IEEE Electron Devices Meeting, pp. 817–820.

[76] Mao B. Y., Sundaresan R., Chen C. E. D., Matloubian M., and Pollack G., (1988) Characteristics of CMOS Devices in oxygen-implanted Silicon-on-Insulator Structures., IEEE Transaction on Electron Devices, 35, 629–633.

[77] Nowak E. D., Ding L., Loh Y. T. and Hu, C., (1994) Speed, power, and yield comparison of thin bonded SOI versus bulk CMOS technologies, IEEE International SOI Conference, pp. 41-42.

[78] Choi J.-Y., and Fossum J. G., (1990) Analysis and control of BJT latch in fully depleted floating-body submicron SOI MOSFETs, IEEE SOS/SOI Technology Conference. Proceedings, pp. 21-22.

[79] Chan M., Yuen S. S., Ma Z., and Hui K. Y., (1994) Comparison of ESD Protection Capability of SOI and BULK CMOS Output Buffers, IEEE International Reliability Physics Symposium, pp. 292–298.

[80] Su L. T., Jacobs J. B., Chung J. E., and Antoniadis D. A., (1994) Deep-Submicrometer Channel Design in Silicon-on-insulator (SOI) MOSFET's, IEEE Electron Device Lett., 15, 183–185.

[81] Lu P.-F., Chuang C.-T., Ji J., Wagner L. F., Hsieh C.-M., Kuang J. B., Hsu L. L.-C., Pelella M. M., Jr., Chu S.-F. S., and Anderson C. J., (1997) Floating-body effects in partially depleted SOI CMOS circuits, IEEE Journal of Solid-State Circuits, 32, 1241–1253.

[82] Valdinoci M., Colalongo L., Baccarani G., Fortunate G., Pecora A., and Policicchio I., (1997) Floating body effects in polysilicon thin-film transistors, IEEE Transaction on Electron Devices, 44, 2234–2241.

[83] Tseng Y.-C., Huang W. M., Ikegami B., Diaz D. C., Ford J. M., and Woo J. C. S., (1997) Local floating body effect in body-grounded SOI nMOSFETs, IEEE International SOI Conference Proceedings, pp. 128–129.

[84] Krishnan S., Fossum J. G., Yeh P. C., Faynot O., Cristoloveanu S., Gautier J., and Gautie J., (1995) Floating-body kinks and dynamic effects in fully depleted SOI MOSFETs, IEEE International SOI Conference Proceedings, pp. 10–11.

[85] Fossum J. G., and Lu Z., (2011) Anomalous floating-body effects in SOI MOSFETs: Low-voltage CMOS?, IEEE International SOI Conference Proceedings, pp. 4–5.

[86] Suh D., and Fossum J. G., (1994) Dynamic floating-body instabilities in partially depleted SOI CMOS circuits, IEEE Electron Devices Meeting, pp. 661-664.

[87] Eminente S., Cristoloveanu S., Clerc R., Ohata A., and Ghibaudo G., (2007) Ultra-thin fully-depleted SOI MOSFETs: Special charge properties and coupling effects, Solid-State Electronics, 51, 239-244.

[88] Moore S. K., (2007) Masters of memory, IEEE spectrum, 44, 45-49.

[89] Okhonin S., Nagoga M., Sallese J. M., and Fazan, P., (2001) A SOI capacitor-less 1T-DRAM concept, IEEE International SOI Conference, pp. 153-154.

[90] Fazan P. C., Okhonin S., Nagoga M., and Sallese J. M., (2002) A simple 1-transistor capacitor-less memory cell for high performance embedded DRAMs, IEEE Custom Integrated Circuits Conference, pp. 99-102.

[91] Giusi G., Alam M. A., Crupi F., and Pierro S., (2010) Bipolar mode operation and scalability of double-gate capacitorless 1T-DRAM cells, IEEE Transactions on Electron Devices, 57, 1743-1750.

[92] Kuo C., King T. J., and Hu C., (2002) A capacitorless double-gate DRAM cell, IEEE Electron Device Letters, 23, 345-347.

[93] Kuo C., King T. J., and Hu C., (2003) A capacitorless double gate DRAM technology for sub-100-nm embedded and stand-alone memory applications, IEEE Transactions on Electron Devices, 50, 2408-2416.

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

[95] Song K. W., Jeong H., Lee J. W., Hong S. I., Tak N. K., Kim Y. T., Choi Y. L., Joo H. S., Kim S. H., Song H. J., and Oh Y. C., (2008) 55 nm capacitor-less 1T DRAM cell transistor with non-overlap structure, IEEE Electron Devices Meeting, pp. 1-4.

[96] Almeida L. M., Aoulaiche M., Sasaki K. R. A., Nicoletti T., de Andrade M. G. C., Collaert N., Simoen E., Claeys C., Martino J. A., and Jurczak M., (2012) Comparison between low and high read bias in FB-RAM on UTBOX FDSOI devices, IEEE International Conference on Ultimate Integration on Silicon, pp. 61-64.

[97] Ranica R., Villaret A., Fenouillet-Beranger C., Malinge, P., Mazoyer P., Masson P., Delille D., Charbuillet C., Candelier P., and Skotnicki T., (2004) A capacitor-less DRAM cell on 75nm gate length, 16 nm thin fully depleted SOI device for high density embedded memories, IEEE Electron Devices Meeting, pp. 277-280.

[98] Ranica R., Villaret A., Malinge P., Mazoyer P., Lenoble D., Candelier P., Jacquet F., Masson P., Bouchakour R., Fournel R., and Schoellkopf J. P., (2004) A one transistor cell on bulk substrate (1T-Bulk) for low-cost and high density eDRAM, IEEE Symposium on VLSI Technology, pp. 128-129.

[99] Hamamoto T., and Ohsawa T., (2009) Overview and future challenges of floating body RAM (FBRAM) technology for 32nm technology node and beyond. Solid-State Electronics, 53, 676-683.

[100] Hubert A., Bawedin M., Guegan, G., Ernst T., Faynot O., and Cristoloveanu S., (2011) SOI 1T-DRAM cells with variable channel length and thickness: Experimental comparison of programming mechanisms. Solid-State Electronics, 65, 256-262.

[101] Rodriguez N., Gamiz F., and Cristoloveanu S., (2010) A-RAM memory cell: Concept and operation, IEEE Electron Device Letters, 31, 972-974.

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

[103] Han J. W., Ryu S. W., Choi S. J., and Choi Y. K., (2009) Gate-induced drain-leakage (GIDL) programming method for soft-programming-free operation in unified RAM (URAM), IEEE Electron Device Letters, 30, 189-191.

[104] Sasaki K. R. A., Almeida L. M., Martino J. A., Aoulaiche M., Simoen E., and Claeys C., (2012) Temperature influence on UTBOX 1T-DRAM using GIDL for writing operation. IEEE Caribbean Conference on Devices, Circuits and Systems, pp. 1-4.

[105] Sasaki K. R. A., Nicoletti T., Almeida L. M., dos Santos S. D., Nissimoff A., Aoulaiche M., Simoen E., Claeys C., and Martino J. A., (2014) Improved retention times in UTBOX nMOSFETs for 1T-DRAM applications, Solid-State Electronics, 97, 30-37.

[106] Yoshida E., and Tanaka T., (2003) A design of a capacitorless 1T-DRAM cell using gate-induced drain leakage (GIDL) current for low-power and high-speed embedded memory, IEEE Electron Devices Meeting, pp. 37-6.

[107] Aoulaiche M., Collaert N., Degraeve R., Lu Z., De Wachter B., Groeseneken G., Jurczak M., and Altimime L., (2010) BJT-mode endurance on a 1T-RAM bulk FinFET device, IEEE Electron Device Letters, 31, 1380-1382.

[108] Lu Z., Fossum J. G., and Zhou Z., (2011) A floating-body/gate DRAM cell upgraded for long retention time, IEEE Electron Device Letters, 32, 731-733.

[109] Collaert N., Aoulaiche M., Rakowski M., Wachter B. De, Bourdelle K., Nguyen B. Y., Boedt F., Delprat D. and Jurczak M. (2009), Analysis of sense margin and reliability of 1T-DRAM fabricated on thin-film UTBOX substrates, IEEE International SOI Conference, pp. 1.

[110] Puget S., Bossu G., Fenouiller-Beranger C., Perreau P., Masson P., Mazoyer P., Lorenzini P., Portal J. M., Bouchakour R., and Skotnicki T., (2009) FDSOI Floating Body Cell eDRAM Using Gate-Induced Drain-Leakage (GIDL) Write Current for High Speed and Low Power Applications, IEEE International Memory Workshop, pp. 1-2.

[111] Hubert A., Bawedin M., Cristoloveanu S., and Ernst T., (2009) Dimensional effects and scalability of Meta-Stable Dip (MSD) memory effect for 1T-DRAM SOI MOSFETs, Solid-State Electronics, 53, 1280-1286.

[112] Okhonin S., Nagoga M., Carman E., Beffa, R., and Faraoni, E., (2007) New generation of Z-RAM, IEEE Electron Devices Meeting, pp. 925-928.

[113] Zhou Z., Fossum J. G., and Lu Z., (2009) Physical insights on BJT-based 1T DRAM cells, IEEE Electron Device Letters, 30, 565-567.

[114] Jeong H., Song K. W., Park I. H., Kim T. H., Lee Y. S., Kim S. G., Seo, J., Cho, K., Lee, K., Shin, H., and Lee, J. D., (2007) A new capacitorless 1T DRAM cell: Surrounding gate MOSFET with vertical channel (SGVC cell), IEEE Transactions on Nanotechnology, 6, 352-357.

[115] Okhonin S., Nagoga M., Lee, C. W., Colinge J. P., Afzalian, A., Yan R., Akhavan N. D., Xiong W., Sverdlov V., Selberherr S. and Mazure C., (2008) Ultra-scaled Z-RAM cell, IEEE International SOI Conference, pp. 157-158.

[116] Almeida L. M., Sasaki K. R. A., Aoulaiche M., Simoen E., Claeys C., and Martino J. A., (2012) One transistor floating body RAM performances on UTBOX devices using the BJT effect, Journal of Integrated Circuit System, 7, 113-118.

[117] Aoulaiche M., Simoen E., Caillat C., Witters L., Bourdelle K. K., Nguyen B. Y., Martino J. A., Claeys C., Fazan P., and Jurczak M., (2016) Understanding and optimizing the floating body retention in FDSOI UTBOX, Solid-State Electronics, 117, 123-129.

[118] Lu Z., Collaert N., Aoulaiche M., De Wachter B., De Keersgieter A., Schwarzenbach W., Bonnin O., Bourdelle K. K., Nguyen B. Y., Mazure C., and Altimime L., (2010) A novel low-voltage biasing scheme for double gate

FBC achieving 5s retention and 10^{16} endurance at 85°C, IEEE Electron Devices Meeting, pp. 12.3.1–12.3.4.

[119] Giusi G., (2014) Physical insights of body effect and charge degradation in floating-body DRAMs, Solid-State Electronics, 95, 1-7.

[120] Bawedin M., Cristoloveanu S. and Flandre D., (2007) Innovating SOI memory devices based on floating-body effects, Solid-State Electronics, 51, 1252-1262.

[121] Cho M. H., (2012) Thin-Body SOI Capacitorless DRAM Cell Design Optimization and Scaling. University of California, Berkeley, Technical Report No. UCB/EECS-2012-176.

[122] Moon D. I., Choi S. J., Han J. W., Kim S., and Choi Y. K., (2010) Fin-width dependence of BJT-based 1T-DRAM implemented on FinFET, IEEE Electron Device Letters, 31, 909-911.

[123] Kim S., Choi S. J., Moon D. I., and Choi Y. K., (2012) Carrier lifetime engineering for floating-body cell memory, IEEE Transactions on Electron Devices, 59, 367-373.

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

[125] Lin J.-T., Huang K. D., and Jheng B.-T., (2008) Performances of a capacitorless 1T-DRAM using polycrystalline silicon thin-film transistors with trenched body, IEEE Electron Device Letters, 29, 1222-1225.

[126] Avci U. E., Ban I., Kencke D. L., and Chang P. L. D., (2008) Floating body cell (FBC) memory for 16-nm technology with low variation on thin silicon and 10-nm BOX, IEEE International Conference on SOI Conference, 2008. pp. 29-30.

[127] Fossum J. G., Lu Z. and Trivedi V. P., (2007) New insights on "capacitorless" floating-body DRAM cells, IEEE electron device letters, 28, 513-516.

[128] Zhang, L., and Chan, M., (2016) Tunneling Field Effect Transistor Technology, Springer International Publishing, Switzerland, (ISBN- 978-3-319-31651-2).

[129] Oda, S. and Ferry, D. K., (2016) Nanoscale Silicon Devices, CRC Press Taylor and Francis Group, Florida (ISBN: 978-1-4822-2868-7).

[130] Packan P., (2007) Device and circuit interactions, IEEE IEDM Short Course

[131] Nowak E. J., (2002) Maintaining the benefits of CMOS scaling when scaling bogs down, IBM Journal of Research and Development, 46, 169-180.

[132] Voorde M. Van. de, Puers R., Livio B., Nooten, S. E. van, (2017) Beyond-CMOS Low-Power Devices: Steep-Slope Switches for Computation and Sensing, in Nanoelectronics: Materials, Devices, Applications, Wiley -VCH Verlag GmbH & Co (ISBN: 978-3-527-34053-8).

[133] Sze S.M., and Ng K. K., (2006) Physics of semiconductor devices, 3rd edition John Wiley & Sons (ISBN: 978-0-471-14323-9).

[134] Ionescu A. M., and Riel H., (2011) Tunnel field-effect transistors as energy-efficient electronic switches, Nature, 479, 329-337.

[135] Seabaugh A. C., and Zhang Q., (2010) Low-voltage tunnel transistors for beyond CMOS logic, Proceedings of the IEEE, 98, 2095-2110.

[136] Boucart K., and Ionescu A. M., (2007) Double-Gate Tunnel FET with High-K Gate Dielectric, IEEE Transactions on Electron Devices, 54, 1725-1733.

[137] Salahuddin S., and Datta S., (2008) Can the subthreshold swing in a classical FET be lowered below 60 mV/decade?, IEEE Electron Devices Meeting, pp. 1-4.

[138] Salahuddin S., and Datta S., (2008) Use of negative capacitance to provide voltage amplification for low power nanoscale devices, Nano letters, 8, 405-410.

[139] Salvatore G. A., Bouvet D., and Ionescu, A. M., (2008) Demonstration of subthrehold swing smaller than 60mV/decade in Fe-FET with P (VDF-TrFE)/SiO 2 gate stack, IEEE Electron Devices Meeting, pp. 1-4.

[140] Ko E., Lee J. W., and Shin C., (2017) Negative Capacitance FinFET with Sub-20-mV/decade Subthreshold Slope and Minimal Hysteresis of 0.48 V, IEEE Electron Device Letters, 38 418-421.

[141] Yuan Z. C., Rizwan S., Wong M., Holland K., Anderson S., Hook T. B., Kienle D., Gadelrab S., Gudem P. S., and Vaidyanathan M., (2016) Switching-speed limitations of ferroelectric negative-capacitance FETs, IEEE Transactions on Electron Devices, 63, 4046-4052.

[142] Kobayashi M., and Hiramoto, T., (2016) On device design for steep-slope negative-capacitance field-effect-transistor operating at sub-0.2 V supply voltage with ferroelectric HfO2 thin film, AIP Advances, 6, 025113.

[143] Jo J., and Shin C., (2016) Negative capacitance field effect transistor with hysteresis-free sub-60-mV/decade switching, IEEE Electron Device Letters, 37, 245-248.

[144] Ionescu A. M., Pott V., Fritschi R., Banerjee K., Declercq M.J., Renaud P., Hibert C., Fluckiger, P., and Racine, G. A., (2002) Modeling and design of a low-voltage SOI suspended-gate MOSFET (SG-MOSFET) with a metal-over-gate architecture, Proceedings International Symposium on Quality Electronic Design, pp. 496-501.

[145] Kam H., Lee D. T., Howe R. T., and King T. J., (2005) A new nano-electro-mechanical field effect transistor (NEMFET) design for low-power electronics, IEEE Electron Devices Meeting, pp. 463-466.

[146] Abelé N., Fritschi R., Boucart K., Casset F., Ancey P., and Ionescu A. M., (2005) Suspended-gate MOSFET: bringing new MEMS functionality into solid-state MOS transistor, In Electron Devices Meeting, pp. 479-481.

[147] Liu T. J. K., Jeon J., Nathanael R., Kam H., Pott V., and Alon E., (2010) Prospects for MEM logic switch technology, IEEE Electron Devices Meeting, pp.424-427.

[148] Xu N., Sun J., Chen I. R., Hutin L., Chen Y., Fujiki J., Qian C., and Liu T. J. K., (2014) Hybrid CMOS/BEOL-NEMS technology for ultra-low-power IC applications, IEEE Electron Devices Meeting, pp. 677-680.

[149] Gopalakrishnan K., Griffin P. B., and Plummer J. D., (2002) I-MOS: A novel semiconductor device with a subthreshold slope lower than kT/q, IEEE Electron Devices Meeting, pp. 289-292.

[150] Sarkar D., Singh N., and Banerjee, K., (2010) A novel enhanced electric-field impact-ionization MOS transistor, IEEE Electron Device Letters, 31, 1175-1177.

[151] Toh E. H., Wang G. H., Chan L., Lo G. Q., Samudra G., and Yeo, Y. C., (2007) Strain and materials engineering for the I-MOS transistor with an elevated impact-ionization region, IEEE Transactions on Electron Devices, 54, 2778-2785.

[152] Choi W. Y., Park B. G., Lee J. D., and Liu T. J. K., (2007) Tunneling field-effect transistors (TFETs) with subthreshold swing (SS) less than 60 mV/dec, IEEE Electron Device Letters, 28, 743-745.

[153] Gandhi R., Chen Z., Singh N., Banerjee K., and Lee S., (2011) Vertical Si-nanowire n-type tunneling FETs with low subthreshold swing (≤ 50 mV/decade) at room temperature, IEEE Electron Device Letters, 32, 437-439.

[154] Roy D. K., (2014) Tunnelling and negative resistance phenomena in semiconductors, Elsevier, (ISBN: 978-0-08-021044-5).

[155] Lilienfeld J. E., (1928) Device for controlling electric current, Patent US 1900018.

[156] Fowler R. H., and Nordheim L., (1928) Electron emission in intense electric fields. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 119, 173-181.

[157] Zener C., (1934) A theory of the electrical breakdown of solid dielectrics. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 145, 523-529.

[158] Müller E. W., (1937) Elektronenmikroskopische beobachtungen von feldkathoden. Zeitschrift für Physik A Hadrons and Nuclei, 106, 541-550.

[159] Chynoweth A. G. and McKay K. G., (1957) Internal Field Emission in Silicon p− n Junctions, Physical Review, 106, 418.

[160] Esaki L., (1958) New phenomenon in narrow germanium p− n junctions. Physical review, 109, 603.

[161] Kane E. O., (1961) Theory of tunneling, J. Appl. Phys., 32, 83-91.

[162] Keldysh L. V., (1958) Behavior of non-metallic crystals in strong electric fields, Soviet Journal of Experimental and Theoretical Physics, 6, 763.

[163] Giaever I., (1960) Energy gap in superconductors measured by electron tunnelling, Physical Review Letters, 5, 147.

[164] Bardeen J., (1961) Tunnelling from a many-particle point of view, Physical Review Letters, 6, 57.

[165] Josephson B. D., (1962) Possible new effects in superconductive tunnelling, Physics letters, 1, 251-253.

[166] Kahng D., and Sze S. M., (1967) A floating gate and its application to memory devices, Bell Labs Technical Journal, 46, 1288-1295.

[167] Chang L., Esaki L., and Tsu R., (1974) Resonant tunneling in semiconductor double barriers, Applied Physics Letters, 24, 593-595.

[168] Binnig G., Rohrer H., Gerber C., and Weibel E., (1982) Surface studies by scanning tunneling microscopy, Physical review letters, 49, 57.

[169] Frensley W. R., (1986) Transient response of a tunneling device obtained from the Wigner function, Physical review letters, 57, 2853.

[170] Van Der Wagt J. P. A., (1999) Tunneling-based SRAM, Proceedings of the IEEE, 87, 571-595.

[171] Bhuwalka K. K., Sedlmaier S., Ludsteck A. K., Tolksdorf C., Schulze J., and Eisele I., (2004) Vertical tunnel field-effect transistor, IEEE Transactions on Electron Devices, 51, 279-282.

[172] Feenstra R. M., Jena D., and Gu G., (2012) Single-particle tunneling in doped graphene-insulator-graphene junctions, Journal of Applied Physics, 111, 043711.

[173] Quinn J. J., Kawamoto G., and McCombe B. D., (1978) Subband spectroscopy by surface channel tunnelling, Surface Science, 73, 190-196.

[174] Banerjee S., Richardson W., Coleman J., and Chatterjee A., (1987) A new three-terminal tunnel device, IEEE Electron Device Letters, 8, 347–349.

[175] Takeda E., Matsuoka H., Igura Y., and Asai S., (1988) A band to band tunneling MOS device B2T-MOSFET, IEEE Electron Devices Meeting, pp. 402–405.

[176] Baba T., (1992) Proposal for surface tunnel transistors, Japanese journal of applied physics, 31, L455.

[177] Koga J., and Toriumi A., (1996) Negative differential conductance in three-terminal silicon tunneling device, Applied Physics Letters, 69, 1435-1437.

[178] Hansch W., Fink C., Schulze J., and Eisele I., (2000) A vertical MOS-gated Esaki tunneling transistor in silicon, Thin Solid Films, 369, 387-389.

[179] Aydin C., Zaslavsky A., Luryi S., Cristoloveanu S., Mariolle D., Fraboulet D., and Deleonibus, S., (2004) Lateral interband tunneling transistor in silicon-on-insulator, Applied Physics Letters, 84, 1780-1782.

[180] Hu C., Patel P., Bowonder A., Jeon K., Kim S. H., Loh W. Y., Kang C. Y., Oh J., Majhi P., Javey A., and Liu T. J. K., (2010) Prospect of tunneling green transistor for 0.1 V CMOS, IEEE Electron Devices Meeting, pp. 387-390.

[181] Appenzeller J., Lin Y.-M., Knoch J., Chen Z., and Avouris P., (2005) Comparing carbon nanotube transistors — the ideal choice: a novel

tunneling device design, IEEE Transaction on Electron Devices, 52, 2568–2576.

[182] Mayer F., Le Royer C., Damlencourt J. F., Romanjek K., Andrieu F., Tabone C., Previtali B., and Deleonibus S., (2008) Impact of SOI, $Si_{1-x}Ge_xOI$ and GeOI substrates on CMOS compatible tunnel FET performance, IEEE Electron Devices Meeting, pp. 163-166.

[183] Jeon K., Loh W. Y., Patel P., Kang C. Y., Oh J., Bowonder A., Park C., Park C. S., Smith C., Majhi P., and Tseng H. H., (2010) Si tunnel transistors with a novel silicided source and 46mV/dec swing. Symposium on VLSI Technology, pp. 121-122.

[184] Appenzeller J., Knoch J., Björk M. T., Riel H., and Riess W., (2008) Toward nanowire electronics, IEEE Transaction on Electron Devices, 55, 2827–2845.

[185] Toh E. H., Wang G.H., Chan L., Samudra G., Yeo Y. C., (2007) Device physics and guiding principles for the design of double-gate tunneling field effect transistor with silicon-germanium source heterojunction, Applied Physics Letters, 91, 243505.

[186] Zhang L., He J., and Chan M., (2012) A compact model for double-gate tunneling field-effect-transistors and its implications on circuit behaviors, IEEE Electron Devices Meeting, pp. 143-146.

[187] Conzatti F., Pala M. G., Esseni D., Bano E., and Selmi L., (2011) A simulation study of strain induced performance enhancements in InAs nanowire Tunnel-FETs, IEEE Electron Devices Meeting, pp. 95-98

[188] Dewey G., Chu-Kung B., Boardman J., Fastenau J. M., Kavalieros J., Kotlyar R., Liu W. K., Lubyshev D., Metz M., Mukherjee N., and Oakey P., (2011) Fabrication, characterization, and physics of III–V heterojunction tunneling field effect transistors (H-TFET) for steep sub-threshold swing, IEEE Electron Devices Meeting, pp. 785-788.

[189] Fiori G., Iannaccone G., (2009) Ultralow-voltage bilayer graphene tunnel FET, IEEE Electron Device Letters, 30, 1096–1098.

[190] Mookerjea S., Mohata D., Mayer T., Narayanan V., and Datta S., 2010. Temperature-Dependent I-V Characteristics of a Vertical In0.53Ga0.47 Tunnel FET, IEEE Electron Device Letters, 31, 564-566.

[191] Born M., Bhuwalka K. K., Schindler M., Abelein U., Schmidt M., Sulima T., and Eisele I., (2006) Tunnel FET: A CMOS device for high temperature applications, IEEE International Conference on Microelectronics, pp. 124-127.

[192] Fulde M., Heigl A., Weis M., Wirnshofer M., Arnim K. V., Nirschl T., Sterkel M., Knoblinger G., Hansch W., Wachutka G., and Schmitt-Landsiedel D., (2008) Fabrication, optimization and application of complementary Multiple-Gate Tunneling FETs, IEEE International Nanoelectronics Conference, pp. 579-584.

[193] Moselund K. E., Bjork M. T., Schmid H., Ghoneim H., Karg S., Lortscher E., Riess W., and Riel H., (2011) Silicon nanowire tunnel FETs: Low-temperature operation and influence of high-k gate dielectric, IEEE Transactions on Electron Devices, 58, 2911-2916.

[194] Singh J., Ramakrishnan K., Mookerjea S., Datta S., Vijaykrishnan N., and Pradhan D., (2010) A novel si-tunnel FET based SRAM design for ultra low-power 0.3 VV DD applications, In Proceedings of Asia and South Pacific Design Automation Conference, pp. 181-186.

[195] Saripalli V., Mohata D. K., Mookerjea S., Datta S., and Narayanan V., (2010) Low power loadless 4T SRAM cell based on degenerately doped source (DDS) In0.53Ga0.47As tunnel FETs, IEEE Device Research Conference, pp. 101-102.

[196] Sedighi B., Hu X. S., Liu H., Nahas J. J., and Niemier M., (2015) Analog circuit design using tunnel-FETs, IEEE Transactions on Circuits and Systems I: Regular Papers, 62, 39-48.

[197] Biswas A., Dagtekin N., Grabinski W., Bazigos A., Le Royer C., Hartmann J. M., Tabone C., Vinet M., and Ionescu A. M., (2014) Investigation of tunnel field-effect transistors as a capacitor-less memory cell, Applied Physics Letters, 104, 092108.

[198] Biswas A., and Ionescu A. M., (2015) 1T capacitor-less DRAM cell based on asymmetric tunnel FET design, IEEE Journal of the Electron Devices Society, 3, 217-222.

[199] Biswas A., and Ionescu A. M., (2014) Study of fin-tunnel FETs with doped pocket as capacitor-less 1T DRAM, IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference, pp. 1-2.

[200] Kim H., Park B.-G., (2016) A 1T-DRAM cell based on a tunnel field-effect transistor with highly-scalable pillar and surrounding gate structure, Journal of the Korean Physical Society, 69, 323-327.

[201] Atlas User's Manual, Silvaco Int., Santa Clara, CA, Jan. 18, 2015.

[202] Badwan A. Z., Chbili Z., Yang Y., Salman A. A., Li Q., and Ioannou D. E., (2013) SOI field-effect diode DRAM cell: Design and operation, IEEE Electron Device Letters, 34,1002-1004.

[203] Badwan A. Z., Li Q., and Ioannou D. E., (2015) On the nature of the memory mechanism of gated-thyristor dynamic-RAM Cells, IEEE Journal of the Electron Devices Society, 3, 468-471.

[204] Yang Y., Gangopadhyay A., Li Q., and Ioannou D. E., (2009) Scaling of the SOI field effect diode (FED) for memory application, IEEE Semiconductor Device Research Symposium, pp. 1-2.

[205] Salman A. A., Beebe S. G., Emam M., Pelella M. M., and Ioannou D. E., (2006) Field Effect Diode (FED): A novel device for ESD protection in deep sub-micron SOI technologies, IEEE Electron Devices Meeting, pp. 1–4.

[206] Raissii F., (1996) A brief analysis of the field effect diode and breakdown transistor, IEEE Transaction on Electron Devices, 43, 362–365.

[207] Wan J., Royer C. Le, Zaslavsky A., and Cristoloveanu S., (2013) Progress in Z^2-FET 1T-DRAM: retention time, writing modes, selective array operation, and dual bit storage, Solid-State Electronics, 84, 147-154.

[208] Dirani E. H., Solaro Y., Fonteneau P., Ferrari P., and Cristoloveanu S., (2016) Properties and mechanisms of Z^2-FET at variable temperature, Solid-State Electronics, 115, 201-206.

[209] Wan J., Zaslavsky A., and Cristoloveanu S., (2015) Comment on Investigation of tunnel field-effect transistors as a capacitor-less memory cell, Applied Physics Letters, 106, 016101.

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

[211] Dirani H. El, Bawedin M., Lee K., Parihar M., Mescot X., Fonteneau P., Galy Ph., Gamiz F., Kim Y-T., Ferrari P., and Cristoloveanu S., (2016) IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference, pp. 1-2.

[212] Wan J., Cristoloveanu S., Le Royer C., and Zaslavsky A., (2012) A feedback silicon-on-insulator steep switching device with gate-controlled carrier injection, Solid-State. Electron, 76, 109–111.

[213] Nemati F., and Plummer J.D., (1999) A novel thyristor-based SRAM cell (T-RAM) for high-speed, low-voltage, giga-scale memories, IEEE Electron Devices Meeting, pp. 283-286.

[214] Cho H.-J., Nemati F., Roy R., Gupta R., Yang K., Ershov M., Banna S., Tarabbia M., Salling C., Hayes D., and Mittal A., (2005) A novel capacitor-less DRAM cell using thin capacitively-coupled thyristor (TCCT), IEEE Electron Devices Meeting, pp. 311-314.

[215] Mulaosmanovic H., Paolucci G. M., Compagnoni C. M., Castellani N., Carnevale G., Fantini P., Ventrice D., Lacaita A. L., Spinelli A. S., and Benvenuti A., (2014) Working principles of a DRAM cell based on gated-thyristor bistability, IEEE Electron Device Letters, 35, 921-923.

Chapter 2

Improving Retention Time in Tunnel Field Effect Transistor based DRAM through Back Gate Engineering

2.1 Introduction

The ever increasing demand of high speed and dense memory necessitates reduction in the size of DRAM cell [1-8]. However, scaling the capacitor is the most critical issue in DRAM as it reduces the charge storage [1-12]. This adversely affects the charge retention, and thus, requires more refresh cycles [3-6]. In the conventional DRAM cell data is stored in the capacitor as electrical charge, but electrical charge leaks over time. Therefore, DRAM must be refreshed periodically to preserve the stored data. Refresh negatively impacts DRAM performance and power dissipation [4]. As the speed and size of DRAM devices continue to increase with each new technology, the performance and power overheads are increasing significantly [3-8]. Moreover, transistor scaling leads to higher leakage current that result into higher power dissipation and also, reduction in retention time. Thus, a high retention is essential to reduce refresh cycles that consumes ~40-50% of energy in off-chip memory hierarchy [3,10]. Due to difficulty in scaling the capacitor associated with the conventional DRAM cell [1-7,9-12], the single transistor (1T) cells [13-30] have been proposed. Further, the quest for DRAM with high retention and low power necessitates use of steep switching devices as dynamic memory [31-43].

2.2 Steep-switching devices as dynamic memory

The devices with steep switching have shown the potential to replace conventional with applicability in logic circuits [44-45] and thus, have been

exploited as dynamic memories [31-43]. These devices include, Thin-Capacitively Coupled Thyristor (TCCT) [31-33], Field Effect Diode (FED) [34-36], Zero sub-threshold swing and Zero impact ionization FET (Z^2-FET) [37-40], and Tunnel Field Effect Transistor (TFET) [41-43]. These devices have different type of dopants for source and drain. TCCT, FED and Z^2-FET operate in forward bias and utilize the positive feedback mechanism for conduction. Z^2-FET, FED and TCCT based dynamic memories have shown high operating current, current sensing margin, retention time and speed [40]. The devices have been evolving with further modifications to improve the performance metrics and needs further investigation to improve the scalability, applicability and operation at higher temperature.

TCCT, FED and Z^2-FET (Figs. 2.1(a)-(c)) are variants of Feedback Field Effect Transistor (FB-FET) [44-46] that exhibits a very steep transition from off-to-on state. The device forms a *p-n-p-n* structure and can be presented as *pnp-npn* latch [47-50] (Fig. 2.1(d)) with an electron (V_n) and hole (V_p) injection barriers (Fig. 2.1(e)) created through various methodologies. The latch state defines a feedback loop, where conduction of a transistor triggers another transistor [48]. FB-FET utilizes positive and negative surface charge densities, generated at underlap regions, in spacers adjacent to the gates to form the barriers [44]. When electrons are injected into the channel, few of the holes accumulated in the barrier reduce the barrier height for electrons and so does the electron accumulation for conduction due to holes [38,44]. The charge reposition increases the conduction that further reduces the barrier heights, and thus, a feedback mechanism is triggered. The feedback mechanism makes the threshold voltage of FBFET strongly dependent on the surface density which is difficult to control [44].

In thyristor based feedback action, the injection barriers are formed through precise doping of *p*-type and *n*-type regions in the channel (Fig. 2.1(a)). This triggers a doping dependent bipolar action [47-49]. The device utilizes the *p*-type doped region in the channel for charge storage. The successor of FB-FET without spacers is FED that utilizes two independent front gates to create different injection barriers (Fig. 2.1(b)) [34-36,44,51,52]. The freedom to operate two gates

independently enable their operation as *p-i-n* diode as well thyristor. FED has been explored for Electrostatic Discharge (ESD) protection [51,52], and recently as dynamic memory [34-36] with semiconductor film under p^+ poly gate as storage region. The anode current characteristic goes through a negative resistance region [34,44], and finally, snaps back to distinguish between the states of memory (Fig. 2.1(f)). The device characteristics are similar to that of TCCT. However, FED exhibits better stability as compared to TCCT due to doping independent operation [44].

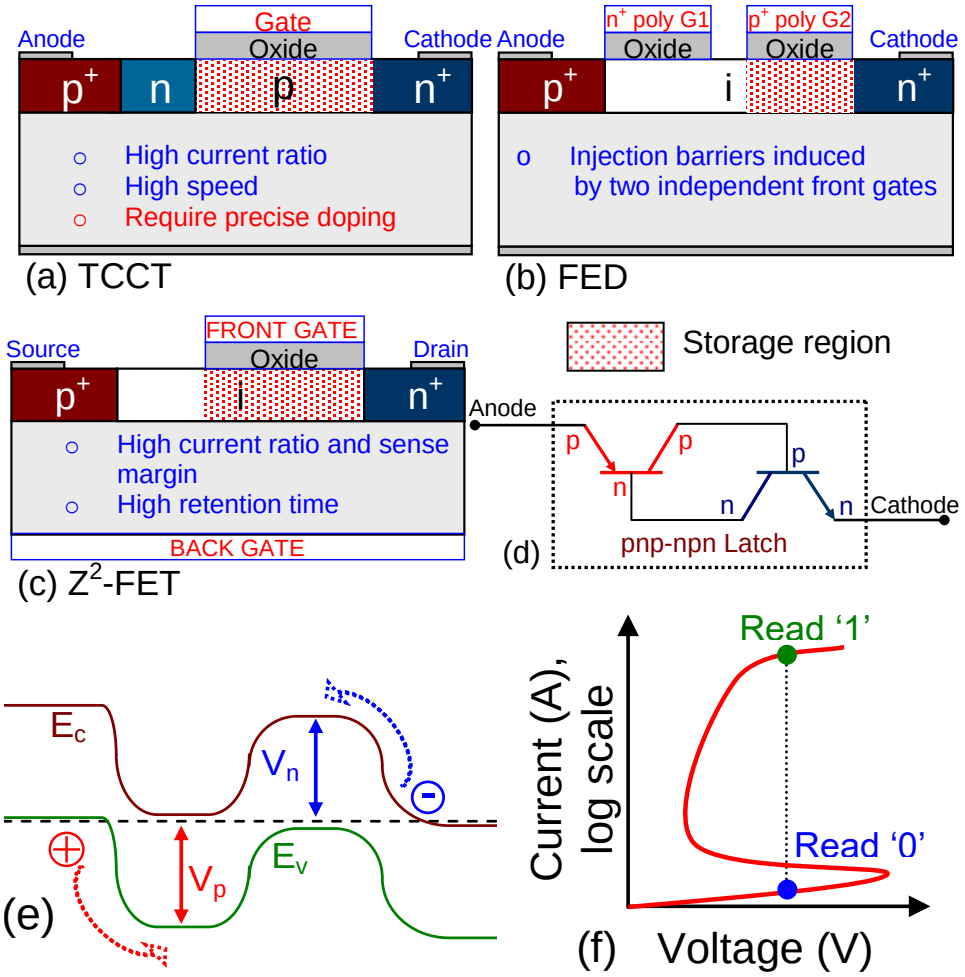

Fig. 2.1 Schematic diagram of (a) TCCT, (b) FED, and (c) Z^2-FET along with their storage region, represented by shaded region [44]. Equivalent schematic representation of (d) these devices along with their (e) feedback mechanism, demonstrated through energy band diagrams, and (f) drain-current characteristics illustrating snapback action with state '1' and '0' of memory device [44]. V_n and V_p are the electron and hole injection barriers, respectively.

Another p^+-i-n^+ structured device that incorporates the feature of FB-FET and FED with operation in forward bias is Z^2-FET (Fig. 2.1(c)) [37-40]. The device uses front and back gates to control the carrier injection barriers, and exploits the region under front gate, near n^+ doped drain region for charge storage. The back gate is biased positively, while the front gate negatively to create a *p*-*n*-*p*-*n* structure. A similar asymmetric Double Gate (DG) TFET utilized as DRAM is demonstrated in Fig. 2.2(a) [37,38]. Unlike Z^2-FET, TFET operates in reverse bias [41] in the read operation and utilizes the underlap region (L_{in}) for charge storage.

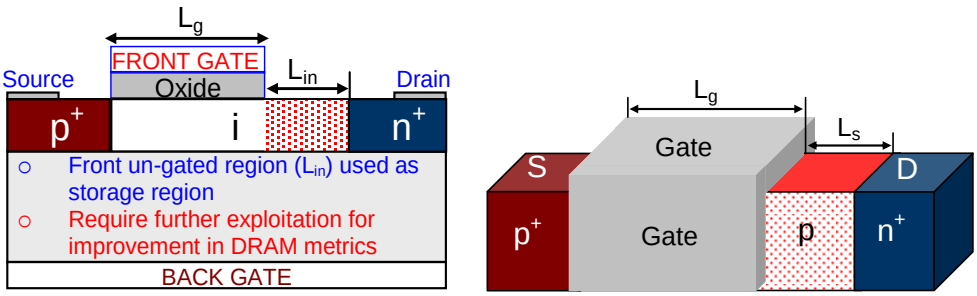

Fig. 2.2 (a) Schematic diagram of asymmetric DG FDSOI TFET with gate length (L_g) of 400 nm and a front un-gated region (L_{in}) of 200 nm [41]. (b) Schematic diagram of a fin-based TFET with L_g = 100 nm and a *p*-doped pocket (L_s) of 25 nm with 10^{18} cm^{-3} doping concentration [43].

Biswas *et al.* presented the concept of exploiting asymmetric TFET for DRAM applications [41,42]. The semiconductor film underneath the front gate of the TFET structure (Fig. 2.2(a)) remains in inversion due to a constant high front gate bias, while the back gate was extended all over the intrinsic region. The holes are accumulated in the induced potential well (created at the front ungated region, L_{in}) by applying a negative bias at the back gate and a small positive bias at the source that push the holes into the body [41]. In another programming scheme, the front gate and source are kept at zero bias, while a negative bias is applied at the back gate along with a small drain bias to prevent the holes from recombining at the drain side [42]. The sense margin, which is defined as the difference in the read currents for state '1' and '0' [14] was reported in the range of 10 nA to 20 nA [41,42], which could be improved to 500 nA by employing a fin based tunnel FET [43]. The time required to reduce the maximum sense margin (ΔI) by 50% is assessed as retention time [14,15]. Few papers define *RT* as the time when *SM* = 0

nA/μm [30,42]. However in our work, the earlier definition of the change in current (ΔI) by 50% [14] is referred for the evaluation of *RT*.

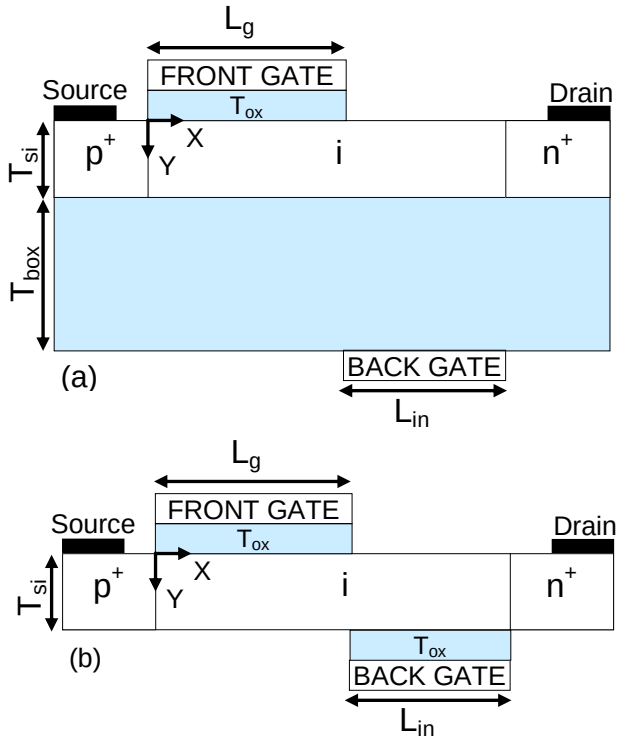

Fig. 2.3 Schematic diagram of (a) UTBOX SOI, and (b) DG TFET device. *X* indicates the direction along the channel length and *Y* indicates the direction perpendicular to the channel length.

Retention time achieved through the first design [41] was in the order of 100's of microseconds to few milliseconds could be achieved at room temperature. The second approach of utilizing a p^+ doped pocket (Fig. 2.2(b)) as storage region required a precise control on doping of p^+ region and the retention time reported is in the orders of 100 μs at 85 °C [43]. Although Biswas *et al.* presented an original concept showing DRAM characteristics using TFET, the retention time attained was still lower than the target specified by ITRS (64 ms) [8] at 85 °C. Therefore, a careful reinvestigation necessitates the use of a misaligned architecture utilized as Ultra-thin Buried Oxide (UTBOX) and Double Gate (DG) FDSOI TFET for improved retention characteristics. The chapter investigates on TFET based DRAM and presents an approach to enhance charge retention DRAM by back gate engineering through misalignment, optimal bias and workfunction values at a temperature of 85 °C. The impact of carrier lifetime, temperature, voltage and gate length scaling are also analysed. Thus, the chapter provides insight into

understanding the functioning of TFET based DRAM along with the physical phenomenon occurring in the device. DRAM optimally utilizes various attributes to improve the retention time in fully depleted SOI technology.

2.3 Device description and simulation

As shown in Figs. 2.3 (a) and (b) TFET structures used in the analysis is a p^+-i-n^+ UTBOX and DG FDSOI n-type TFET device, respectively, with intrinsic channel having a total length of 600 nm of which the front gate region (L_g) is 400 nm and the back gate region (L_{in}) is 200 nm. The n^+ polysilicon front gate controls the partial region of the silicon film while the p^+ poly back gate controls the front un-gated region near the drain that results in the formation of an electrically induced potential well. A device with n^+ poly front gate and p^+ poly back gate can be fabricated as suggested by Tanaka *et al.* [53] with several gate materials available for achieving higher workfunction values [20,54]. The device specifications are given in Table 2.1.

Device analysis has been carried out using Atlas [55] simulation software with calibrated models for tunneling phenomenon. The work throughout the thesis is based on TFET [56-60], and therefore, it is imperative to analyse various tunneling models used by the simulator. The tunneling models are basically classified as local and non-local model [55]. The local tunneling models, assume either an average or maximum electric field which is constant throughout the tunneling path [55,56,60]. In non-local model, the current depends on the band-edge profile along the entire path, and thus, the electric field is evaluated at each point [57]. The electric field is changing dynamically and thus, it makes tunneling a non-local process (Fig. 2.4). Under the assumption of a uniform electric field, the generation/recombination rate is reduced to the well-known Kane and Keldysh models [61,62]. The tunneling generation rate (G_{BBT}) is [55] expressed as:

$$G_{BBT} = D\ BB.A\ E^{BB.GAMMA} exp\left(\frac{-BB.B}{E}\right) \qquad (2.1)$$

where E is the magnitude of the electric field, D is a statistical factor, and *BB.A*, *BB.B,* and *BB.GAMMA* are user-definable parameters. The model parameters can be set by specifying *BBT.STD* or *BBT.KL* on the as the standard tunneling models [55]. In applications, *BBT.STD* is generally used with direct transitions while

BBT.KL [63] with indirect. The other commonly used local model for tunneling includes models proposed by Kane [61], Schenk [64], and Hurkx [65].

Table 2.1 Device specifications

Specifications	***UTBOX TFET***	***DG TFET***
Front gate region (L_g)	400 nm	400 nm
Back gate region (L_{in})	200 nm	200 nm
Width (W)	1 µm	1 µm
Source/Drain (S/D) doping	10^{20} cm^{-3}	10^{20} cm^{-3}
Silicon thickness (T_{si})	20 nm	20 nm
Oxide thickness (T_{ox})	3 nm (HfO_2)	3 nm (HfO_2)
Box thickness (T_{box})	10 nm	-
Front gate workfunction (φ_{m1})	n^+ poly (4.15 eV)	n^+ poly
Back gate workfunction (φ_{m2})	p^+ poly (5.25 eV)	p^+ poly
Temperature (T)	85 °C	85 °C

Tunneling can thus, be analysed using local or non-local methods. However, default parameters of both the models might require calibration with the available experimental data. Therefore, in this work, a more realistic approach of specifying a local model along with a non-local model is adopted, as was suggested by few authors [59,60]. The local model is not applied inside the region specified by quantum meshes [55]. Thus, to capture the essential characteristics of TFET devices as shown by experimental results, non-local model simultaneously with Klaassen model [63] is utilized. The user defined parameters (Table 2.2) *BB.A* and *BB.B* are modified to achieve the best fit to the experimental data [56]. This is similar to that suggested by Biswas *et al.*, [57] and Fukuda *et al.*, [58]. Therefore, in our approach, the tunneling parameters of Klaassen model are calibrated to capture the essential characteristics of TFET. The other modules used are band gap narrowing, Shockley Read Hall (SRH) recombination model [63], Fermi statistics [66] and Lombardi mobility model.

Our simulation results along with the available experimental data published in the literature [56] are shown in Fig. 2.5. Fig. 2.5(a) shows the schematic of TFET

being simulated and its drain-current characteristics in Fig. 2.5(b). To study the DRAM characteristics, electron and hole lifetimes are very crucial as they govern generation/recombination [63,67] and are modeled according to the SRH model, where the default values of both, electron and hole lifetime is assumed to be equal (with $\tau_{n0/p0}$ (300 K) = 100 ns) as utilized in various works [19,27,28]. The temperature (T) dependent carrier lifetime was as described in equation 2.2 by Schenk with α = 1.5 [67],

$$\tau_{n0/p0}(T\ in\ Kelvins) = \tau_{n0/p0}(300\ K)\left(\frac{300}{T}\right)^{\alpha} \qquad (2.2)$$

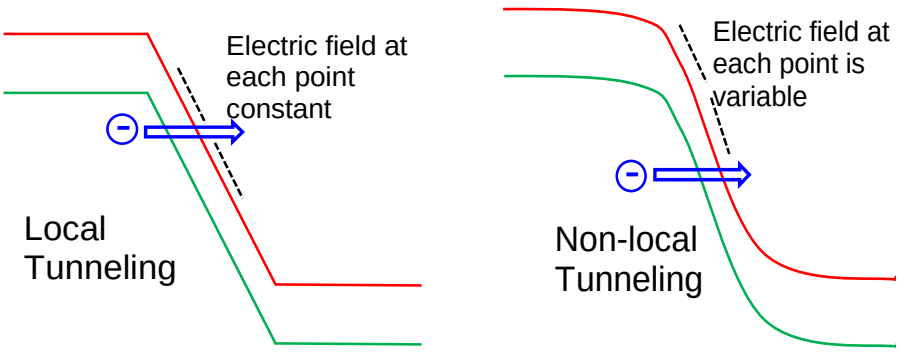

Fig. 2.4 Schematic representation of energy band diagrams illustrating variation in the evaluation of electric field with local and non-local band-to-band tunneling models.

Table 2.2 User-defined Klaassen band-to-band tunneling parameters [55,63].

Parameter name	***Default value***	***Units***
BB.A	4×10^{14}	$cm^{-3}s^{-1}$
BB.B	1.9×10^{7}	Vcm^{-1}
BB.GAMMA	2.5	none

Further, their doping dependence was according to the Scharfetter relation in equation 2.3 (with $\tau_{max} = \tau_{n0/p0}(T)$ τ_{min} = 0 ns, N_{reff} = 5 x 10^{16} cm^{-3} and γ = 1) [52],

$$\tau_{n/p} = \frac{\tau_{max} - \tau_{min}}{1 + (N/N_{reff})^{\gamma}} \qquad (2.3)$$

where N is the total doping density. Incorporation of all the models discussed in this section lead to a systematic analysis of DRAM characteristics, illustrating various and unique attributes of TFET for functionality as DRAM, which will be discussed in the next sections.

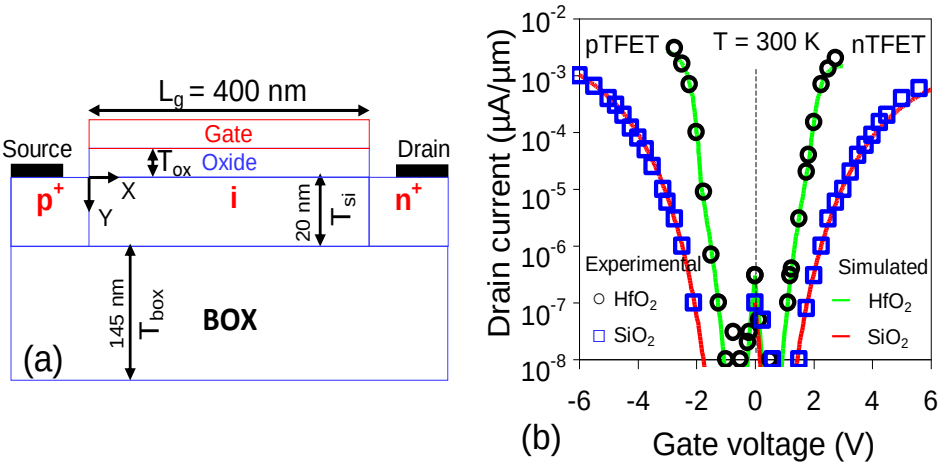

Fig. 2.5 (a) Schematic diagram of Tunnel FET along with its (b) drain current-gate voltage characteristics calibrated with experimental data [56] at 300 K. Parameters: L_g = 400 nm, T_{si} = 20 nm, T_{ox} = 3 nm (HfO_2) and 6 nm (SiO_2) T_{box} = 140 nm, and V_d = 1 V. V_d is the drain voltage.

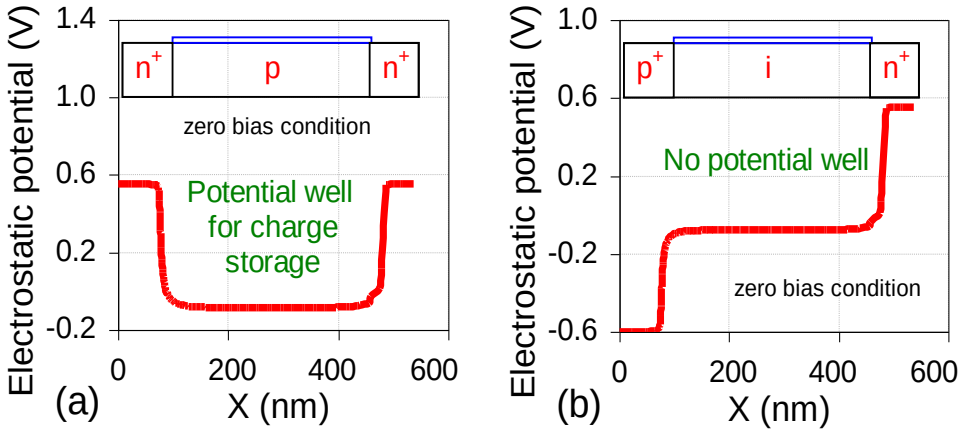

Fig. 2.6 Variation in electrostatic potential along the channel direction (X) for (a) conventional n-type MOS and (b) TFET along with their schematic diagrams.

2.4 Misaligned TFET as dynamic memory

2.4.1 Device design for operation as DRAM

An essential requirement for all types of dynamic memories is the storage area for charge carriers. An electrostatic potential well can be created with a low potential region amidst two regions with higher potential. *n*MOSFETs have an n^+-p-n^+ structure in which the p-type body with holes accumulated behaves as low potential region between two heavily doped n^+ region having high potential [25]. Thus, p-type body implicitly behaves as the storage region (Fig. 2.6 (a)) in conventional MOS devices [25]. For TFET, with a p^+-i-n^+ structure, the creation of potential well is critical as electrostatic potential well is not implicitly formed,

as shown in Fig. 2.6(b). Therefore, the architecture is modified and used as asymmetric FET, where in previous work [41,42] front gate is aligned at a partial portion of intrinsic film and back gate at complete intrinsic region. In our work, the back gate is misaligned and positioned at front un-gated region. The basic principles governing the operation of back gate engineered TFET as a DRAM have been analysed through an UTBOX TFET structures, and then the same are applied to the DG TFET. The use of a thin box at the back increases the electrostatic control of the back gate [47] that permits voltage scaling as well as formation of a deeper potential well that could sustain charges for longer duration. Therefore, by back gate engineering through location, workfunction and bias values, the performance of TFET for DRAM cell can be optimized.

Table 2.3 Programming Scheme for UTBOX TFET

Operation	V_d	V_{fg}^{a}	V_{bg}^{b}	***Time***
Write 1	0 V	0 V	-5 V	50 ns
Write 0	0 V	0 V	5 V	50 ns
Read	1 V	4 V	1 V	50 ns
Hold	0 V	0 V	-0.25 V	-

V_{fg} is the front gate voltage.
V_{bg} is the back gate voltage.

2.4.2 Operating mechanism

The programming scheme adopted for the UTBOX FDSOI TFET to perform as DRAM is given in Table 2.3 and the operating mechanism is described as below:

2.4.2.1. Write operation

Storage of the majority excess carriers (holes) is defined as write '1' or program operation and the removal of excess carriers is termed as write '0' or erase operation [13,14]. Write '1' operation is performed by applying a negative bias at the back gate that increases the potential depth (Fig. 2.7(a)) that signifies an increase in the hole concentration (~10^{18} cm^{-3}) in the storage region (Fig. 2.7(b)). The erase operation is carried out by applying a positive back bias that raises the potential as shown in Fig. 2.7(a). It expels holes from the potential well which recombines at the drain side [14] and the concentration is reduced (~ 10^{4} cm^{-3}).

Thus, the difference in the hole concentration in the storage area after write '1' and '0' operation can be clearly observed in Fig. 2.7(b).

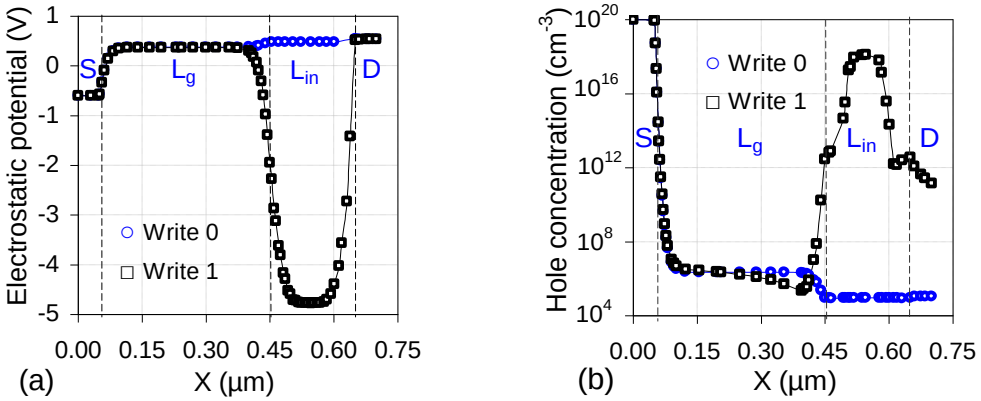

Fig. 2.7 Variation of (a) electrostatic potential, and (b) hole concentration along the channel direction (X) after state '1' and '0' are stored. The cutlines are taken at a cross-section 3 nm above the back interface ($Y = T_{si}$).

2.4.2.2. Read operation

The read mechanism is based on BTBT between source and channel regions. Electrons tunnel from the p^+ source into the channel region. In order to drift them into the drain region, the energy barrier at the L_{in} that resists the flow of electrons has to be lowered. Therefore, a positive potential is applied at the back gate. The presence of more holes at the back, increase the effective potential at the front interface of the L_{in} region for state '1' compared to state '0' [13]. This results into a lower barrier, and hence, into a higher drain current for read '1' as compared to read '0', as shown in Fig. 2.8(a). Thus, the presence or absence of majority carriers affects the drain current, whereas the concentration of the holes depends on the bias applied during the hold operation as demonstrated in Fig. 2.8(b).

2.4.2.3. Hold operation

Hold time is defined as the maximum time between write and read, where state is still being read correctly [13]. The generation of holes in the storage region degrades state '0', while their recombination leads to degradation of state '1'. For lower back gate voltage, state '0' is degraded due to thermal generation and tunneling at the drain junction that accumulates holes in the storage region, while for higher back gate voltage due to thermal recombination and diffusion of holes

from the potential well, the state ‘1’ degrades [13-15]. Thus, based on the bias applied disturb ‘1’ and ‘0’ conditions prevail.

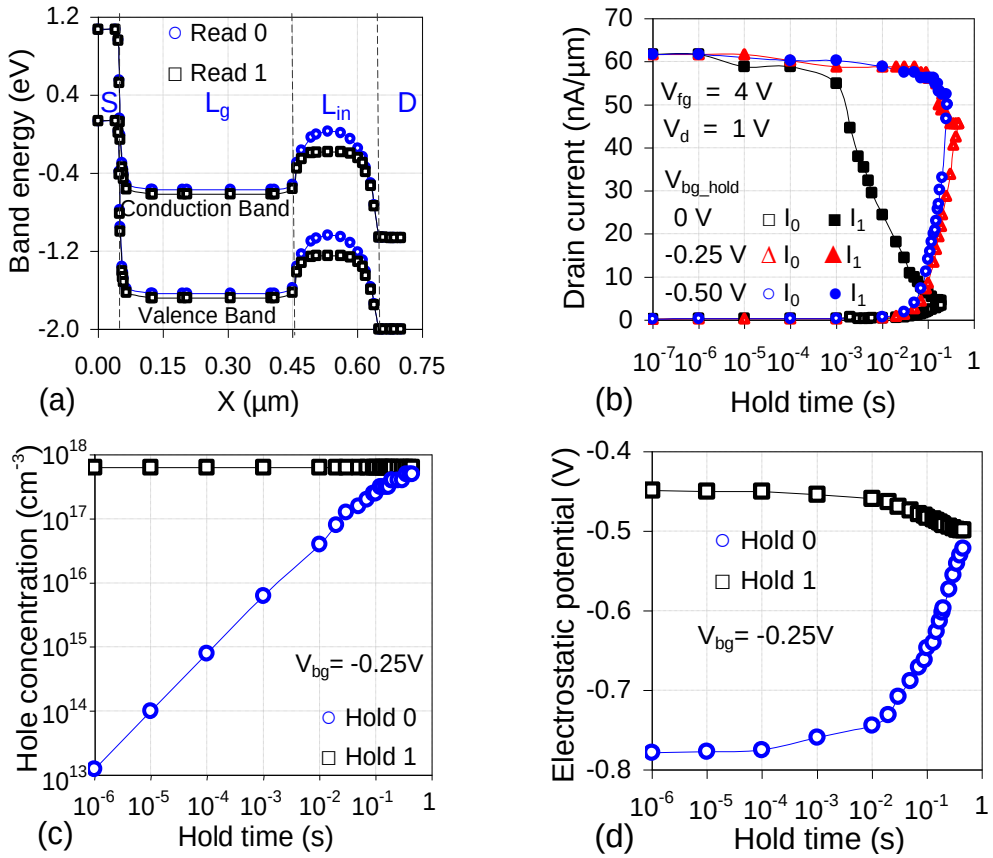

Fig. 2.8 (a) Variation in the band energy for read ‘1’ and ‘0’ along the channel direction (X) at a cross-section 3 nm below Y = 0. (b) Dependence of read currents, I_1 and I_0 on the back gate bias during hold state, with the time. Variation of (c) hole concentration, and (d) electrostatic potential with the hold time.

2.4.3 Optimizing back gate bias

2.4.3.1. Retention Time (RT)

During hold, recombination and generation of holes are the governing phenomenon for the retention of states. The same is reflected in Fig. 2.8(b) where for a back gate bias (V_{bg}) of 0 V during hold, I_1 decreases due to recombination of holes and for V_{bg} equal to or less than -0.25 V, I_0 increases due to generation of holes. Thus, the sense margin reduces with time as shown in the Fig. 2.8(b). A small negative voltage preserves the potential well that sustains the holes for a longer duration, thus maintaining state ‘1’. While for state ‘0’, holes inject into the storage region, as shown in Fig. 2.8(c). After write ‘0’, the hole concentration

observed in the storage region is ~10^4 cm^{-3} (Fig 2.7(a)) that increases to 10^{13} cm^{-3} while holding it for 1 µs at a back gate bias of -0.25 V. The presence of more holes keeps state '1' at a higher potential [13] in the L_{in} region, as shown in Fig. 2.8(d). Thus, state '0' degrades while state '1' maintains at a constant hole concentration of ~ 10^{17} cm^{-3} (Fig. 2.8(c)). This is also evident from Fig. 2.8(d) where we can observe that state '0' degrades, and finally, around 450 ms both the states are indistinguishable. An optimized bias (back gate voltage of -0.25 V during hold operation) attain the best retention time of 170 ms (Fig. 2.9(a)). The results have been compared with UTBOX SOI TFET with back gate positioned at complete intrinsic region (Fig. 2.9(a)) similar to that analysed by Biswas *et al.* [41]. The optimized bias shows a lower retention time of ~ 80 µs (V_{bg_hold} = 0 V) which verifies that along with the optimized bias, the topology of the device utilized as DRAM is crucial.

2.4.3.2. Sense Margin (SM)

The back gate controls the energy barrier formed at the storage region. A higher back gate voltage (> 1 V) lowers the energy barrier, increasing the read currents (I_0 and I_1) that can lower the sense margin ($I_1 - I_0$) whereas a lower back gate voltage (< 0.5 V) would have an enhanced barrier that would result into lower read currents and thus, a degraded *SM*. Therefore, we need to optimize the bias value to achieve a good *SM*. As shown in Fig. 2.9(b) for a more negative bias hole generation degrades *SM* [18,26]. For bias below 1 V the read current for state '1' is low, decreasing the sense margin. As the bias is increased beyond 1 V, read current for state '0' increases, thereby decreasing the sense margin. Therefore, the best *SM* is obtained at 1 V with I_0 having the value of 0.3 nA and I_1 as ~61 nA with difference of about 60 nA between the two states.

The memory operations were performed in the sequence as shown in Fig. 2.9(c) with programming scheme specified in Table 2.3. The read currents for state '1' and '0' are shown in the Fig. 2.9(d). It shows a sense margin of ~60 nA and the retention time achieved is ~170 ms. Although the sense margin obtained is low in the range of nano-Amperes, we can observe that the current ratio (I_1/I_0) obtained is ~ 10^2 (Fig. 2.9(c)) which provides a good read sensitivity. Thus, by optimizing the back gate voltages, an acceptable sense margin along with high retention time at

85 °C could be achieved, which is higher compared to the previous work [42] (range of few ms at room temperature). Even for same set of bias values, the *RT* achieved for misaligned architecture is higher than architecture with back gate at complete intrinsic region (Fig. 2.9(a)). Thus, back gate position is crucial and will be discussed in the subsequent section.

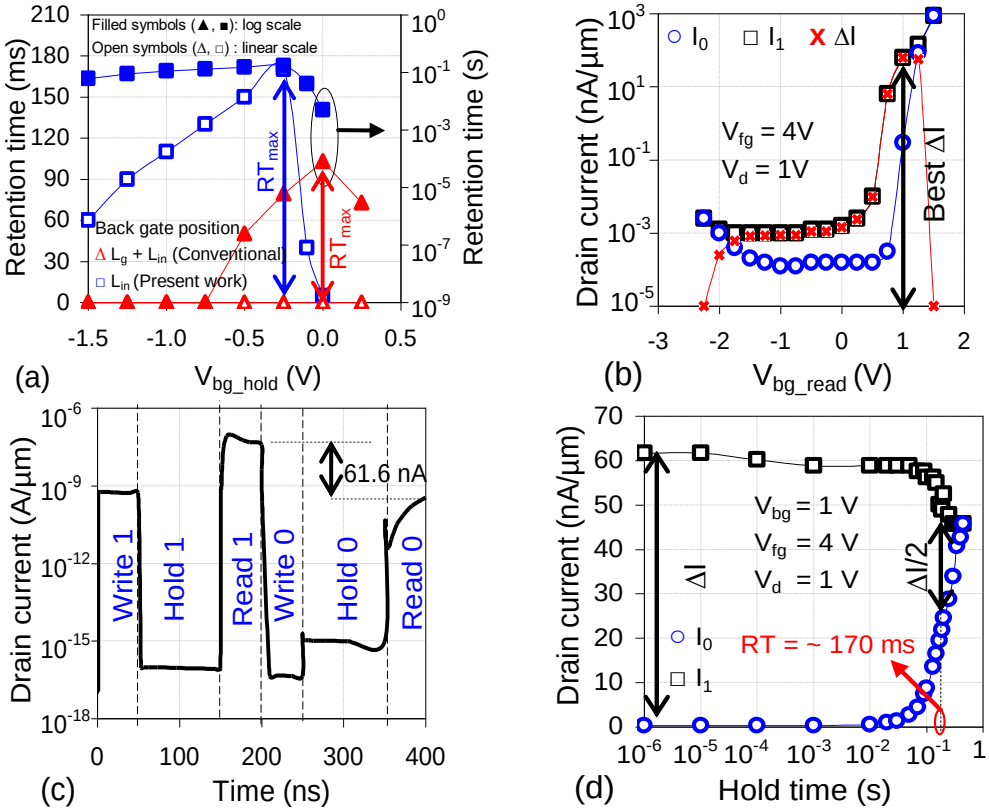

Fig. 2.9 (a) Variation of *RT* with the V_{bg} applied during hold operation on UTBOX TFET with back gate at the underlap region (our work) and at the complete intrinsic region (conventional) [41]. (b) Dependence of I_1, I_0 and ΔI on V_{bg} during read operation. (c) Drain current transients in the sequence of operation. (d) I_1 and I_0 showing a *RT* of 170 ms estimated when read margin is $\Delta I/2$.

2.4.4 Impact of back gate position

Fig. 2.10 highlights the impact of back gate alignment on the potential well, indicating the significance of back gate position to operate as DRAM. A perfectly aligned back gate with respect to front gate (Case A) doesn't form a dedicated volume for charge storage and, thus, the design is not suited as DRAM (Fig. 2.10(a)). Figs. 2.10(b) show asymmetric TFET with back gate at complete intrinsic region (Case B).

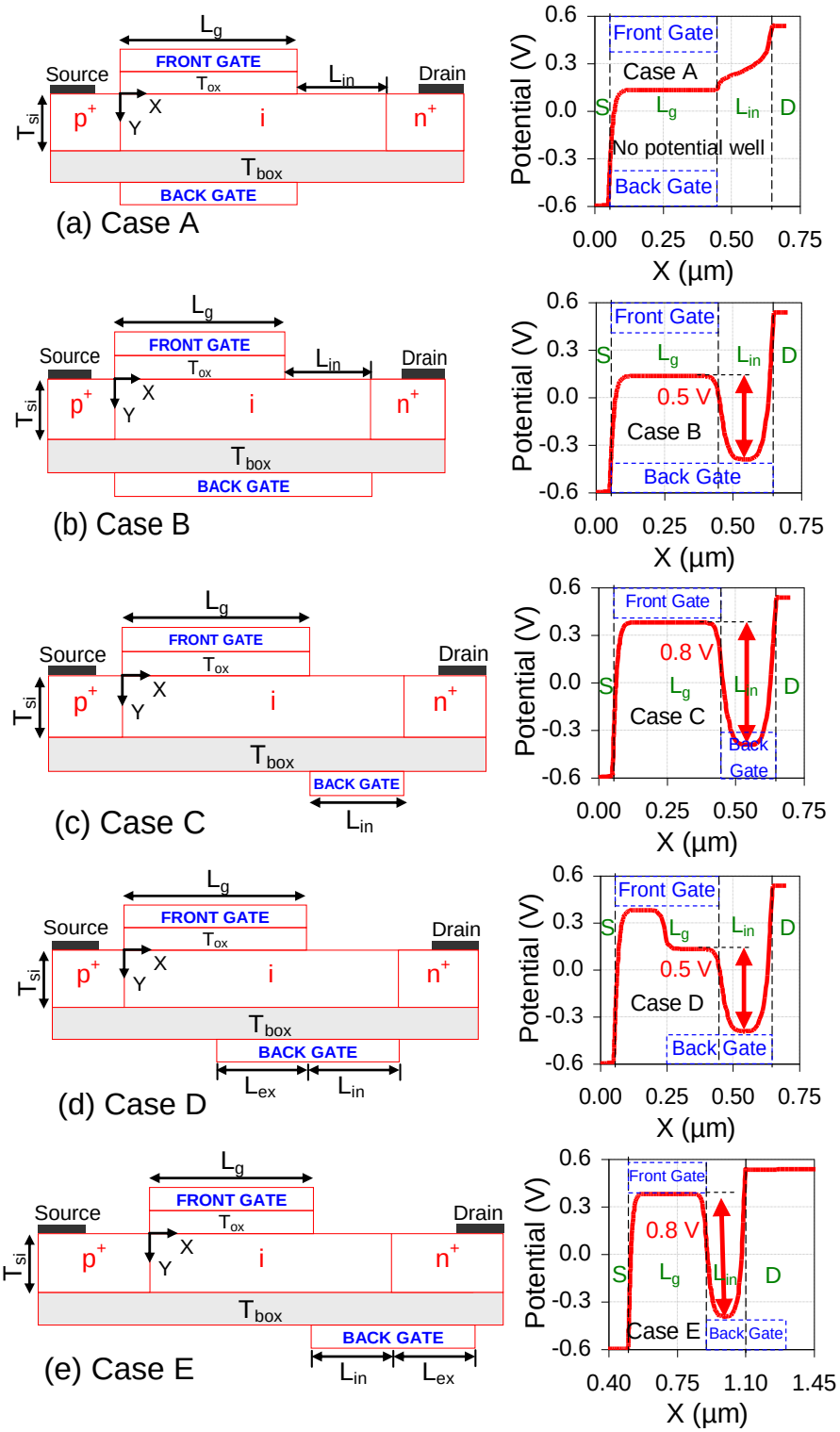

Fig. 2.10 Schematic diagrams for various possibilities of back gate alignment in UTBOX TFET with their respective electrostatic potential. The back gate is aligned at (a) position symmetric to front gate (Case A) (b) complete intrinsic channel region (Case B) (c) front un-gated region (Case C) underlap region elongated partially towards (d) front gate (Case D) and (e) drain (Case D).

Fig. 2.10(c) shows a completely misaligned back gate (front un-gated region: Case C). Case C results into more profound well (~0.8 V) as compared to back gate located at complete intrinsic region.Conventionally, the back gate covering the complete intrinsic region ($L_g + L_{in}$) as suggested by Biswas *et al.* [41] results in a potential well of depth 0.5 V (Case B) ranging from 0.1 V to -0.4 V. The potential depth (ΔV) can be expressed as in equation 2.4,

$$\Delta V = \Phi_{L_{g\_back}} - \Phi_{L_{in\_back}} \tag{2.4}$$

where, Φ_{Lg_back} and Φ_{Lin_back} are the potential at the back surface of the front gated (L_g) and front ungated (L_{in}) region, respectively. The front gate being n^+ poly (~4.2 eV) creates a high potential region (`0.4 V) in the film corresponding to its location, while back gate being p^+ poly (~5.2 eV) forms a low potential region (~ -0.4 V). The overlapping of these two gates, result into a potential of ~0.1 V due to coupling. The same is reflected in Table 2.4.

Table 2.4 Potential depth (ΔV) with varying gate positions

Gate position	Φ_{Lg_back}	Φ_{Lin_back}	ΔV	***Case***
Front and back gate symmetric	~ 0.1 V	No gate	0 V	A
Front and back gate overlap (partial)	~ 0.1 V	~ -0.4 V	~ 0.5 V	B,D
Front and back gate misaligned completely	~ 0.4 V	~ -0.4 V	~ 0.8 V	C,E

Figs. 2.10(d) and (e) indicates the potential well depth when the back gate extends at a partial region (L_{ext} = 200 nm) of front gate (Case D) and drain (Case E) showing depth of 0.5 V and 0.8 V, respectively. Thus, back gate alignment is an essential component to evaluate the profoundness of physical well, which highlights the back gate positioned with no overlap with the front gate region (Case C and E) results into a deeper storage region. The retention time for all different back gate positions, illustrated in Fig. 2.10 have been evaluated in Fig. 2.11 with optimized bias set, and it confirms the deeper potential well (Case C and E) result into enhanced retention characteristics (RT = 170 ms). However, Case E with back gate extended towards drain results into an increased tunneling during hold ‘0’ for a more negative bias and thinner back gate oxide. This degrades state ‘0’ at a quicker rate and hence, the retention time.

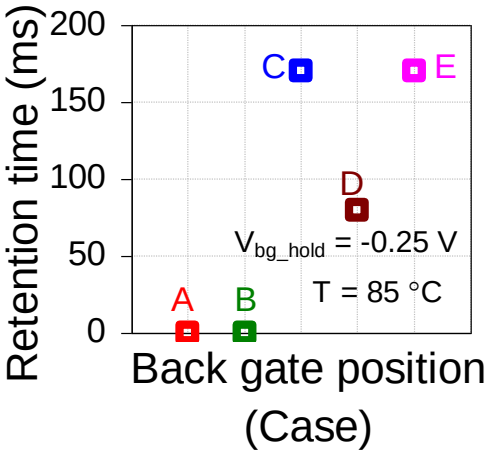

Fig. 2.11 *RT* as a function of back gate position for different cases in Fig. 2.10.

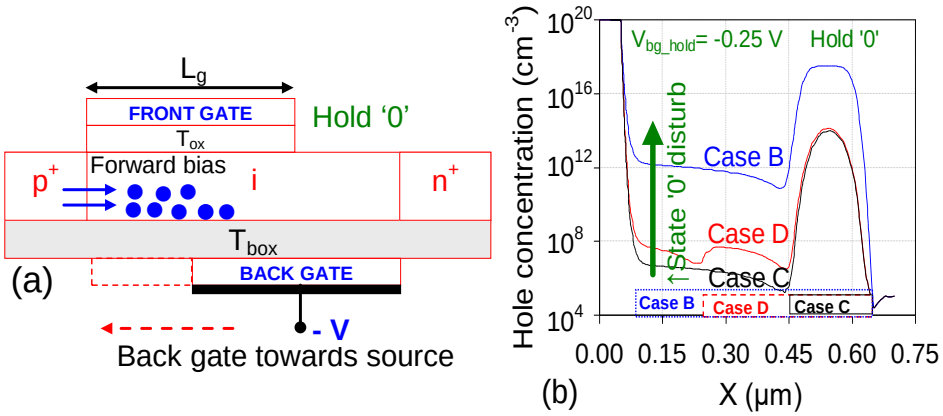

Fig. 2.12 (a) Schematic illustration of UTBOX TFET during hold '0'. (b) Variation in hole concentration as a function of back gate alignment with position shifting towards source along *X*.

Although for Case B and D, potential well depth is same (0.5 V) the retention time is different due to variation in the extent of alignment with front gate region. Case D with back gate positioned at $L_{in} + L_{ext}$ has partial overlap with front gate region (L_{ext}) that result into higher *RT* (in ms) than Case B (in μs) with back gate positioned at $L_g + L_{in}$. This is governed by the hole generation during hold '0'. A negative back gate bias is applied during Hold operation that forward biases the p^+/i region and holes are accumulated in the storage region (Fig. 2.12(a)). A higher degree of back gate alignment towards source result into increased hole generation in the back gate region. This is evident from Fig. 2.12(b) that shows increased hole concentration with increase in the length of back gate towards source region. The above observations clearly support the choice of the position of the back gate in the present work. Further analysis is based on the design

shown in Case C with complete back gate misalignment with respect to front gate positioned at intrinsic region.

The possibility of back gate misalignment for dual bit operation exhibiting retention characteristics was also shown in Z^2-FET [37]. While gate misalignment is often considered as a drawback for device design due to loss of gate controllability, the present work highlights the potential benefit of the back gate misalignment (position) in TFET which results into creation of deeper well which is advantageous in improving the retention time of the proposed DRAM. Intentional misalignment between the front and back gates can be obtained by shifting the electrical vernier as demonstrated by Widiez *et al.*[68]. Using the same approach, it is possible to shift back gate as a whole with the assistance of the electron beam position accuracy.

2.4.5 Impact of back gate workfunction

Apart from location, the workfunction of the back gate is also crucial. A p^+ poly back gate (~5.25 eV) accumulates holes in the storage region. The use of a p^+ poly back gate have been utilized in dynamic memory to enhance the retention time, [18,25] and also to form the injection barriers, [34] used as potential well. The present work uses a p^+ poly gate to create as well as maintain the physical well for charge storage and to replace the conventional TFET based DRAM that utilized a *p*-doped region as the storage area [42]. Therefore, without the need of conventional pocket implantation or doping the *p*-region, a hole concentration of ~ 10^{17} cm^{-3} can be obtained by the use of a p^+ poly back gate. While for an n^+ poly gate, the hole concentration in the underlap region is very low (~10^6 cm^{-3}).

Fig. 2.13(a) shows that a p^+ poly gate forms a potential well of depth 0.8 V whereas an n^+ poly gate (4.15 eV) doesn't form the potential well at the back. Thus, a p^+ poly back gate is opted to have a deeper storage region that could facilitate longer retention, while an n^+ poly front gate is selected to control the tunneling mechanism during read operation. A higher workfunction of the back occurs into an increased barrier height over the L_{in} region. Fig. 2.13(b) shows an injection barrier of 0.8 eV for p^+ poly gate while an n^+ poly gate has an energy

barrier of only 0.1 eV. This would result into lower current values for p^+ poly gate, affecting the sense margin.

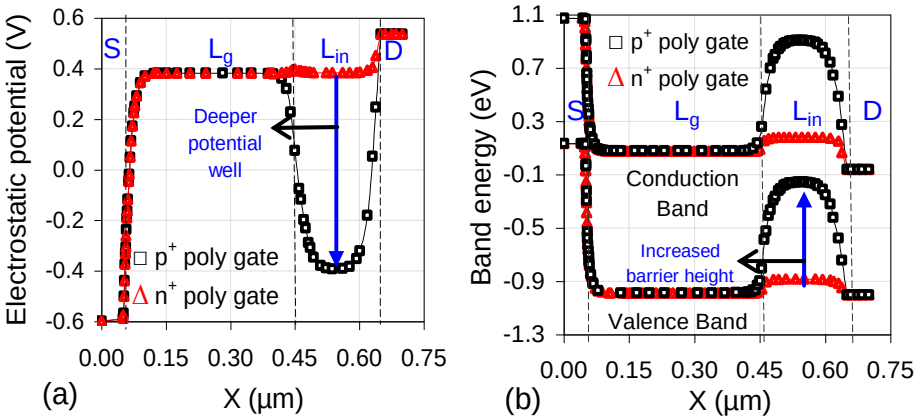

Fig. 2.13 Variation of (a) electrostatic potential, and (b) energy band profile along X for n^+ poly and p^+ poly back gates. The cutline for (a) is taken at a cross-section 3 nm above the back interface. The cutline for (b) is taken at a cross-section 3 nm below the front interface, where conduction occurs.

The same has been verified using the topology being proposed previously [41,42] having back gate at complete intrinsic region and box thickness of 145 nm (Fig. 2.14(a)). A higher back gate workfunction (p^+ poly) results into a deeper potential (Fig. 2.14(b)) that preserve charges for longer duration, and hence, results into improved retention time of 20 ms as compared to n^+ poly back gate with RT of 2 ms (Fig. 2.14(c)). The storage region serves as the barrier for electrons during read operation, and, therefore, the deeper potential well shows increased barrier that reduces the read currents for state '0' and state '1', thereby resulting into reduced sense margin (Fig. 2.14(d)).

Although sense margin degrades for p^+ poly back gate, the current ratio is well maintained above 10^2 that shows an acceptable read sensitivity. The results also demonstrate, a trade-off between sense margin and current ratio and also, between sense margin and retention time. Thus, the significance of back gate workfunction and intentional misalignment has been justified for charge retention in TFET based DRAM which needs to be accompanied with optimized bias values to achieve an improved performance. Thus, a p^+ poly back gate positioned at front un-gated region is selected for further investigation.

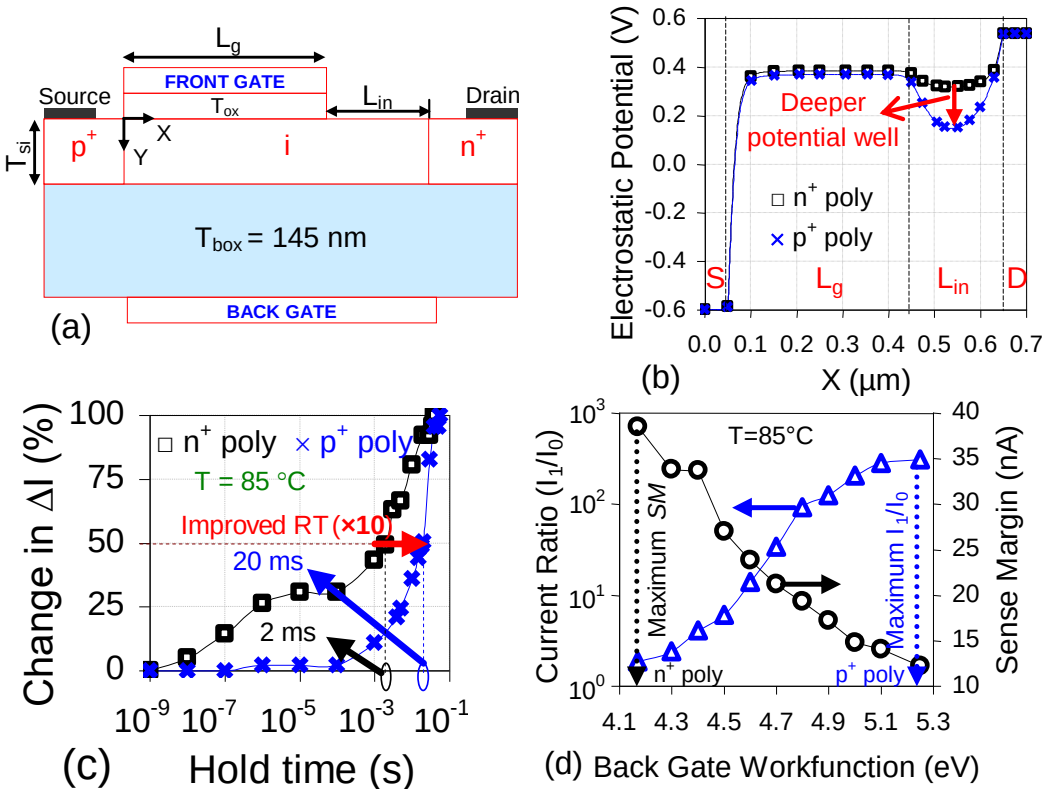

Fig. 2.14 (a) Schematic diagram of TFET as DRAM. (b) Variation of electrostatic potential with back gate workfunction along *X*. (c) Change in ΔI with time where 50% variation indicates *RT*. (d) Dependence of *SM* and *CR* on the back gate workfunction. Bias scheme: write '1': V_{bg} = -15 V, write '0': V_{bg} = 15 V, hold: V_{bg} = -3 V, read: V_{bg} = -1.5 V, V_{fg} = 4 V, V_d = 1.5 V.

2.5 Improved retention time using Double Gate TFET

As mentioned before, the back gate is responsible for the formation of the storage region. Therefore, DG TFET (Fig. 2.3(b)) having a thinner back oxide (3 nm) with HfO_2 as gate dielectric has a higher influence of back gate over the storage region as compared to the UTBOX DG FDSOI and permits use of lower voltages [47]. This results into a deeper potential well for DG architecture (Fig. 2.15(a)). Therefore, for write '1' operation, the bias applied at the back gate is -2 V that shows hole concentration of 3×10^{18} cm^{-3}, whereas for an UTBOX TFET a back gate bias of -5 V resulted into hole concentration of 10^{18} cm^{-3} (Fig. 2.7(b)). The underlap region with a larger storage capacitance results into a higher hole concentration (~10^{18} cm^{-3}) for DG architecture as compared to UTBOX TFET (~ 10^{17} cm^{-3}) at zero bias condition (Fig. 2.15(b)). Thus, a deeper potential well with a higher hole concentration retains the charges for longer duration, and therefore,

a higher retention time could be achieved by the double gate architecture. The programming scheme for the DG architecture is shown in Table 2.5.

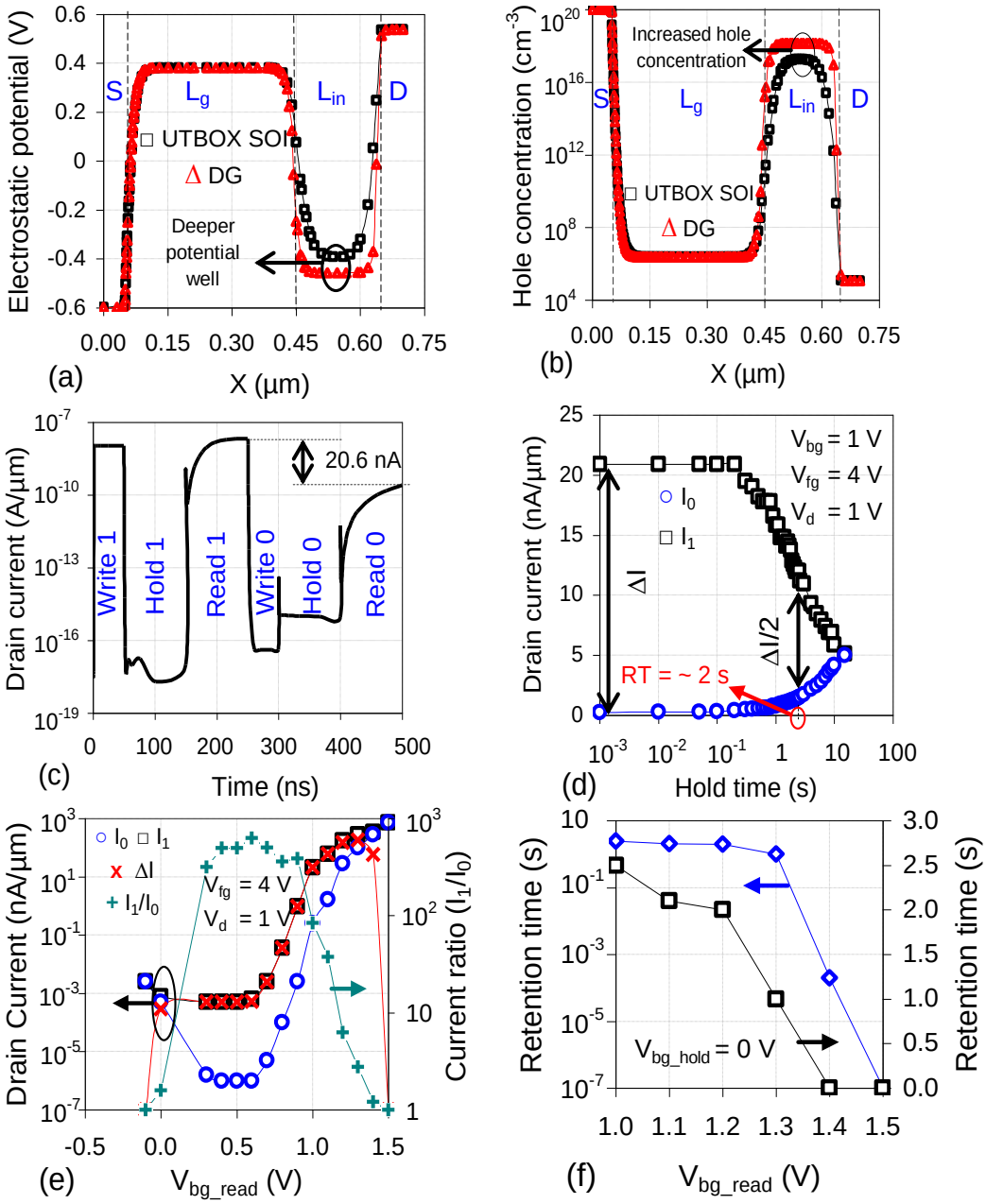

Fig. 2.15 Variation of (a) electrostatic potential, and (b) hole concentration at zero bias condition along *X* for UTBOX SOI and DG TFET. (c) Drain current transients for DG TFET in the sequence of the operation. (d) I_1 and I_0 showing a *RT* of ~2 s. Dependence of (e) I_1 and I_0, *SM* and *CR*, and (f) *RT* on the back gate bias applied during read operation. The cutlines for (a) and (b) are taken at a cross-section 3 nm above the back interface.

The read operation is performed with the same set of bias values as used for UTBOX FDSOI TFET. As shown in Figs. 2.15(c) and 2.15(d) the sense margin obtained is 20 nA and the *RT* achieved is ~ 2 seconds. A higher sense margin could be achieved by optimizing the back gate voltage during read operation. As shown in Fig. 2.15(e) the back gate bias in the range of 1 V-1.4 V can be applied to achieve an acceptable sense margin (in nano-Amperes) for TFET based DRAM. A maximum *SM* of ~ 170 nA is achieved with a back gate bias of 1.3 V during read operation. Although the sense margin is higher in the voltage range of 1.1 V-1.4 V, there is a gradual degradation in the retention time in this voltage range as compared to the back gate bias of 1 V applied during the read operation (Fig. 2.15(f)). Thus, the *SM* obtained with a back gate voltage of 1 V is low (~ 20 nA) but the current ratio (I_1/I_0) obtained is ~ 10^2 (Figs. 2.15(c) and 2.15(e)) and also, the retention time attained is high (~ 2 s) (Figs. 2.15(d) and 2.15(f)). Thus, a DG TFET structure provides a larger storage capacitance that maintains the charges for a longer period, and hence, improves the *RT*.

Table 2.5 Programming Scheme for DG TFET based DRAM

Operation	V_d	V_{fg}	V_{bg}	***Time***
Write 1	0 V	0 V	-2 V	50 ns
Write 0	0 V	0 V	2 V	50 ns
Read	1 V	4 V	1 V	50 ns
Hold	0 V	0 V	0 V	-

2.6. Other parameters influencing DRAM retention

This section introduces various factors, other than device architecture and bias that influence the retention characteristics of DRAM. The device parameters that affect retention include doping, gate length and operating temperature.

2.6.1. Impact of carrier lifetime

The temperature (equation 2.2), material, and the doping (equation 2.3) affects the carrier lifetime, which impacts the recombination rate [50,63]. Doping dependent carrier lifetime (equation 2.3) shows an intrinsic channel (10^{15} cm^{-3}) doesn't degrade carrier lifetime and thus, is beneficial for attaining longer charge

sustenance [34]. For silicon as channel material, previous literatures have reported carrier lifetime in the range of 10 ns -1 μs [15,19,24,27,28,34,37]. The reduced carrier lifetime decreases state '1' retention due to increased recombination rate [63]. Although for the present work, the simulations are done with $\tau_{n0/p0}$ as 100 ns at 27 °C, the dependence of *RT* on the carrier lifetime in DG TFET with low $\tau_{n0/p0}$ of 10 ns at 85 °C, has been shown in the Fig. 2.16(a). This default values are further varied according to temperature [67] and doping by using Scharfetter relation [38,55] described in section 2.3. The carrier lifetime is also influenced by the traps [24], where the traps lead to decrease in carrier lifetime and thus, reduced retention. This affect can also be observed through use of lower carrier lifetime than that proposed in previous works [15,24], in the range of μs. A degraded *RT* of ~ 350 ms with $\tau_{n0/p0}$ of 10 ns in comparison to 2 sec with $\tau_{n0/p0}$ = 100 ns (Fig. 2.15(d)) is achieved, but it is still higher than 64 ms. This shows a significant influence of carrier lifetime on *RT*, where an order decrease in $\tau_{n0/p0}$, reduces *RT* by ~ 6 times.

2.6.2. Impact of gate length scaling

As shown in the Fig. 2.16(b) TFET operates successfully as DRAM with *RT* > 64 ms [8], with the front gate region (L_g) and back gate region (L_{in}) scaled down to 100 nm. Further scaling of L_g and L_{in} degrade the performance of the dynamic memory. Scaling modifies the barriers formed by the front gate (n^+ poly) and back gate (p^+ poly) that regulates *SM* and *RT*. The back gate region (L_{in}) serves as the physical well for the charge storage. Therefore, its scaling reduces the storage area, and thus, the storage capacitance (Fig. 2.16(c)) which decreases the retention time. This is evident from Fig. 2.16(b) that shows the degradation of *RT* from few seconds (L_{in} = 200 nm) to few ms (L_{in} = 50 nm) with reduced storage area.

The downscaling of L_g results in the formation of a narrow potential barrier (Fig. 2.16(d)). However, with the scaling of L_g till 100 nm, DRAM qualitatively performs similar to the longer devices (L_g = 400 nm) (Fig. 2.16(b)). Further Fig. 2.16(d) demonstrates downscaling L_g below 100 nm, decreases the barrier height [39] that results in decrease in the depth of potential well, and thus, reduces the *RT*. Also, the reduced gate controllability due to narrow barrier affects the read mechanism and degrades the *SM*. The decrease in effective gate length increases

the read currents with significant impact on read current of state '0' more significantly, and thus, reduces *SM*. The influence of barrier narrowing or SCE that decreases the energy barrier between the front and back gates has been explained in detail in the subsequent chapters. Thus, the scalability of the TFET DRAM depends on the control of the regions formed by the gates and considering *RT* > 64 ms as criteria, both the gates can be scaled down till 100 nm.

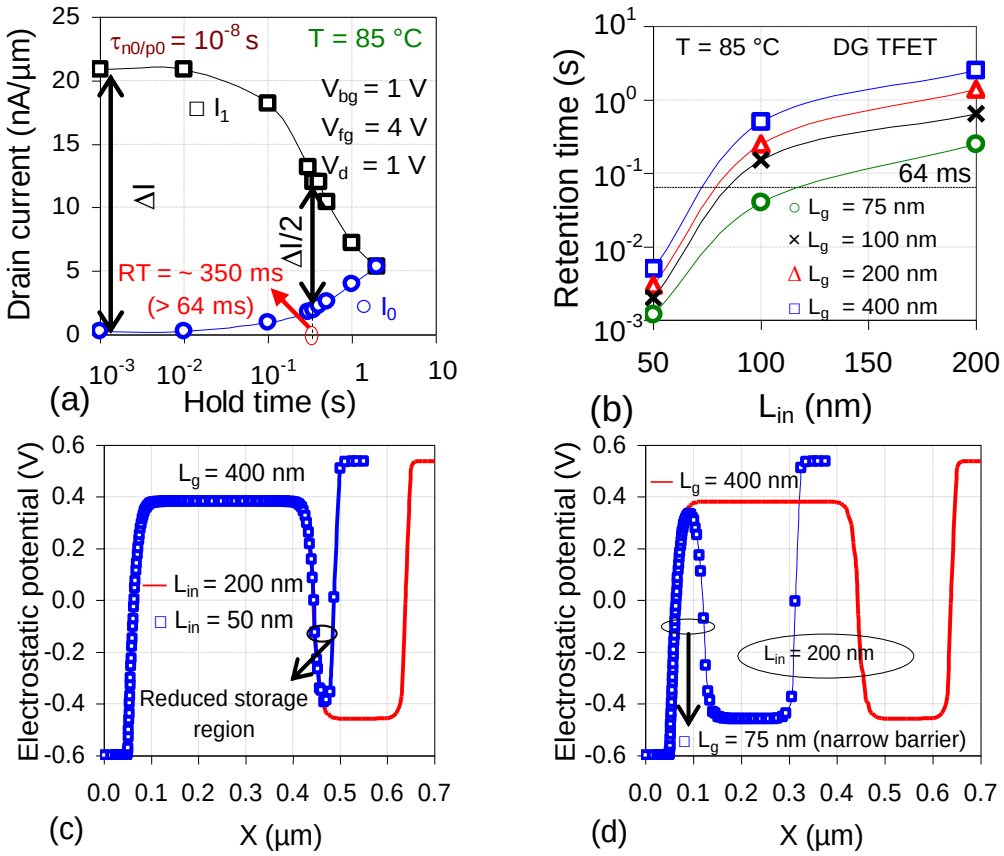

Fig. 2.16 (a) I_1 and I_0 showing *RT* of ~ 350 ms, estimated with $\tau_{n0/p0}$ = 10 ns. (b) Dependence of *RT* on storage region (L_{in}) for various values of front gate lengths (L_g). Variation of electrostatic potential as a function of (c) L_{in}, and (d) gate length (L_g) along the channel direction (*X*).

2.6.3 Impact of operating temperature

The generation/recombination increases with temperature that degrades charge retention [39,50,63]. The functionality of DG TFET DRAM at 125 °C is similar to that at 85 °C, but its performance deteriorates. This is evident from Fig. 2.17(a) which shows decrease in retention time to 120 ms at 125 °C from 2 seconds at 85 °C. The reduction in *RT* at higher temperatures of DRAM has also been observed

in the conventional MOSFET structures [13-30] and other p^+-i-n^+ structures [34-40]. For DG TFET based DRAM, although the performance degrades, the retention time attained is still higher than 64 ms. Moreover, the advantage of utilizing TFET device with weak temperature dependency is reflective in sense margin, which will be discussed in detail in the later part of thesis.

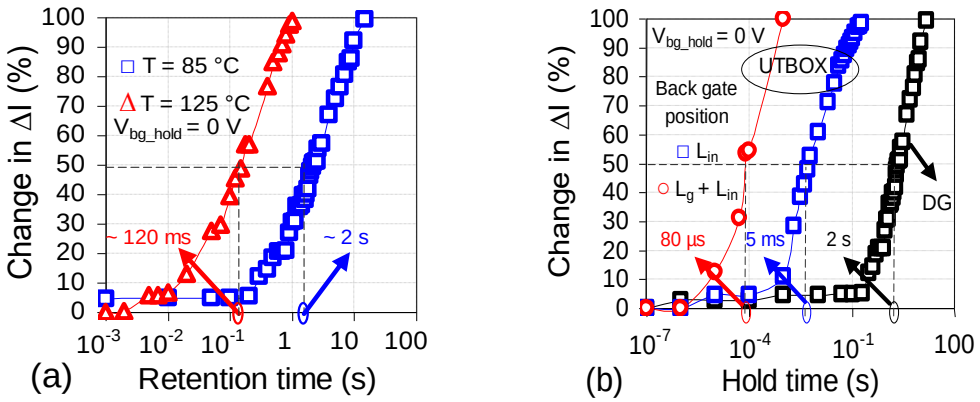

Fig. 2.17 Variation in the *RT* as a function of (a) temperature, and (b) device architecture (UTBOX with two different locations of the back gate, and DG) estimated when maximum sense margin (ΔI) is changed by 50%.

2.7 Comparative analysis

Fig. 2.17(b) shows the comparison of conventional UTBOX FDSOI (back gate is across L_g + L_{in} region) misaligned UTBOX and DG TFET (back gate is across L_{in} region) highlighting the transition of retention time from ~ 80 µs to ~ 2 s, from conventional TFET (UTBOX) to misaligned DG TFET. For conventional UTBOX TFET, along with thermal generation and tunneling at the drain junction, holes are accumulated in the storage region due to forward biased p^+-i junction leading to rapid degradation of state '0'. Therefore, TFET device with deeper potential well (misaligned p^+ poly back gate) is implemented which increases retention time to 5 ms, with the same bias values. Further, the retention characteristics can be improved with DG architecture.

The results, also demonstrate higher retention capability than similar architecture that includes FED [34] and Z^2-FET [37]. While use of Z^2-FET [34] for a total length (L_g+ L_{in}) of 600 nm report retention time of ~5.5 s at 27 °C, FED with total length of 840 nm show retention of ~1 s at 27 °C, use of TFET have also, resulted

in charge sustenance for ~2 s at 85 °C (Fig. 2.18). However, few issues that need to be resolved for tunneling based read mechanism are low sense margin and use of higher bias values. Although, sense margin is low, high current ratio compensates for it. The work explores the possibility of exploiting TFET further as dynamic memory, highlighting its capability to compete with other similar architectures. The flow for analysing and optimizing the operation as DRAM can be summarized as in Fig. 2.19.

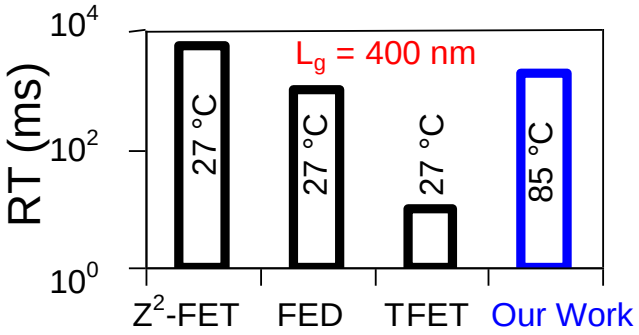

Fig. 2.18 Comparison of *RT* of various p^+-*i*-n^+ based architectures (FED [34], Z^2-FET [37], TFET [42]) with a gate length of 400 nm with the proposed misaligned DG TFET architecture (our work).

2.8 Conclusion

An in-depth analysis of different attributes of FDSOI TFET governing the DRAM characteristics has been presented. The results have demonstrated that through an insightful optimization of the position, workfunction and bias values of back gate, a retention time higher than the target of 64 ms can be achieved in DG tunnel FETs. The n^+ poly front gate primarily controls the read mechanism based on band-to-band tunneling. The p^+ poly back gate is intentionally misaligned away from the front gate and positioned at the front un-gated region near the drain to create a deeper potential well for the charge storage. The deeper physical well results into prolonged storage of charges, thereby increasing the retention time which can be further improved by the selection of optimal back gate bias. The retention time of ~ 170 ms is attained for an UTBOX FDSOI TFET, which can be enhanced through a DG TFET structure that operates similar to UTBOX FDSOI TFET and achieves a retention time of ~ 2 seconds at a temperature of 85 °C.

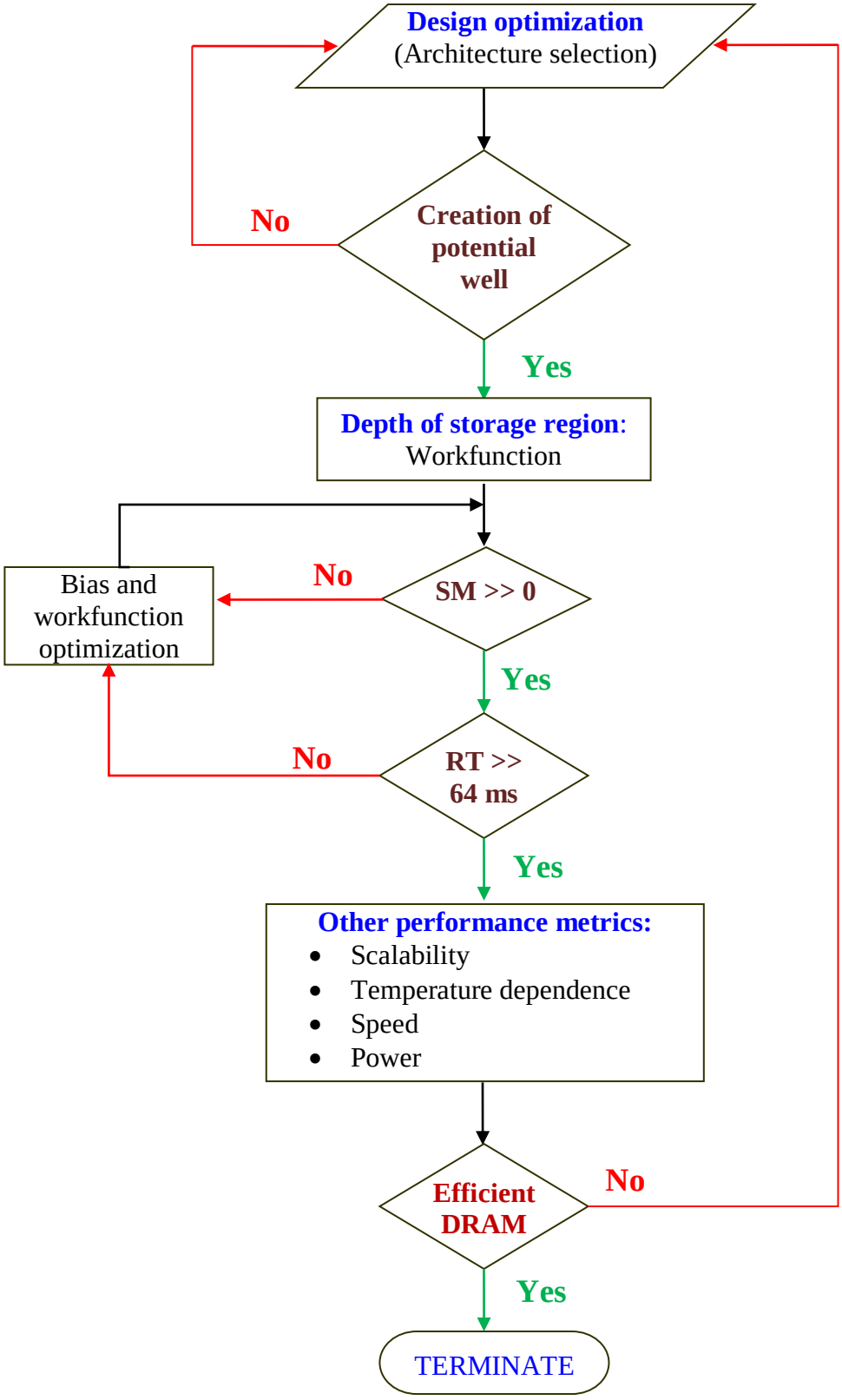

Fig. 2.19 Flow chart demonstrating a probable methodology, to evaluate device performance as dynamic memory.

The proposed optimized TFET based DRAM has performed better in terms of retention time as compared to the previously reported (in ms at room temperature) DRAMs utilizing TFET. The work highlights the opportunity in the design of an optimized TFET for dynamic memory with improved retention characteristics, achieved through back gate engineering. Further evaluation is based on enhancing other DRAM performance metrics that includes device scalability, sense margin, current ratio, speed and power which will be discussed in subsequent chapters.

2.9 References

[1] Keeth B., Baker J. R., Jacob B., Lin F., (2007) DRAM circuit design: fundamental and high-speed topics 2nd ed. John Wiley & Sons Inc., New Jersey (ISBN- 978-0-4701-8475-2).

[2] Joodaki, M., (2012) Selected advances in nanoelectronic devices: logic, memory and RF, Springer Science and Business Media, Heidelberg (ISBN-978-3-642-31349-3).

[3] Mutlu O., (2013) Memory scaling: A systems architecture perspective. IEEE International Memory Workshop, pp. 21-25.

[4] Bhati I., Chang M. T., Chishti Z., Lu S. L., and Jacob B., (2016) DRAM refresh mechanisms, penalties, and trade-offs, IEEE Transactions on Computers, 65, 108-121.

[5] Ohsawa, T., Hamamoto, T., (2012) A Novel Capacitor-less DRAM: Floating Body Cell, CRC Press, Taylor & Francis Group, Florida (ISBN: 978-9-81430-308-8).

[6] Itoh K., (2008) The history of DRAM circuit designs–at the forefront of DRAM development, IEEE Solid-State Circuits Society Newsletter, 13, 27-31.

[7] Kim K., (2000) Perspectives on giga-bit scaled DRAM technology generation, Microelectronics reliability, 40, 191-206.

[8] ITRS International Technology Working Groups. International Technology Roadmap for Semiconductors, http://www.itrs.net (2015).

[9] Lee S. H., (2016) Technology scaling challenges and opportunities of memory devices, IEEE Electron Devices Meeting, pp. 1-8.

[10] Hassan H., Vijaykumar N., Khan S., Ghose S., Chang K., Pekhimenko G., Lee D., Ergin O., and Mutlu O., (2017) SoftMC: A flexible and practical open-source infrastructure for enabling experimental DRAM studies, IEEE International Symposium on High Performance Computer Architecture, pp. 241-252.

[11] Sharroush S. M., (2016) Performance optimization of 1T-1C DRAMs: A quantitative study, Microelectronics Journal, 52, 147-164.

[12] Faul J. W., and Henke D., (2005) Transistor challenges–A DRAM perspective. Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 237, 228-234.

[13] Giusi G., Alam M. A., Crupi F., and Pierro S., (2010) Bipolar mode operation and scalability of double-gate capacitorless 1T-DRAM cells, IEEE Transactions on Electron Devices, 57, 1743-1750.

[14] Okhonin S., Nagoga M., Carman E., Beffa, R., and Faraoni, E., (2007) New generation of Z-RAM, IEEE Electron Devices Meeting, pp. 925-928.

[15] Kuo C., King T. J., and Hu C., (2002) A capacitorless double-gate DRAM cell, IEEE Electron Device Letters, 23, 345-347.

[16] Bawedin M., Cristoloveanu S. and Flandre D., (2007) Innovating SOI memory devices based on floating-body effects, Solid-State Electronics, 51, 1252-1262.

[17] Kuo C., King T. J., and Hu C., (2003) A capacitorless double gate DRAM technology for sub-100-nm embedded and stand-alone memory applications, IEEE Transactions on Electron Devices, 50, 2408-2416.

[18] Sasaki K. R. A., Nicoletti T., Almeida L. M., dos Santos S. D., Nissimoff A., Aoulaiche M., Simoen E., Claeys C., and Martino J. A., (2014) Improved retention times in UTBOX nMOSFETs for 1T-DRAM applications, Solid-State Electronics, 97, 30-37.

[19] Lu Z., Fossum J. G., and Zhou Z., (2011) A floating-body/gate DRAM cell upgraded for long retention time, IEEE Electron Device Letters, 32, 731-733.

[20] Han J.-W., Ryu S.-W., Kim D.-H., Kim C.-J., Kim S., Moon D.-I., Choi S.-J., and Choi Y.-K., (2009) Fully depleted polysilicon TFTs for capacitorless 1T-DRAM, IEEE Electron Device Letters., 30, 742–744.

[21] Giusi G., (2014) Physical insights of body effect and charge degradation in floating-body DRAMs, Solid-State Electronics, 95, 1-7.

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

[23] Fossum J. G., Lu Z., and Trivedi V. P., (2007) New insights on "capacitorless" floating-body DRAM cells, IEEE Electron Device Letters, 28, 513-516.

[24] Cho M. H., (2012) Thin-Body SOI Capacitorless DRAM Cell Design Optimization and Scaling. University of California, Berkeley, Technical Report No. UCB/EECS-2012-176.

[25] Ranica R., Villaret A., Fenouillet-Beranger C., Malinge, P., Mazoyer P., Masson P., Delille D., Charbuillet C., Candelier P., and Skotnicki T., (2004) A capacitor-less DRAM cell on 75nm gate length, 16 nm thin fully depleted SOI device for high density embedded memories, IEEE Electron Devices Meeting, pp. 277-280.

[26] Almeida L. M., Sasaki K. R. A., Caillat C., Aoulaiche M., Collaert N., Jurczak M., Simoen E., Claeys C., and Martino J. A., (2013) Optimizing the front and back biases for the best sense margin and retention time in UTBOX FBRAM, Solid-State Electronics, 90, 149-154.

[27] Kim S., Choi S. J., Moon D. I., and Choi Y. K., (2012) Carrier lifetime engineering for floating-body cell memory, IEEE Transactions on Electron Devices, 59, 367-373

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

[29] Aoulaiche M., Collaert N., Degraeve R., Lu Z., De Wachter B., Groeseneken G., Jurczak M., and Altimime L., (2010) BJT-mode endurance on a 1T-RAM bulk FinFET device, IEEE Electron Device Letters, 31, 1380-1382.

[30] Collaert N., Aoulaiche M., Rakowski M., Wachter B. De, Bourdelle K., Nguyen B. Y., Boedt F., Delprat D. and Jurczak M. (2009), Analysis of sense margin and reliability of 1T-DRAM fabricated on thin-film UTBOX substrates, IEEE International SOI Conference, pp. 1.

[31] Nemati F., and Plummer J.D., (1999) A novel thyristor-based SRAM cell (T-RAM) for high-speed, low-voltage, giga-scale memories, IEEE Electron Devices Meeting, pp. 283-286.

[32] Cho H.-J., Nemati F., Roy R., Gupta R., Yang K., Ershov M., Banna S., Tarabbia M., Salling C., Hayes D., and Mittal A., (2005) A novel capacitor-less DRAM cell using thin capacitively-coupled thyristor (TCCT), IEEE Electron Devices Meeting, pp. 311-314.

[33] Mulaosmanovic H., Paolucci G. M., Compagnoni C. M., Castellani N., Carnevale G., Fantini P., Ventrice D., Lacaita A. L., Spinelli A. S., and Benvenuti A., (2014) Working principles of a DRAM cell based on gated-thyristor bistability, IEEE Electron Device Letters, 35, 921-923.

[34] Badwan A. Z., Chbili Z., Yang Y., Salman A. A., Li Q., and Ioannou D. E., (2013) SOI field-effect diode DRAM cell: Design and operation, IEEE Electron Device Letters, 34,1002-1004.

[35] Badwan A. Z., Li Q., and Ioannou D. E., (2015) On the nature of the memory mechanism of gated-thyristor dynamic-RAM Cells, IEEE Journal of the Electron Devices Society, 3, 468-471.

[36] Yang Y., Gangopadhyay A., Li Q., and Ioannou D. E., (2009) Scaling of the SOI field effect diode (FED) for memory application, IEEE Semiconductor Device Research Symposium, pp. 1-2.

[37] Wan J., Royer C. Le, Zaslavsky A., and Cristoloveanu S., (2013) Progress in Z^2-FET 1T-DRAM: retention time, writing modes, selective array operation, and dual bit storage, Solid-State Electronics, 84, 147-154.

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

[39] Dirani E. H., Solaro Y., Fonteneau P., Ferrari P., and Cristoloveanu S., (2016) Properties and mechanisms of Z^2-FET at variable temperature, Solid-State Electronics, 115, 201-206.

[40] Wan J., Zaslavsky A., and Cristoloveanu S., (2015) Comment on Investigation of tunnel field-effect transistors as a capacitor-less memory cell, Applied Physics Letters, 106, 016101.

[41] Biswas A., Dagtekin N., Grabinski W., Bazigos A., Le Royer C., Hartmann J. M., Tabone C., Vinet M., and Ionescu A. M., (2014) Investigation of

tunnel field-effect transistors as a capacitor-less memory cell, Applied Physics Letters, 104, 092108.

[42] Biswas A., and Ionescu A. M., (2015) 1T capacitor-less DRAM cell based on asymmetric tunnel FET design, IEEE Journal of the Electron Devices Society, 3, 217-222.

[43] Biswas A., and Ionescu A. M., (2014) Study of fin-tunnel FETs with doped pocket as capacitor-less 1T DRAM, IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference, pp. 1-2.

[44] Luryi S., Xu J., and Zaslavsky A., (2013) Future trends in microelectronics: Frontiers and Innovations, John Wiley & Sons (ISBN- 9781119069171).

[45] Zhang, L., and Chan, M., (2016) Tunneling Field Effect Transistor Technology, Springer International Publishing, Switzerland, (ISBN- 978-3-319-31651-2)

[46] Yeung C. W., Padilla A., Liu T. J. K., and Hu C., (2009) Programming characteristics of the steep turn-on/off feedback FET (FBFET), IEEE Symposium on VLSI Technology, pp. 176-177.

[47] Colinge, J.-P., (1997) Silicon-on-Insulator Technology: Materials to VLSI, third ed. Springer Science Business Media, New York, (ISBN-978-1-4613-4795-8).

[48] Fossum J. G., (1990) Analysis and control of BJT latch in fully depleted floating-body submicron SOI MOSFETs, IEEE SOS/SOI Technology Conference. Proceedings, pp. 21-22.

[49] Chen C. D., Matloubian M., Sundaresan R., Mao B. Y., Wei C. C., and Pollack G. P., (1988) Single-transistor latch in SOI MOSFETs, IEEE Electron Device Letters, 9, 636-638.

[50] Streetman, B. G., and Banerjee, S. K., (2009) Solid State Electronic Devices, sixth ed. Pearson Education Inc., New Jersey 07458, (ISBN-978-81 -203-3020-7).

[51] Salman A. A., Beebe S. G., Emam M., Pelella M. M., and Ioannou D. E., (2006) Field Effect Diode (FED): A novel device for ESD protection in deep sub-micron SOI technologies, IEEE Electron Devices Meeting, pp. 1–4.

[52] Raissii F., (1996) A brief analysis of the field effect diode and breakdown transistor, IEEE Transaction on Electron Devices, 43, 362–365.

[53] Tanaka T., Suzuki K., Horie H., and Sugii T., (1994) Ultrafast Operation of &-Adjusted p+-n+ Double-Gate SO1 MOSFET's, IEEE Electron Device Letters, 15, 386-388.

[54] Skotnicki T., Fenouillet-Beranger C., Gallon C., Boeuf F., Monfray S., Payet F., Pouydebasque A., Szczap M., Farcy A., Arnaud F., and Clerc S., (2008) Innovative materials, devices, and CMOS technologies for low-power mobile multimedia, IEEE Transactions on Electron Devices, 55, 96-130.

[55] Atlas User's Manual, Silvaco Int., Santa Clara, CA, Jan. 18, 2012.

[56] Wan J., Le Royer C., Zaslavsky A., and Cristoloveanu S., (2011) Tunneling FETs on SOI: Suppression of ambipolar leakage, low-frequency noise behavior, and modeling, Solid-State Electronics, 65, 226-233.

[57] Biswas A., Dan S.S., Le Royer C., Grabinski W., and Ionescu A. M., (2012) TCAD simulation of SOI TFETs and calibration of non-local band-to-band tunneling model, Microelectronic Engineering, 98, 334-337.

[58] Fukuda K., Mori T., Mizubayashi W., Morita Y., Tanabe A., Masahara M., Yasuda T., Migita S., and Ota H., (2013) A compact model for tunnel field-effect transistors incorporating nonlocal band-to-band tunneling, Journal of Applied Physics, 114,144512.

[59] Boucart K., and Ionescu A. M., (2007) Double-Gate Tunnel FET With High-K Gate Dielectric, IEEE Transactions on Electron Devices, 54, 1725-1733.

[60] Cho S.J., O'uchi S., Endo K., Kim S. W., Son Y. H., Kang I. M., Masahara M., Harris Jr J. S., and Park B. G., (2010) Rigorous design of 22-nm node 4-terminal SOI FinFETs for reliable low standby power operation with semi-empirical parameters, JSTS: Journal of Semiconductor Technology and Science, 10, 265-275.

[61] Kane E. O., (1961) Theory of tunneling, Journal of Applied Physics, 32, 83-91.

[62] Keldysh L. V., (1958) Behavior of non-metallic crystals in strong electric fields, Soviet Journal of Experimental and Theoretical Physics, 6, 763.

[63] Hurkx G. A. M., Klaassen D. B. M., and Knuvers M. P. G., (1992) A new recombination model for device simulation including tunneling, IEEE Transactions on Electron Devices, 39, 331-338.

[64] Schenk A., (1993) Rigorous Theory and Simplified Model of the Band-to-Band Tunneling in Silicon, Solid-State Electronics, 36, 19-34.

[65] Hurkx G. A. M., (1989) On the modelling of tunnelling currents in reverse-biased pn junctions, Solid-State Electronics, 32, 665-668.

[66] Li D., Zhang B., Lou H., Zhang L., Lin X., and Chan M., (2015) Comparative analysis of carrier statistics on MOSFET and tunneling FET characteristics, IEEE Journal of the Electron Devices Society, 3, 447-451.

[67] Schenk A., (1992) A model for the field and temperature dependence of Shockley-Read-Hall lifetimes in silicon, Solid-State Electronics, 35, 1585-1596.

[68] Widiez J., Lolivier J., Vinet M., Poiroux T., Previtali B., Daugé F., Mouis M., and Deleonibus S., (2005) Experimental evaluation of gate architecture influence on DG SOI MOSFETs performance, IEEE Transactions on Electron Devices, 52, 1772-1779.

Chapter 3

Sub-100 nm Misaligned Double Gate Tunnel FET based DRAM: Design Perspective for Improved Retention and Scalability

3.1 Introduction

The prediction of Moore has been credited with being the engine of the revolution in electronics [1-6], that led to components with improved speed and high density. Predictions using the law became the basis for future production goals, which in turn showcases the validity of the law as a measurement of industry progress [6]. DRAM and microprocessors were the two most crucial semiconductor products that followed the Moore's prediction. DRAM has been a critical component in semiconductor market due to utility as large repetitive cells with easier design, and thus, has been often used as indicators to reflect the progress in semiconductor industry [3]. In fact, in early 1980s, the Moore's curve was viewed as rule about the densities of dynamic memories [3].

DRAM being a stable technology was utilized for standardization as well as validation of the Moore's law. Thus, DRAM progress has been continuously fuelled with aggressive scaling, thereby following Moore's predictions. In the past decade, the problems associated with capacitor scaling in the conventional DRAMs [7-11], introduced capacitorless DRAM [12-14] with single transistor for charge storage as well as accessing the data stored. In real world, one of the most critical aspects is the low power operation that necessitates device scaling [3,15]. Therefore, it is essential to analyse the scaling capability of the DRAM cell while maintaining the retention time of 64 ms [6]. This facilitates the use of TFET as dynamic memory, which is focus of our work. The innovative design has been optimally structured as well as biased to function efficiently as dynamic memory.

Although, misaligning the front and back gates have shown the way forward to improve retention time, the applicability of TFET for DRAM is limited by the lower *RT* values, especially for sub-100 nm regime. Also, the use of a high front gate voltage during read operation in the misaligned DG TFET (chapter 2) requires voltage scaling.

The previous chapter had presented insights into device operation, highlighting the influence of back gate for functionality as DRAM. In this chapter, the role of each gate is discussed, focusing on methodologies to overcome limitation for scaling down that would serve as a design guideline in future. The chapter investigates on the scaling capability of memory, considering retention as the criteria. Through an in-depth analysis and design optimization, a systematic analysis shows the significance of total length as well as individual gate lengths (L_{g1} and L_{g2}) lateral spacing between the gates and underlap length (L_{un}) on two important parameters of DRAM, namely *SM* and *RT*. The work provides new opportunities to realize TFET based DRAM with improved *RT*, scalability and high temperature operation.

3.2 Design and operating principle

3.2.1 Device description

Fig. 3.1(a) illustrates the DG TFET based p^+-i-n^+ structure used in the analysis using Atlas [16] with parameters as introduced in chapter 2. The front gate (L_{g1}) and back gate (L_{g2}) are both 100 nm. The distinct functionality of the gates is the principle basis for operation of TFET based DRAM. The device design is further modified to achieve better DRAM characteristics by using a lateral spacing (L_{gap}) between the gates and an underlap (L_{un}) between the drain and back gate which can be implemented using a dual spacer process [17]. The operation of DG TFET based DRAM (Fig. 3.1(a)) is based on the storage of charges at the back gate capacitor, that serves as the potential well (Fig. 3.1(b)). The sustenance of charges stored in the physical well determines the retention time, where an increased storage region results into improved retention characteristics. This concept is implied by the use of a lateral spacing (L_{gap}) between the gates (Fig. 3.1(c)) designated as Case-II that increases the effective length of the front as well as

back gate region (Fig. 3.1(d)) thereby, contributing to larger storage capacitance. The idea is extended by including an underlap (L_{un}) between the back gate and drain (Fig. 3.1(e) and 3.1(f)) designated as Case-III that further increases the effective channel length associated with back gate, and thus, the retention time.

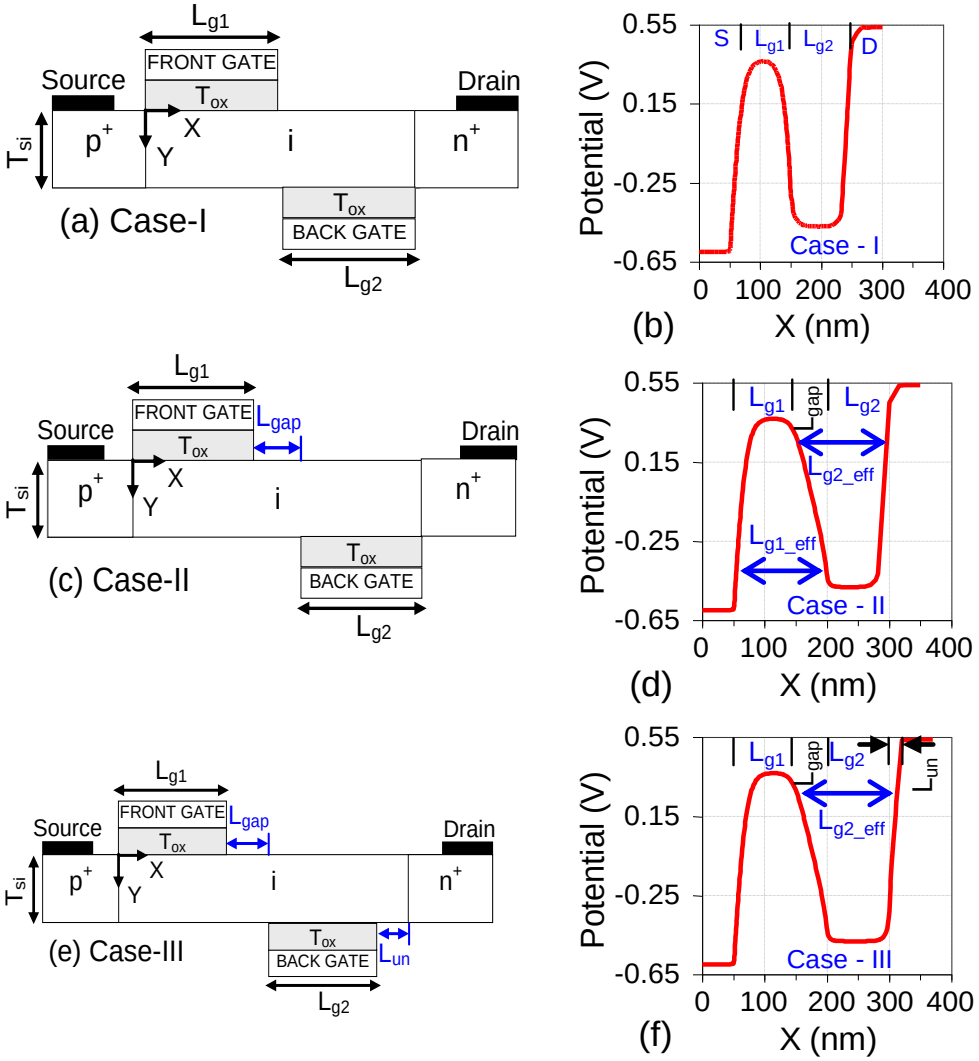

Fig. 3.1 Schematic diagram of a misaligned Double Gate (DG) TFET (a) without gap between the gates (L_{gap}) and an underlap (L_{un}) region between the back gate and drain (Case-I) (c) with L_{gap} (Case-II) (e) with L_{gap} and L_{un} (Case-III). Variation of electrostatic potential for (b) Case-I, (d) Case-II, and (f) Case-III. The cutline is taken at a cross-section 2 nm above the back interface for all the cases.

Other than storage capacitor being determined by device dimension, the retention time is also dependent on the applied bias and temperature that governs the leakage current due to thermal generation and BTBT between the gates and back gate/drain region [18-22] for state '0' (Fig. 3.2(a)) and recombination for state '1'

(Fig. 3.2(b)). Thus, the present work utilizes structural modification with optimized bias to increase the storage capacitance and reduce the leakage current that improves the retention time. The equivalent schematic representation of the device is as shown in Fig. 3.1(c)) where C_{G1}, and C_{G2} are the gate capacitance with storage capacitance, $C_s = C_{G2}$.

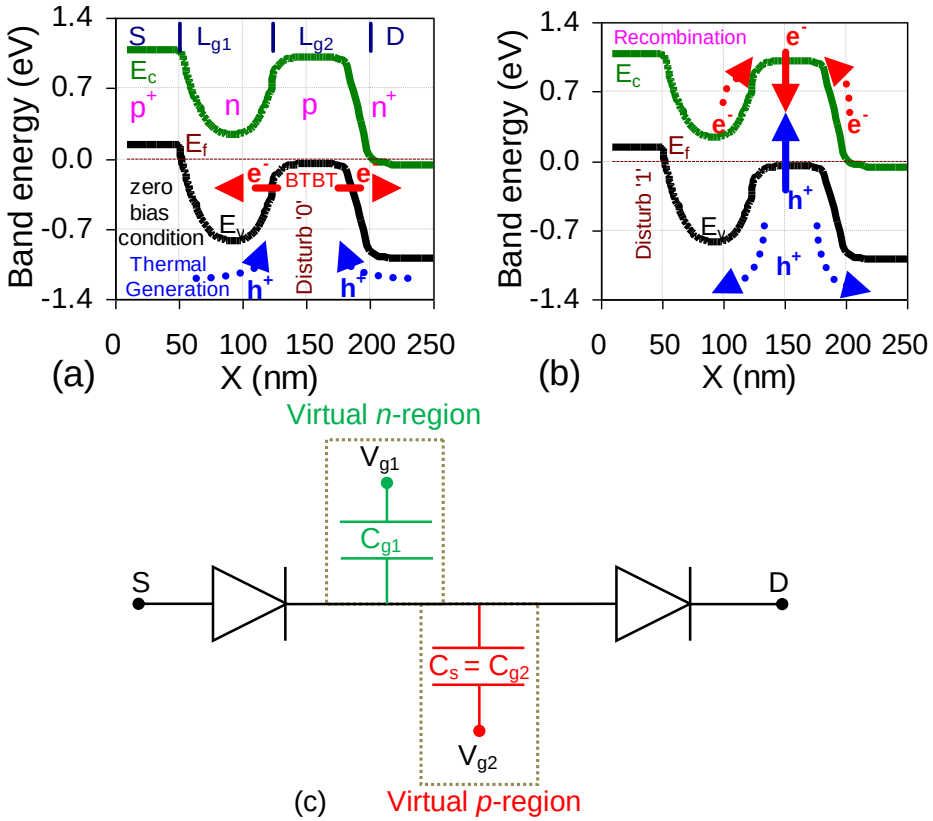

Fig. 3.2 Band energy demonstrating (a) disturb '0' due to hole generation in the storage region based on thermal generation and band-to-band tunneling of electrons towards drain, and (b) disturb '1' due to hole recombination based on thermal recombination and hole diffusion from the storage region. The band energy is at zero bias at cutline 1 nm above the back interface ($Y = T_{si}$). (c) The equivalent schematic representation of the misaligned DG TFET. C_{g1} and C_{g2} (= C_s) are the gate capacitances.

3.2.2 Operating Mechanism

The operating principle is similar to that described in chapter 2 with bias scheme as illustrated in Fig. 3.3. The figure illustrates the schematic representation of TFET device during write '1', where capacitance associated with back gate, serves as storage capacitance (C_s). This gate capacitance is similar to that

described in [23,24]. The change in the stored charge is reflective in the effective potential of the body ($\Delta V_b = \Delta Q_s / C_s$) [25-27], where Q_s is the stored hole charge. This impacts the read currents and thus, differentiates the two states. Fig. 3.4(a) reflects the same, where presence of higher hole concentration results into higher effective potential at complete channel region. Consequently, it shows a higher conduction density for read '1' compared to read '0' (Fig. 3.4(b)).

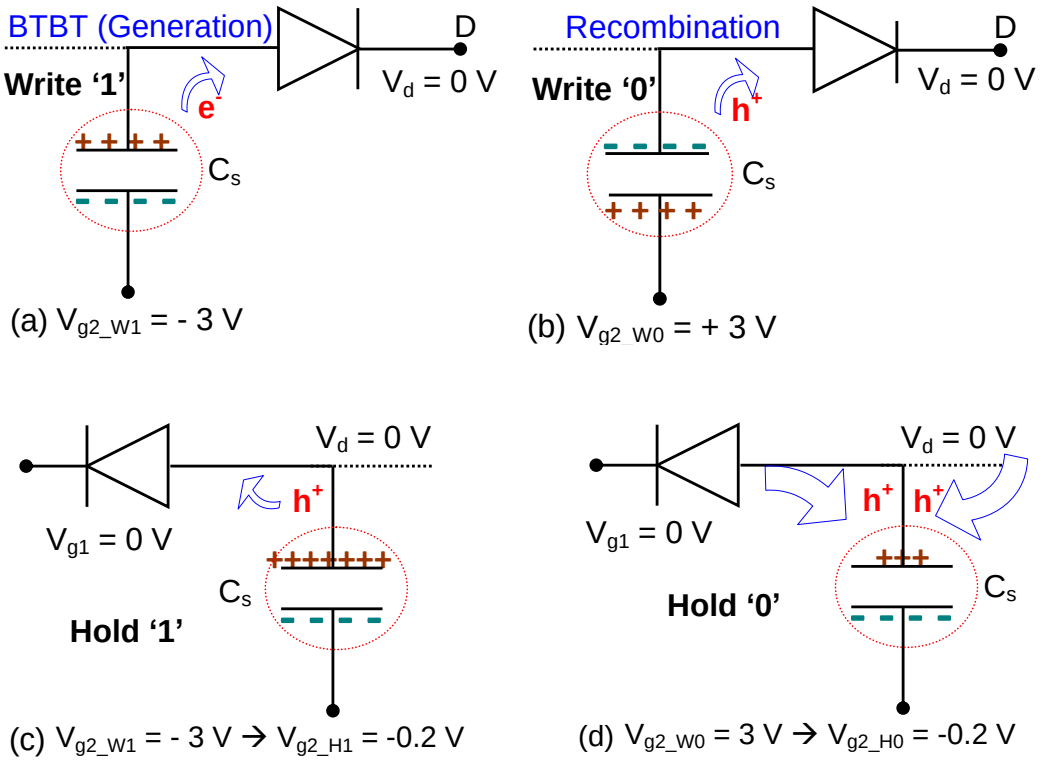

Fig. 3.3 Schematic illustration of DG TFET indicating (a) write '1' (W1) (b) write '0' (W0) (c) hold '1' (H1) (d) hold '0' (H0) through charging and discharging of the storage capacitor (C_s) associated with back gate.

The discussion demonstrates that the sense margin is estimated based on the positive charge stored in the capacitor associated with back gate during read operation. The charge present at read operation depends on (i) the charge stored in the capacitor in the previous operation and (ii) the bias applied in the previous as well as performed operation. The same has been demonstrated in Fig. 3.5 where the maximum difference in the charge stored between state '1' and state '0' is during write operation ($\Delta Q_W = Q_{W1} - Q_{W0}$) which decays during hold operation ($\Delta Q_H = Q_{H1} - Q_{H0;}$ $\Delta Q_W > \Delta Q_H$) due to hole recombination ($-\Delta Q_{hh1}$) for state '1' and hole generation ($+\Delta Q_{hh0}$) for state '0'. The maintenance of charges (Q_{H1} and

Q_{H0}) during hold determine the retention characteristics. Further during read, the hole concentration decreases for state '1' due to diffusion and thermal recombination and the charge difference observable during read ($\Delta Q_R = Q_{R1} - Q_{R0}$; $\Delta Q_W > \Delta Q_H > \Delta Q_R$) determine the sense margin. The choice of bias at back gate during read (1.2 V) is based on the maximum values of sense margin and retention time achieved, as shown in Figs. 3.6(a)-(c).

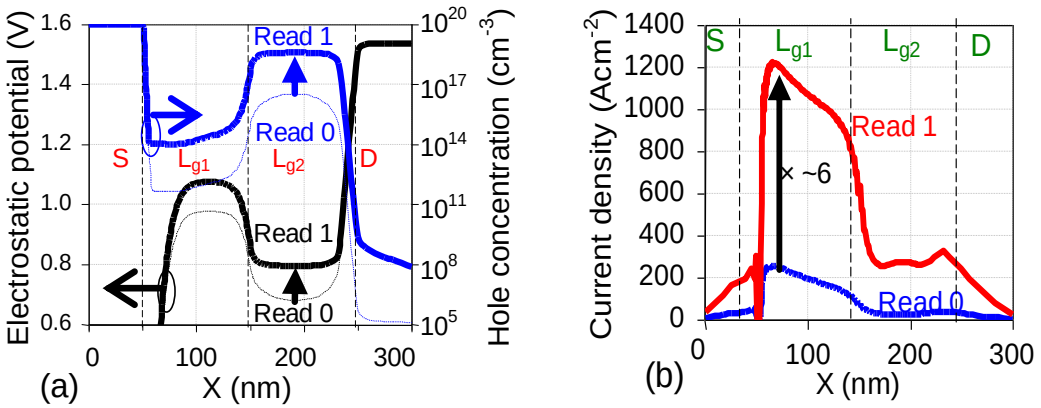

Fig. 3.4 Variation of (a) hole concentration and potential (after a hold time of 1 ms) (b) current density for state '1' and '0', during read operation along the channel length at a cross-section 1 nm below $Y = 0$.

3.2.3 Significance of bias during read

The bias applied affects the recombination and generation that directly influences the read currents. For a lower back gate bias (< 0 V) the electron and hole generation at back gate region lead to an increase in read currents (Fig. 3.6(a)). State '0' is depleted of charge carriers, and therefore, the impact of generation is more prominent for state '0' (Fig. 3.7(a)). Thus, for lower bias values, a higher degree of change is observed for read '0' as compared to read '1'. For a higher bias voltage (> 1.2 V) the increasing bias voltage drifts more holes towards the front gate. The significant change in the hole concentration is observed for bias > 1.2 V which lowers the band energy at front gate region, irrespective of the same bias applied at V_{g1}. It is influenced by the holes drifted towards front gate due to application of higher bias at back gate, and thus, results into higher conduction.

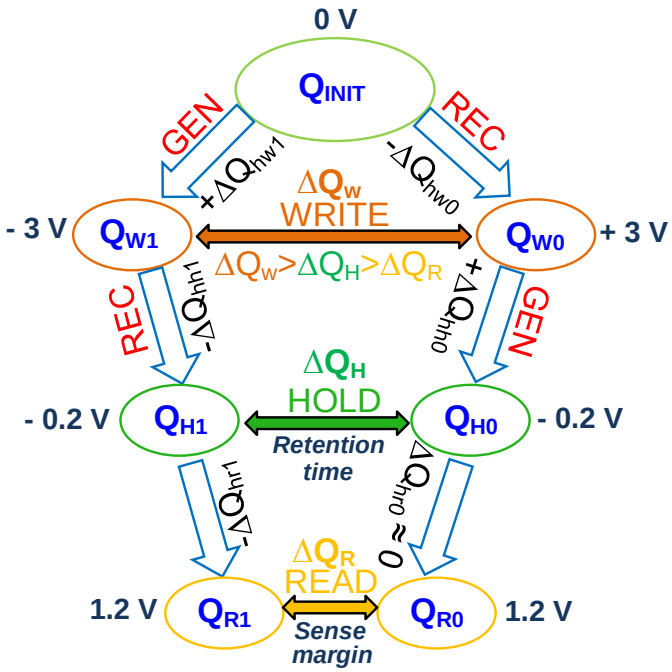

Fig. 3.5 Schematic illustration of charge distribution in the potential well with consecutive operation in the sequence as write, hold and read operation. Q_{INIT} indicates the charge in the storage region at zero bias, which increases to $Q_{W1} = Q_{INIT} + \Delta Q_{hw1}$ during write '1' and decreases to $Q_{W0} = Q_{INIT} - \Delta Q_{hw0}$ during write '0'. Further charge during hold '1' decreases to $Q_{H1} = Q_{W1} - \Delta Q_{hh1}$ and during hold '0' increases to $Q_{H0} = Q_{W0} + \Delta Q_{hh0}$. The charge during read '1' is Q_{R1} and during read '0' is $Q_{R0,}$ and the difference between the two is used to indicate *SM* in terms of current. GEN and REC represent generation and recombination, respectively.

The same is reflected in Fig. 3.7(b) that highlights lower band energy for electrons to drift towards drain for higher back gate bias values and the change is significant for state '0' and thus, read current for state '0' shows more increase when compared to state '1'. Thus, a low sense margin is observed for V_{g2} < -0.4 V and V_{g2} > 1.3 V (Fig. 3.6(b)). Figure also demonstrates decrease in I_1 and I_0 for V_{g2} between 0 V - 0.8 V. However, the influence is more for $I_{0,}$ which is reflective through current ratio. The analysis shows a wider programming window with V_{g2} ranging from -0.4 V to 1.3 V, highlighting the values of sense margin to be the range of 50-60 nAs. Fig. 3.6(b) shows a high current ratio (I_1/I_0) for a back gate bias of 0.8 V. However, state retention is negligible. *RT* > 64 ms is observable for back gate bias of 1.1-1.2 V. A back gate bias of 1.2 V is opted that yields the maximum retention time (Fig. 3.6(c)).

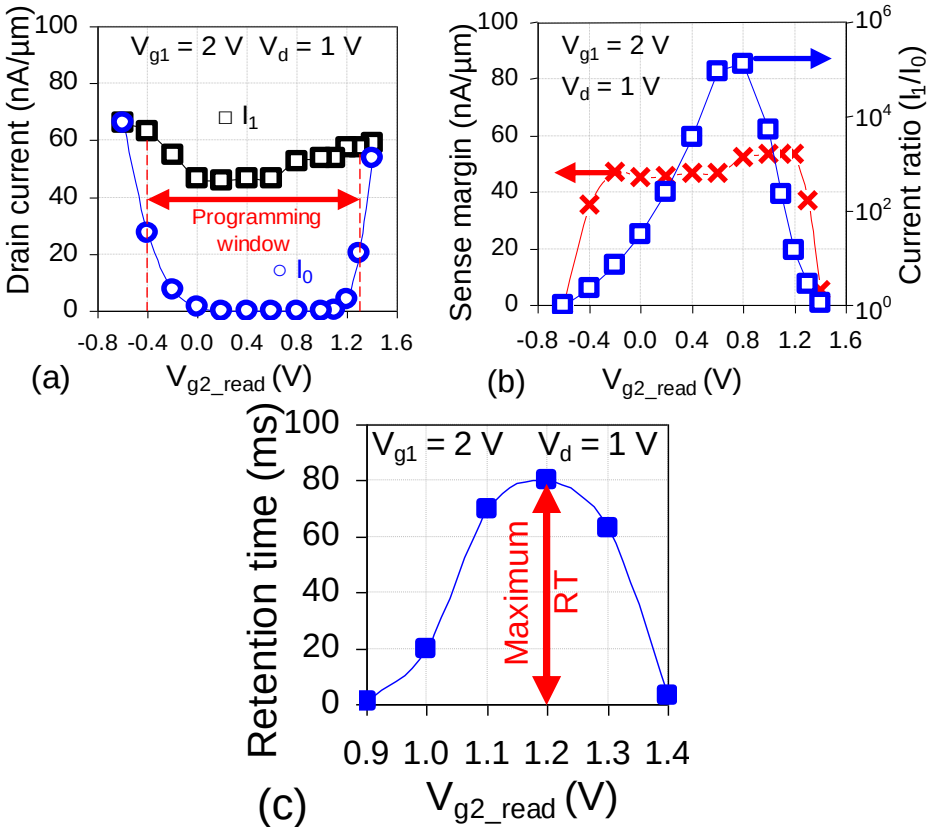

Fig. 3.6 Variation in (a) I_1 and I_0, (b) *SM* and *CR*, and (c) *RT* as a function of back gate voltage applied during read operation.

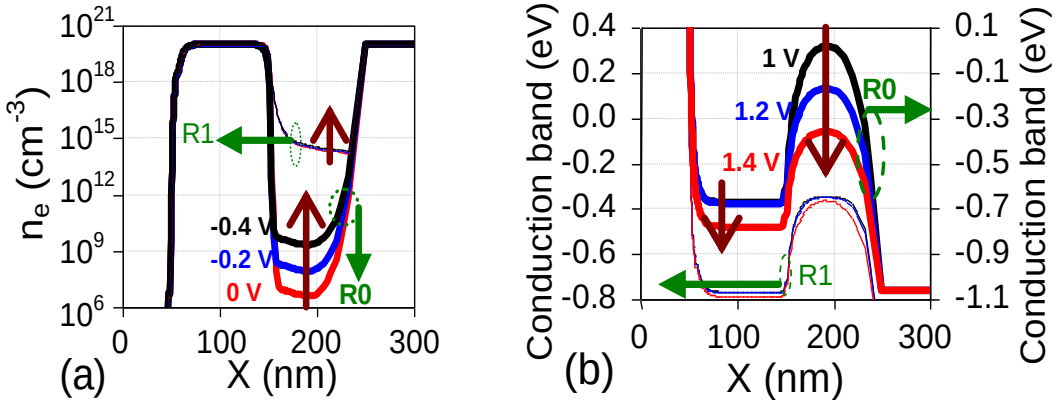

Fig. 3.7 Variation in (a) electron concentration (n_e) and conduction band energy for R1 and R0 as a function of back gate bias, < 0 V for (a) and > 1.2 V for (b) along *X* at a cross-section 1 nm below the front interface.

The *RT* is pre-dominantly controlled by the charge sustenance in the storage region which is governed by bias applied during hold operation. However, the bias applied during read is also crucial for DRAM characteristics. The charge loss in the physical well is determined by the generation and recombination phenomenon [18-26], governed by the back gate bias (Fig. 3.8(a)). Therefore, an

optimized back gate bias of -0.2 V is applied that gives the maximum *RT* of ~ 80 ms at 85 °C and ~ 300 ms at room temperature (Fig. 3.8(b)) for L_{g1} = L_{g2} = 100 nm, which highlights the device capability to operate as DRAM. The *SM* observed is ~ 60 nA at 85 °C (Fig. 3.8(c)). Thus, recombination and generation phenomenon regulated by capacitance associated with back gate and bias plays a vital role in charge storage and retention ($\Delta Q \propto C_s \Delta V_{g2}$).

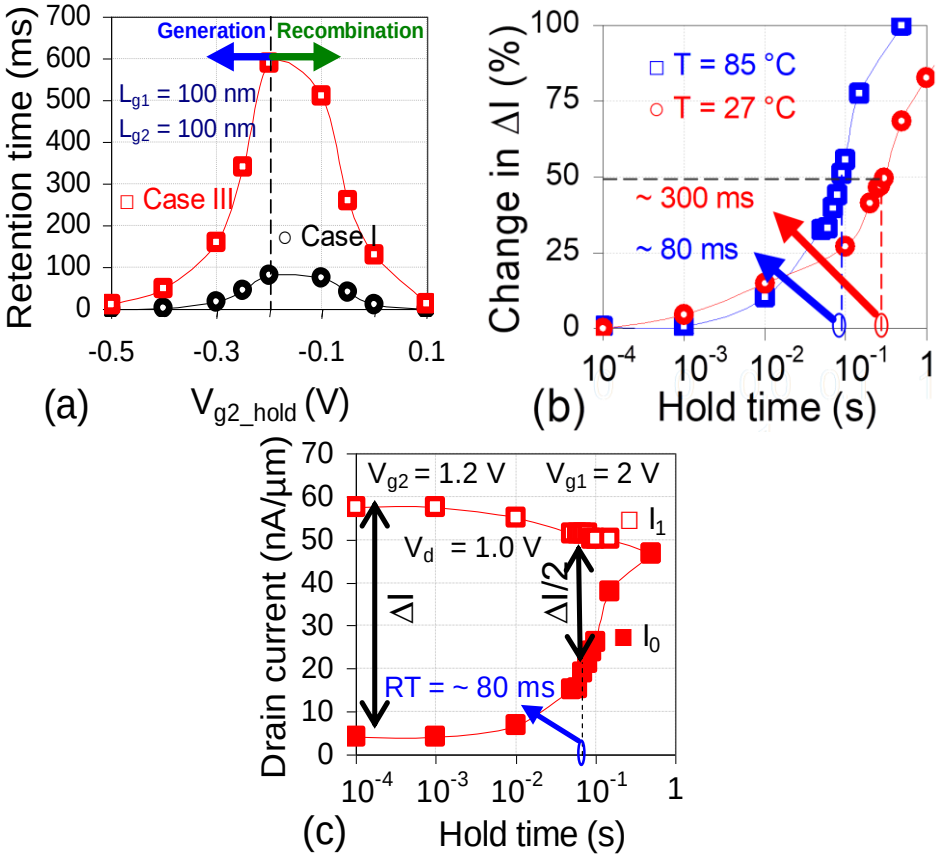

Fig. 3.8 (a) Dependence of *RT* on the back gate bias applied during hold operation for Case-I and Case III (with L_{gap} and L_{un}). (b) Variation in *SM* for Case-I at 27 °C and 85 °C, showing *RT* at a change of Δ*I* by 50%. (c) Variation in read currents for state '1' and '0' with hold time.

3.3 Design Optimization for improved performance metrics

3.3.1 Incorporation of lateral spacing (L_{gap}) and underlap region (L_{un})

3.3.1.1 Impact on retention time

The misaligned DG device for L_{g1} = L_{g2} = 100 nm (Fig. 3.1(a)) has shown significant improvement as compared to previous TFET based dynamic memory [27-30], but the present chapter focuses on adopting innovative techniques to

further enhance the retention characteristics (RT > 80 ms) and device scalability at 85 °C. This is accomplished by introducing a lateral spacing (L_{gap}) between the gates and including an underlap region (L_{un}) between the back gate and drain interface. Inclusion of L_{gap}, designated as Case-II (Fig. 3.1(c)) results into a wider potential well due to longer effective length of storage region (Fig. 3.1(d)). A longer length increases the storage capacitance, hence the charge retention. More significantly, it reduces the generation of holes due to leakage from virtually induced *n*-type front gate region, as demonstrated in Fig. 3.9 thereby, increasing the retention time. This is evident from Fig. 3.10(a) that shows inclusion of L_{gap}, irrespective of its length (25-80 nm) increases the retention time from 80 ms (Case-I) to ~ 300 ms (Case-II).

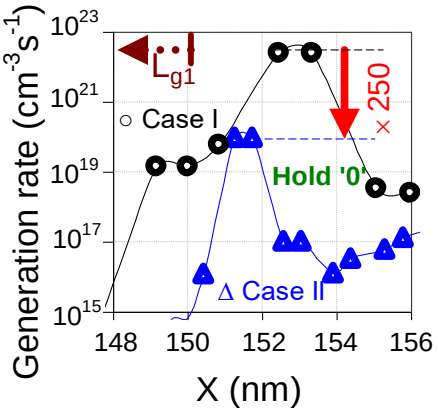

Fig. 3.9 Variation in generation rate for Case-I and Case-II for hold ‘0’ along the channel length.

Further, including an underlap region (L_{un}) of 20 nm (Case-III) between back gate and drain along with L_{gap} (Fig. 3.1(e)) as adopted in conventional MOS based DRAM [18,19,30] enhances retention characteristics. Fig. 3.10(a) clearly highlights the impact of using an L_{gap} and L_{un} on the sense margin and retention time. Fig 3.8(a) shows similar retention characteristics for Case-III as obtained in Case-I, for bias varied during hold operation and maximum RT at -0.2 V, but with an enhanced performance. Fig 3.10(b) summarizes the improved retention characteristics from 80 ms with a misaligned gate (Case-I) to ~300 ms with the use on a lateral spacing between the gates (Case-II) and finally, to ~600 ms with an underlap (Case –III) for $L_{g1} = L_{g2} = 100$ nm at 85 °C.

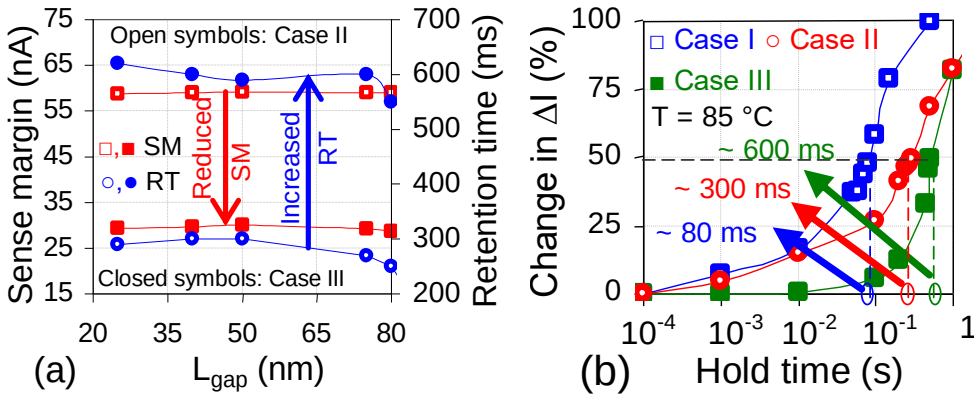

Fig. 3.10 (a) Dependence of *RT* and *SM* on L_{gap} and L_{un} regions. (b) Percentage change in *SM* with hold time, where 50% variation reflect *RT* of 80 ms, 300 ms, and 600 ms for design with Case I, II, III, respectively.

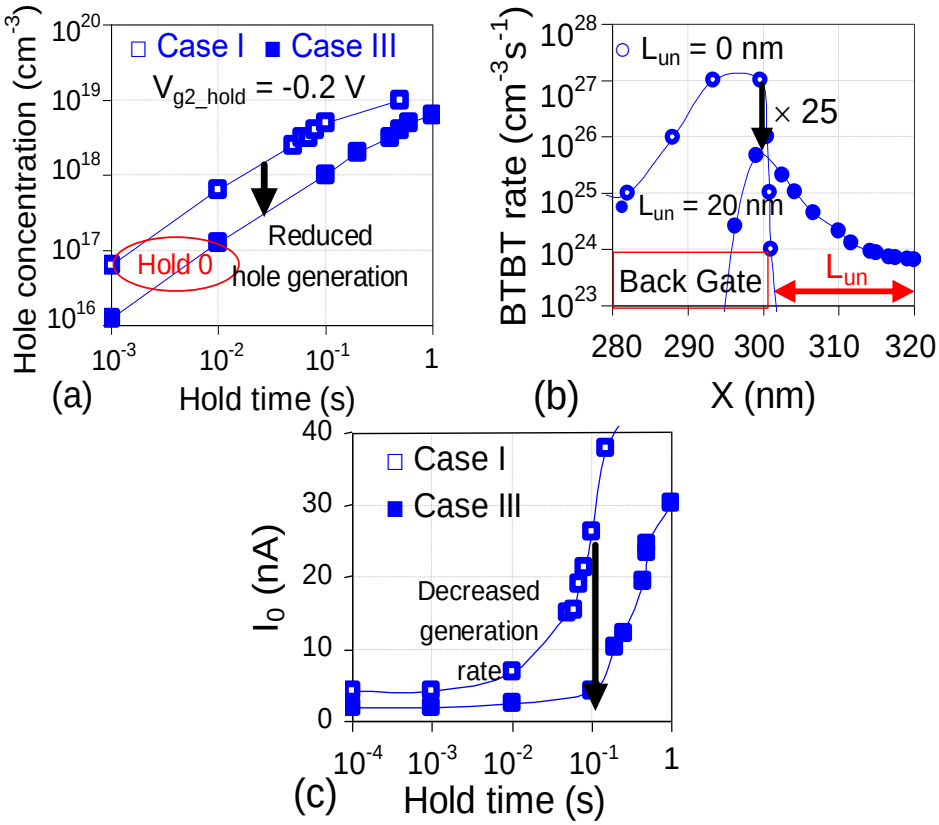

Fig. 3.11 Variation in (a) hole concentration for Case-I and Case-III for state '0' with the time. (b) Variation of BTBT rate at the back gate and drain interface during write '1' operation with and without the underlap regions. (c) Variation in I_0 for Case-I and Case-III with hold time.

The use of an underlap between drain and back gate results into reduced hole generation during hold '0', as shown in Fig 3.11(a). This is due to presence of an intrinsic region between a virtually induced *p*-type region, associated with back

gate and heavily doped n^+ drain region. It results into increased depletion width at the junction, and thus, lowers the electric field at G1/G2 and G2/drain interface. This reduces BTBT of electrons into drain for hold state '0'. This decrease the hole generation rate (Fig. 3.11(b)), and hence, lower the current during read '0', as demonstrated in Fig. 3.11(c). The longer sustenance of state '0' (which was significantly degraded for a back gate bias of -0.2 V) during hold, increases the *RT* to ~ 600 ms.

3.3.1.2 Impact on sense margin

Further, focusing on the another DRAM metric, sense margin, Fig. 3.10(a) shows the use of L_{gap} is benefited with a higher retention while maintaining the sense margin as in Case-I (~ 60 nA). This is due to tunneling based read mechanism that doesn't vary the read currents for a longer effective length. Also, the barrier height between both the gates that governs diffusion is not affected with the presence of L_{gap} for L_{g1} = L_{g2} = 100 nm. The impact on sense margin is prominent at shorter gate lengths, which has been demonstrated in subsequent chapters, but doesn't show variation for the gate length investigated in the present chapter. Thus, as shown in Fig. 3.10(a) L_{gap} doesn't affect the sense margin as the potential barriers are unaltered; however including a L_{un} reduces the *SM*.

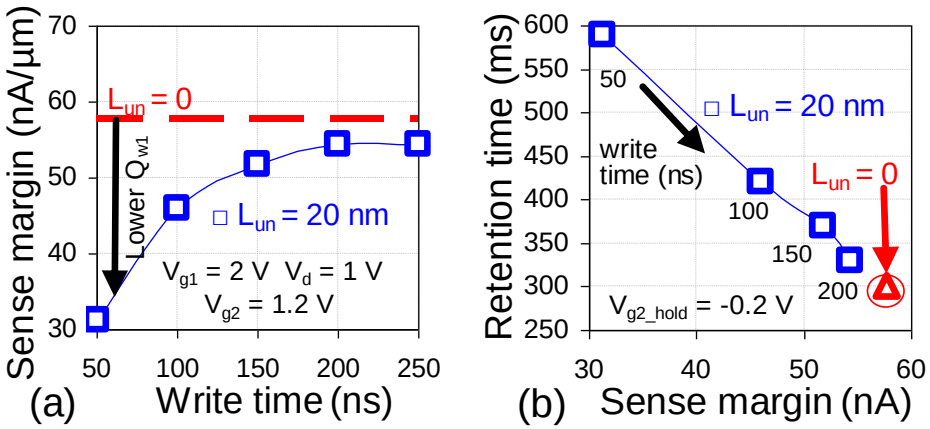

Fig. 3.12 (a) Variation in *SM* with write time for Case-II and Case-III. (b) Dependence of *RT* and *SM* on write time for Case-II and Case-III. Q_{W1} indicates the hole charge generation during write '1'.

The write mechanism is based on band-to-band tunneling of electrons from back gate to drain that facilitates hole accumulation. The inclusion of an underlap

between these region, reduces the tunneling rate, thereby reducing the hole generation at the storage region. The charge generation during write '1' (Q_{W1}), lead to an increase in hole concentration, which reduces during read operation (Q_{R1}). Although, hole concentration at read is lower than that at write '1', it is higher than the concentration at steady state. These excess charge carriers, increase the effective potential, and thus, result into higher read current that defines state '1'. As shown in Fig. 3.12(a) for Case-II, that is for L_{un} = 0, for a write time > 50 ns, a constant read current for state '1' is obtained. This is due to the fact that the hole concentration during read saturates (~ 8 × 10^{18} cm^{-3}) for a particular bias applied at the back gate (1.2 V) during read [14]. Thus, irrespective of the holes generated during write '1', the read current is constant with write time; if a hole concentration of ~ 8 × 10^{18} cm^{-3} is retained during read '1'. However, for the case with underlap, the reduced hole generation during write '1' leads to a lower hole concentration during read (< ~ 8 × 10^{18} cm^{-3}) and hence, result into a reduced sense margin (Fig. 3.12(a)).

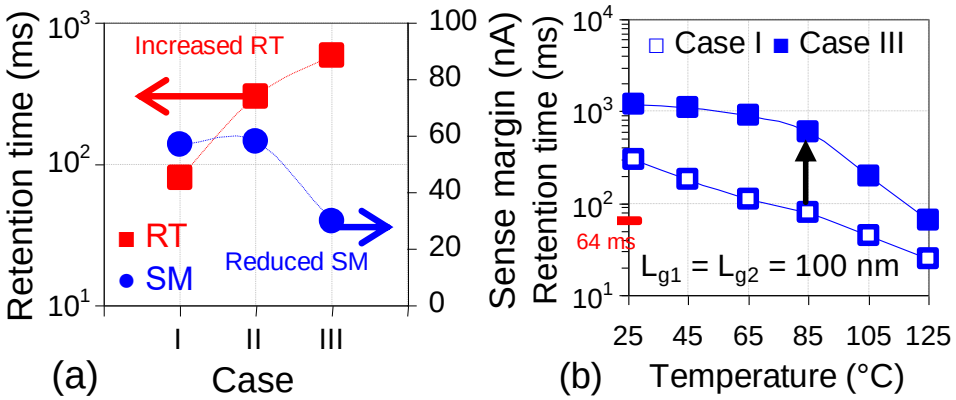

Fig. 3.13 (a) Variation in retention time and sense margin for (i) Case-I, (ii) Case-II (L_{gap} = 50 nm) and (iii) Case-III (L_{gap} = 50 nm and L_{un} = 20 nm). (b) Variation in *RT* with the temperature for Case-I and Case-III.

The deficiency of the holes can be compensated by increasing the write time that accumulates more holes and, thus, reflect a higher read current (Fig. 3.12(a)). The hole concentration at read '1' for a write time = 200 ns is same as observed for read '1' in structure without underlap (L_{un} = 0). However, even with the same hole concentration during read for a write time = 200 ns, the sense margin is ~ 3 nA lower which is due to the underlap. Thus, it can be inferred that reduction in sense

margin (~30 nA) is more prominent due to reduced hole generation (for write time = 50 ns) rather than that due to increased resistance associated with underlap. Although, *SM* enhances with write time, the rapid degradation of state '1' with increased recombination due to underlap result into a reduced *RT* (Fig. 3.12(b)). The incorporation of underlap, thus, leads to a trade-off between sense margin and retention time (Fig. 3.13(a)). However, the investigation gives a broader perspective of selecting the bias values and write time to achieve DRAM characteristics. This is similar to MOS with underlap in [30] that shows a trade-off between high operating bias for charge generation (based on impact ionization) and improved *RT*. However, the primary focus is improving *RT* for which the proposed write time of 50 ns is adopted that shows an acceptable *SM* and a significant improvement in *RT*. Additionally, the results highlight that device with an L_{gap} (Case-II) can be utilized, which maintains the *SM* and a higher *RT* (compared to Case-I) thereby, balancing between the two important parameters of DRAM.

3.3.1.3 Impact of temperature

DRAM operation may be limited at higher temperatures due to the increased recombination/generation rate that leads to loss of retention characteristics [18]. Many military as well as industry applications require DRAM with operation at higher temperatures. Therefore, the performance of dynamic memory at elevated temperatures has been investigated and it is observed that with the adopted architecture that includes an L_{gap} and L_{un}, *RT* > 64 ms at temperature till 125 °C is observed. As shown in Fig. 3.13(b) for device design without L_{gap} and L_{un} (Case-I) *RT* > 64 ms could be achieved till 85 °C while for Case-III, operation till 125 °C is permissible which gives a wide range of operating temperature. Thus, along with down scaling capability, functionality at elevated temperatures demonstrates additional benefits of using L_{un} and L_{gap} in TFET based DRAM.

3.3.2 Significance of gate lengths

3.3.2.1 Gate length scaling

The scaling of device is considerably limited by BTBT at the drain/source and gate region [26], which can be resolved by using an underlap between drain and

back gate region [30] and a lateral spacing between the gates. Analysis shows L_{g1} can be scaled down to 75 nm, while reducing L_{g1} below it doesn't distinguish the two read currents. Thus, utilizing L_{g1} = 75 nm, for Case-III, *RT* > 64 ms can be attained with storage region of 50 nm. Therefore, Fig. 3.14(a) demonstrates the scaling limit of 25 nm and 50 nm with L_{g1} as 100 nm, and 75 nm, respectively. Thus, the device design incorporated with L_{un} and L_{gap} upgrade the device scalability, evident from Fig. 3.14(b) that shows with the front gate (L_{g1}) of 100 nm, the storage region (L_{g2}) can be scaled down to 75 nm for Case-I, 50 nm for Case-II and finally, downscaling till 25 nm is possible for Case-III, achieving *RT* > 64 ms [6]. Fig. 3.14(c) shows impact of reduced L_{g2} on *RT* and *SM*. L_{g2} acts as potential barrier for the electrons and also, potential well for charge storage. Its reduction decreases the storage capacitance, hence, the retention time, and also, reduces the barrier for electrons, thereby increasing the read currents, and hence, sense margin. The results demonstrate a remarkable improvement in scaling as compared to previous work [27-29] and also, the design presented in the previous chapter, highlighting the significance of design optimization.

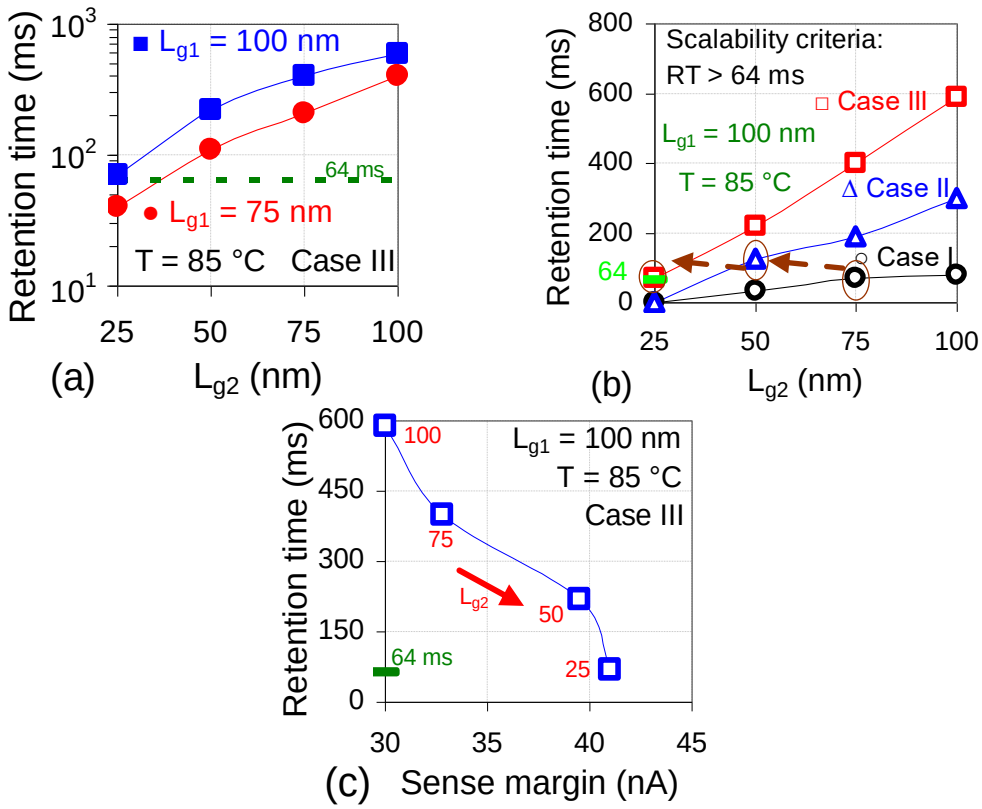

Fig. 3.14 Dependence of *RT* on L_{g2} for (a) various values of L_{g1} (b) different cases with L_{g1} = 100 nm. (c) Dependence of *RT* and *SM* on L_{g2} for L_{g1} = 100 nm.

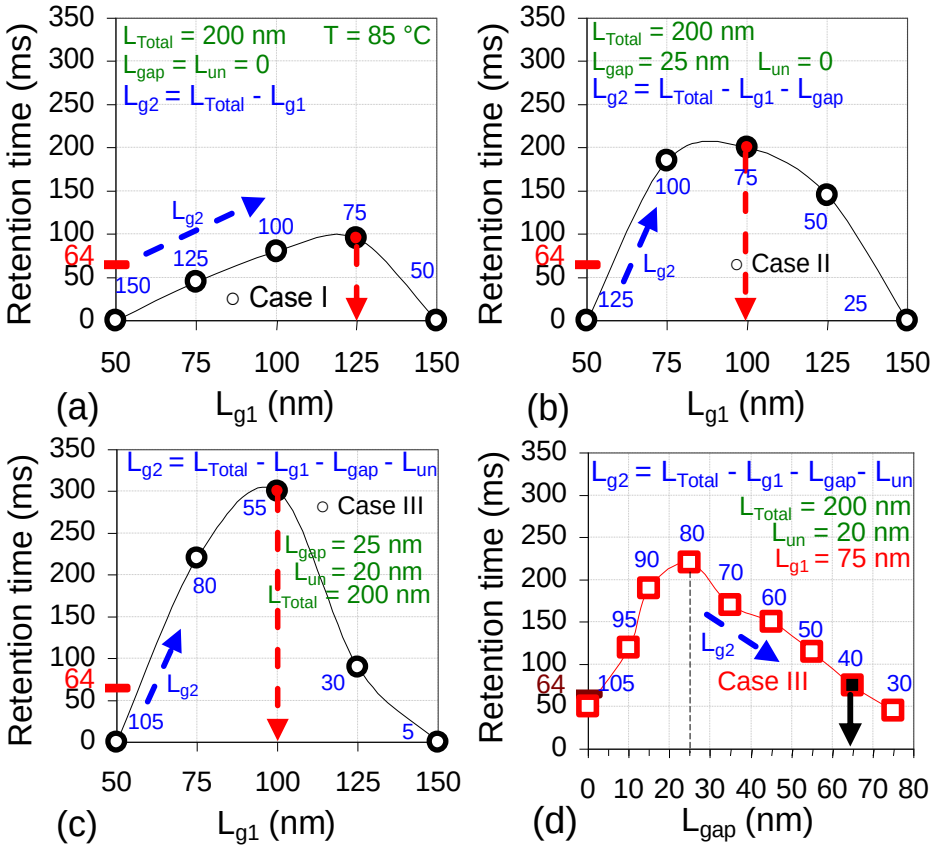

Fig. 3.15 Variation in *RT* as a function of gate lengths (L_{g1} and L_{g2}) with total length (L_{Total}) as 200 nm for (a) Case-I, (b) Case-II, and (c) Case-III. (d) Dependence of *RT* on lateral spacing (L_{gap}) and back gate length (L_{g2}) for L_{g2} + L_{gap} + L_{un} (= 20 nm) = 125 nm and front gate length (L_{g1}) of 75 nm with a total length of 200 nm. **Case I**: $L_{Total} = L_{g1} + L_{g2}$, **Case II**: $L_{Total} = L_{g1} + L_{gap}$ (= 25 nm) + L_{g2}, and **Case III**: $L_{Total} = L_{g1} + L_{gap} + L_{g2} + L_{un}$ (= 20 nm).

3.3.2.2 Total length as scaling criteria

The above discussion highlights that for DRAM scalability, along with the systematic approach that utilized L_{gap} and L_{un}, the optimization of individual lengths as well as total length ($L_{Total} = L_{g1} + L_{g2} + L_{gap} + L_{un}$) is also crucial. The significance of each with a constant total length of 200 nm is evaluated that shows the extent of downscaling for L_{g1}, L_{g2} and L_{gap}. As shown in Figs. 3.15(a)-(c) for Case I, the front gate length of 125 nm with a back gate region of 75 nm shows the maximum *RT* of ~ 100 ms, while for Case II, maximum *RT* (200 ms) is attained with L_{g1} = 100 nm, L_{g2} = 75 nm and L_{gap} = 25 nm. In Case III, having L_{g1}=100 nm, L_{g2}= 55 nm, L_{gap} = 25 nm and L_{un}= 20 nm, the maximum *RT* of 300 ms is achieved. The result shows the device progresses from a *RT* of 100 ms to

300 ms from Case I to Case III, with total length of 200 nm. Figs. 3.15(a)-(c) also demonstrate that retention time is regulated by front as well as back gate with the front gate predominantly governing the read mechanism.

Fig. 3.15(c) shows the front gate length (L_{g1}) cannot be scaled below 75 nm. The scaling of front gate length narrows the *n*-type induced barrier due to n^+ poly front gate, resulting into reduced gate controllability [24], and thus, short channel effects result into decreased sense margin. Therefore, for further estimation, $(L_{g1})_{min}$ = 75 nm with constant $L_{g2} + L_{gap} + L_{un}$ (= 20 nm) = 125 nm is considered. Fig. 3.15(d) illustrates that L_{g2} can be scaled down to 40 nm and L_{gap} to 25 nm. For, $L_{gap} \geq 25$ nm, the reduction in L_{g2} reduces the storage capacitance and thus, leads to decrease in retention time. However, reducing L_{gap} below 25 nm, permits a higher tunneling towards front gate region, thereby degrading state '0' and thus, reduces retention time. The variation in *RT* for a constant total length of 200 nm shows individual gate lengths are crucial for *RT* evaluation while $L_{gap} \geq 25$ nm (Fig. 3.10(a)) suffices the minimum requirement, as *RT* decreases below L_{gap} = 25 nm. Thus, rather than the length of lateral spacing (for ≥ 25 nm) its utilization affects *RT* due to reduction in the generation rate for state '0'.

Considering, total length as the scaling criteria, Fig. 3.16(a) demonstrates Case-I ($L_{Total} = L_{g1} + L_{g2}$) and Case-III ($L_{Total} = L_{g1} + L_{gap} + L_{g2} + L_{un}$) with L_{g1} = 75 nm. The figure illustrates improved *RT* as well scalability with the total length scaled down to 160 nm for Case-III, in comparison to Case-I that shows L_{Total} to be scaled down to 220 nm for *RT* to be higher than 64 ms. Although, inclusion of L_{gap} = 25 nm and underlap L_{un} = 20 nm increases the effective gate length, they allow back gate to be scaled down to 40 nm which is much higher when compared to Case-I having back gate scaled to 145 nm with front gate length of 75 nm. Focusing on *RT* > 64 ms, Fig. 3.16(b) demonstrates variation in *RT* with the ratio of gate lengths (L_{g1}/L_{g2}) and also, signifies the maximum limit of scaling that can serve as a guideline for device design. The device can efficiently perform as DRAM with *RT* > 64 ms for $L_{g1}/L_{g2} \leq 4$ for L_{g1} = 100 nm and $L_{g1}/L_{g2} \leq 2$ for L_{g1} = 75 nm. Although, total gate length is important for practical area scaling, but individual gate lengths also serve as key scaling parameters. Thus, the scalability

of the TFET DRAM depends on the control of the regions formed by the gates and the device can perform efficiently i.e. (*RT* > 64 ms) with minimum value of L_{g1} = 75 nm, L_{g2} = 40 nm, L_{gap} = 25 nm and L_{un} = 20 nm, showing L_{Total} = 160 nm.

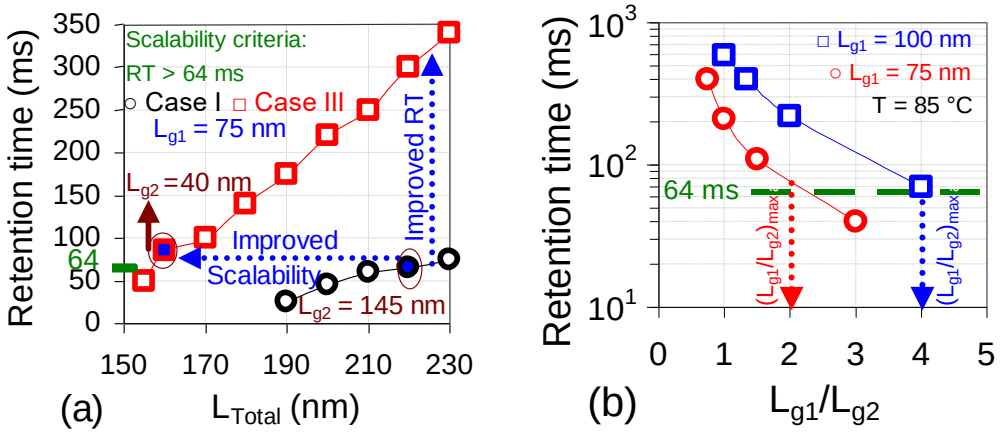

Fig. 3.16 (a) Dependence of *RT* on total length for a front gate length (L_{g1}) of 75 nm for Case-I and Case-III (L_{gap} = 25 nm and L_{un} = 20 nm). (b) Variation in *RT* with ratio of gate lengths (L_{g1}/ L_{g2}) for Case-III, indicating the maximum ratio of gate lengths for memory design with *RT* > 64 ms.

The present chapter highlights an innovative device design to improve the retention characteristics and scalability, showing an enhanced performance as compared to previous TFET based DRAM [27,28], evident from Fig. 3.17(a). In the previous work with misaligned DG TFET, the *RT* achieved was ~ 2 s with the gate length of 400 nm and storage region of 200 nm at 85 °C, which showed improvement over the TFET with same gate length and storage region but with an over-sized back gate [27]. Also, for shorter gate length of 100 nm with a storage region of 25 nm, the fin-based TFET proposed in the past could only attain a retention time of ~ 100 µs at 85 °C [29]. However, in the present work considerable improvement in terms of higher *RT* for scaled lengths at 85 °C. Moreover, the proposed TFET advances as the device that can probably compete with similar structures, FED [31,32] and Z^2-FET [23-24] in terms of *RT*, although they operate in forward bias with a feedback mechanism whereas TFET operates in reverse bias based on BTBT for read mechanism. The tunneling mechanism limits on-current, resulting into lower sense margin compared to the other similar structures (Fig. 3.17(b)) but has shown a significant improvement as DRAM in terms of *RT* and scalability.

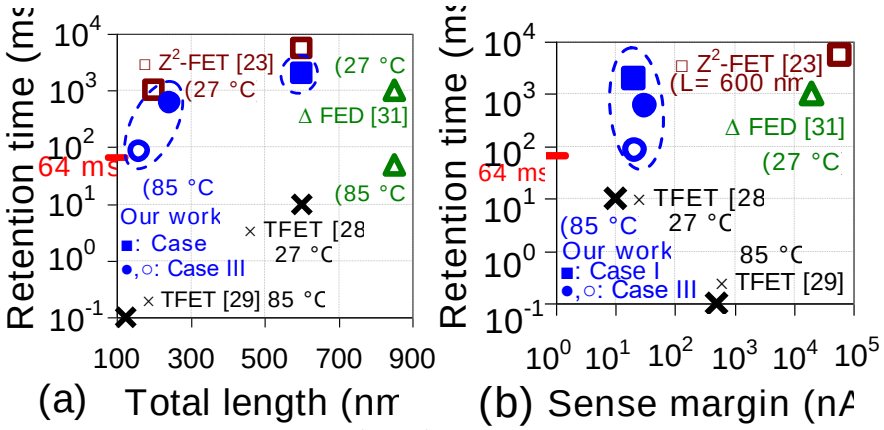

Fig. 3.17 Comparison of various p^+-i-n^+ structures in terms of *RT* as a function of (a) total length and (b) *SM*. In our work, Case I: $L_{Total} = L_{g1} + L_{g2}$, Case III: $L_{Total} = L_{g1} + L_{g2} + L_{gap} + L_{un}$ with L_{gap}= 25 nm and L_{un} = 20 nm has been considered and filled symbol (•) indicate L_{g1} = 100 nm, and open symbol (o) represent a device with L_{g1} = 75 nm. (b) show *SM* for the total length indicated in (a).

3.4 Conclusion

The chapter demonstrates the functionality of a sub-100 nm double gate TFET as dynamic memory where the capacitance associated with back gate is utilized to store the charges. The front gate is positioned at a partial region of the intrinsic film and primarily controls the read mechanism based on BTBT, while the back gate is located beneath the front un-gated region. The work showcases a methodological assessment that highlights optimization of device design in three steps: misalignment, lateral spacing between the gates (L_{gap}) and an underlap (L_{un}) between the back gate and drain region that improves:

(i) the retention time to ~ 600 ms (from ~ 80 ms without L_{gap} and L_{un}) at 85 °C for gate length (L_{g1}) and storage region (L_{g2}) of 100 nm;

(ii) the device scalability with L_{g2} being scaled down to 25 nm with L_{g1} = 100 nm;

(iii) the total length scalability from 220 nm (without L_{gap} and L_{un}) to 160 nm for L_{g1} = 75 nm;

(iv) the device functionality at elevated temperatures with *RT* > 64 ms till 125 °C.

The chapter systematically analyses significance of individual gate lengths, L_{gap} and L_{un}, their extent of scaling and impact on *SM* and *RT*. It shows for a fixed L_{un} (= 20 nm) and L_{gap} (= 25 nm) $L_{g1}/L_{g2} \cong 4$ @ L_{g1} = 100 nm, or $L_{g1}/L_{g2} \cong 2$ @ L_{g1} =

75 nm must be maintained, with minimum L_{Total} (= L_{g1} + L_{g2} + L_{gap} + L_{un}) of 160 nm to 170 nm for DRAM functionality. Thus, minimum L_{Total} and individual lengths (L_{g1}, L_{g2}, L_{gap}, L_{un}) are critical for RT > 64 ms. The device performance is better, attributed to increased storage capacitance due to increase in effective length of storage region with the integration of L_{gap}, supported by L_{un} and most importantly, reduced leakage current associated with tunneling during hold state '0'. Through a methodical analysis, insightful analysis, and device design, the investigation highlights the potential of TFET to operate as DRAM with optimal bias values in sub-100 nm regime. The work emphasizes the opportunity for TFET for dynamic memory with improved retention characteristics and better scaling perspective, achieved through design optimization.

3.5 References

[1] Moore G. E., (1975) Progress in digital integrated electronics, IEEE Electron Devices Meeting, 21, pp. 11-13.

[2] Schaller R. R., (1997) Moore's law: past, present and future, IEEE spectrum, 34, 52-59.

[3] Mollick E., (2006) Establishing Moore's law, IEEE Annals of the History of Computing, 28, 62-75.

[4] Wong H.-S. P., (2002) Beyond the conventional transistor, IBM Journal of Research and Development, 46, 133-168.

[5] Chen T. C., (2006) Where CMOS is going: trendy hype vs. real technology, IEEE Solid-State Circuits Conference, pp. 1-18.

[6] ITRS International Technology Working Groups. International Technology Roadmap for Semiconductors, http://www.itrs.net (2015).

[7] Sharroush S. M., (2016) Performance optimization of 1T-1C DRAMs: A quantitative study, Microelectronics Journal, 52, 147-164.

[8] Faul J. W., and Henke, D., (2005) Transistor challenges–A DRAM perspective, Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, 237, 228-234.

[9] Itoh K., (2008) The history of DRAM circuit designs–at the forefront of DRAM development, IEEE Solid-State Circuits Society Newsletter, 13, 27-31.

[10] Kim K., (2000) Perspectives on giga-bit scaled DRAM technology generation, Microelectronics reliability, 40, 191-206.

[11] Joodaki M., (2012) Selected advances in nanoelectronic devices: logic, memory and RF, Springer Science and Business Media, Heidelberg (ISBN-978-3-642-31349-3).

[12] Moore S. K., (2007) Masters of memory, IEEE spectrum, 44, pp.45-49.

[13] Ohsawa, T., and Hamamoto, T., (2012) A Novel Capacitor-less DRAM: Floating Body Cell, CRC Press, Taylor & Francis Group, Florida (ISBN: 978-9-81430-308-8).

[14] Giusi G., (2014) Physical insights of body effect and charge degradation in floating-body DRAMs, Solid-State Electronics, 95, 1-7.

[15] Bohr M., (2007) A 30 year retrospective on Dennard's MOSFET scaling paper, IEEE Solid-State Circuits Society Newsletter, 12, 11-13.

[16] Atlas User's Manual, Silvaco Int., Santa Clara, CA, Jan. 18, 2015.

[17] Park J. S., (2002) Dual Gate MOSFET Fabrication US Patent 6168998 B1.

[18] Sasaki K. R. A., Nicoletti T., Almeida L. M., dos Santos S. D., Nissimoff A., Aoulaiche M., Simoen E., Claeys C., and Martino J. A., (2014) Improved retention times in UTBOX nMOSFETs for 1T-DRAM applications, Solid-State Electronics, 97, 30-37.

[19] Collaert N., Aoulaiche M., Rakowski M., Wachter B. De, Bourdelle K., Nguyen B. Y., Boedt F., Delprat D. and Jurczak M. (2009), Analysis of sense margin and reliability of 1T-DRAM fabricated on thin-film UTBOX substrates, IEEE International SOI Conference, pp. 1.

[20] Giusi G., Alam M. A., Crupi F., and Pierro S., (2010) Bipolar mode operation and scalability of double-gate capacitorless 1T-DRAM cells, IEEE Transactions on Electron Devices, 57, 1743-1750.

[21] Almeida L. M., Sasaki K. R. A., Aoulaiche M., Simoen E., Claeys C., and Martino J. A., (2012) One transistor floating body RAM performances on UTBOX devices using the BJT effect, Journal of Integrated Circuit System, 7, 113-118.

[22] Kuo C., King T. J., and Hu C., (2002) A capacitorless double-gate DRAM cell, IEEE Electron Device Letters, 23, 345-347.

[23] Wan J., Royer C. Le, Zaslavsky A., and Cristoloveanu S., (2013) Progress in Z^2-FET 1T-DRAM: retention time, writing modes, selective array operation, and dual bit storage, Solid-State Electronics, 84, 147-154.

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

[25] Fossum J. G., Lu Z. and Trivedi V. P., (2007) New insights on "capacitorless" floating-body DRAM cells, IEEE electron device letters, 28, 513-516.

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

[27] Biswas A., Dagtekin N., Grabinski W., Bazigos A., Le Royer C., Hartmann J. M., Tabone C., Vinet M., and Ionescu A. M., (2014) Investigation of tunnel field-effect transistors as a capacitor-less memory cell, Applied Physics Letters, 104, 092108.

[28] Biswas A., and Ionescu A. M., (2015) 1T capacitor-less DRAM cell based on asymmetric tunnel FET design, IEEE Journal of the Electron Devices Society, 3, 217-222.

[29] Biswas A., and Ionescu A. M., (2014) Study of fin-tunnel FETs with doped pocket as capacitor-less 1T DRAM, IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference, pp. 1-2.

[30] Aoulaiche M., Collaert N., Degraeve R., Lu Z., De Wachter B., Groeseneken G., Jurczak M., and Altimime L., (2010) BJT-mode endurance on a 1T-RAM bulk FinFET device, IEEE Electron Device Letters, 31, 1380-1382.

[31] Badwan A. Z., Chbili Z., Yang Y., Salman A. A., Li Q., and Ioannou D. E., (2013) SOI field-effect diode DRAM cell: Design and operation, IEEE Electron Device Letters, 34,1002-1004.

[32] Badwan A. Z., Li Q., and Ioannou D. E., (2015) On the nature of the memory mechanism of gated-thyristor dynamic-RAM Cells, IEEE Journal of the Electron Devices Society, 3, 468-471.

Chapter 4

Twin Gate Tunnel FET based Capacitorless Dynamic Memory

4.1 Introduction

The need of portable smart electronic products demands low power devices compatible with conventional CMOS along with technological competency [1-13]. The choice is governed by the selection of applications such as cellular phones, personal digital assistants, medical electronics, and other portable computing and smart devices [8,9] that not only demands high density, but also, minimization of energy and improved speed per operation [14-17]. One of the crucial components in these electronic devices is the memory. The usage of memory devices has extensively widened with the introduction of data intensive applications. Along with computing segment, the various mobile multimedia applications, advanced cell phones, applications based on big data, cloud computing, virtual reality, and Internet of Things (IoT) [18-24] require innovation in memory to achieve high performance and density with low power consumption.

Focusing on dynamic memory, the need for high speed has always been the quest that led to invention of various series of Double Data Rate (DDR) Synchronous Dynamic Random Access Memory (SDRAM). This led to development of laptop and notebook and further, smartphones [1,8,9,13]. Thus, DRAM application is not only limited to various applications for computers, including email, video streaming, but also, wide range of "smart" products, including mobile devices, high definition television and MP3 players [8]. Among all integrated circuits, embedded memory [2-6] has an important in today's application as high performance computing, low power consumer electronics [9]. Therefore, the focus has been diverted towards improving eDRAM, being denser and power efficient than Static RAM [25]. However, speed is a critical issue. The development is aimed at using eDRAM to replace SRAM to achieve cheaper and lower voltage embedded chips. The present

DRAM technology utilizes LPDDR4 (Low Power Double Data Rate-4) as 1T 1C DRAM. LPDDR4 show enhanced speed than its predecessor with operation a lower voltage of 1.1 V [1,10,13,27]. The device is compatible with existing cell-libraries, and are considered well suited for present mobile technologies and IoT based devices, but along with cost overhead [25]. Therefore, the devices with elimination of external storage capacitor could serve as a substitute.

Capacitorless DRAMs can therefore be utilized due to its simple structure with low cost, if it meets the following expectations [28],

(i) Low write time,

(ii) High retention,

(iii) Ultra low voltage operation.

However, the aggressive scaling of conventional MOS devices worsened the device performance, thereby increasing the standby power [7,12,29]. To address this, innovative low power devices have been commonly introduced [30-50] as FETs with polarity control [30-32], FED [33-37], Z^2-FET [38-41], TFET [41-44], Heterostructure-FED (H-FED) [45], Modified-FED (M-FED) [46], Side-Contacted FED (SFED) [49]. With an intrinsic channel, metal gates with suitable workfunction are required to achieve application specific performance. The analysis based on these devices use two independent gates that will substantially benefit in replacing the conventional MOS devices.

The devices have been implemented for logic circuits and have shown applicability for high speed as well as low power operation [46-50]. In few digital applications, the device with three gates has been proposed for steep switching and has been implemented as basic logical operations [48]. Thus, the incorporation of two gates independently has been proposed for various applications that include its utilization for:

i) Electrostatic Discharge (ESD) protection [33,34],

ii) Memory devices[35-44],

iii) Steep-switching devices [33-41],

iv) Digital and logic circuits [45-50].

Expending the idea of incorporating additional electrodes permits the selection of charge-carrier types in nanowire transistors, and thus, creation of compact logic

gates [31]. The devices utilize the ambipolar conduction to enhance logic functionality, where results have shown fabrication of complex circuits (XOR, a full adder) using less resources than conventional CMOS [31,50]. The investigation has shown improved gate controllability with independent operation that result into faster operation at reduced gate lengths. Moreover with scaling, ESD of integrated circuits is a concern, where FED is suitable as a local clamping ESD device due to high current capability and fast operation [32,33]. Thus, devices with independent gate operation can be used for power management and innovative logical function design [31]. Therefore, the applicability of such devices at circuit level may be welcomed as post-CMOS devices.

Among these devices, TFET needs exploitation to perform as high speed, energy efficient dynamic memory for embedded memory applications that will also be the focus on this chapter. Thus, along with stand-alone applications, the device capability to operate at low power with low write time and high retention is studied that could be useful for embedded applications as well. The ideology presented in Chapter 2 and 3 is further realized as a twin gate topology in this chapter. The analysis is based on controlling the energy barriers, regulated by device structure, dimensions, bias values and temperature. The concept, design and operation as DRAM are explained through band diagrams by decoupling each operation and highlighting the underlying device physics. The twin gate architecture has structure similar to FED [33-37], but operates with different mechanism. Also, the operating principle of twin gate is similar to misaligned DG TFET, introduced in chapter 2, but with a different gate positioning. Therefore, the device operation is compared with two architectures, with first focus on variation with FED and subsequently, with misaligned DG TFET at the concluding section after discussing the operating principles and results of twin gate.

4.2 Comparison with Field Effect Diode (FED)

Twin gate TFET has similar architecture as that of FED [37], utilized for dynamic memory applications. The twin gate structure can be fabricated as shown by [51-57]. FEDs have been previously explored for Electrostatic Discharge (ESD)

protection and DRAM applications. The FED resembles a thyristor when operated as a memory device [37], where current characteristics exhibit a snapback to distinguish between the states. For DRAM, FED showed *RT* of ~ 1 s with L_{g1} = L_{g2} = 400 nm at 27 °C [36]. The two front gates, Gate-1 (G1) and Gate-2 (G2) are utilized to virtually induce *n*-region, and *p*-region, respectively. The virtually induced *p*-type region is used as storage region (Fig. 4.1). However, the operating principle for both the devices is different.

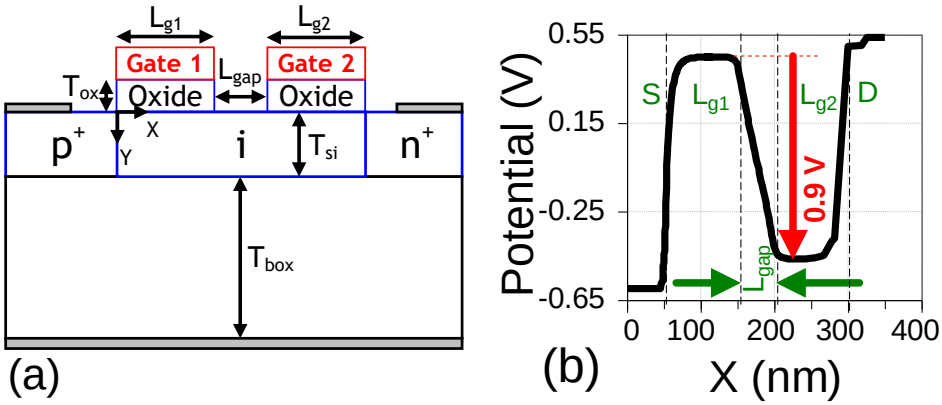

Fig. 4.1 Schematic diagram of twin gate TFET. (b) Variation of electrostatic potential profile at zero bias along the channel direction (*X*). The cutline is taken at 2 nm below the front silicon surface (*Y* = 0).

The prime difference is in the read mechanism, where FED operates with a feedback mechanism in forward blocking state [33-37,40], with a positive bias at anode (V_A) while a positive and negative bias are applied at G1 and G2, respectively to maintain the induced *pn* regions (Fig. 4.2). The cathode voltage is ramped down to zero volts from a positive bias in the previous operation (write '1') that diffuses electrons towards G2. The electrons from cathode accumulate under G1, being biased positive and further lowers the barrier for hole diffusion from anode towards G1. Similarly, positive bias at anode reduce the hole barrier between anode and G1 region, the holes accumulate under G2, lowering barrier for electrons originating from cathode. The electrons transit from cathode towards anode, while the holes from anode towards cathode [40]. The feedback mechanism result into a high on-current. Thus, the read operation in FED, as well as in a similar architecture, Z^2-FET [38-40] is based on forward mechanism, while TFET operates in reverse bias [41-44].

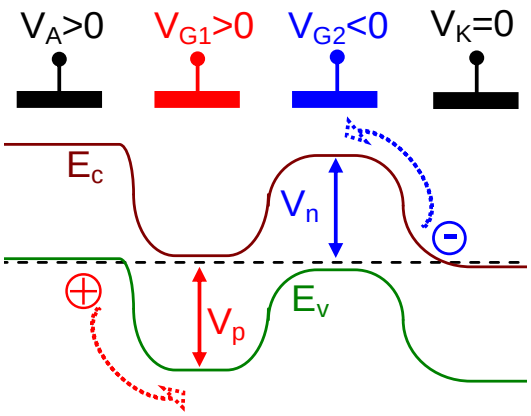

Fig. 4.2 Schematic representation showing read mechanism in FED based DRAM through energy band diagram. Read is based on drift-diffusion mechanism with a feedback action that results into high on-current.

The virtually induced *n*-type region in TFET is positively biased that tunnel electron from source towards G1 (Fig. 4.3). These electrons diffuse towards silicon film under G2, the positive bias at G2 lowers the barrier for higher electron diffusion, and finally, the electrons are drifted towards drain with a positive bias at drain. The tunneling based read mechanism limits the on-current of the memory. In FED, during write '1', the bias at the G1 which is held at positive bias is pulsed down to zero volt and then back to positive value. This lowers the hole barrier from anode and the hole are accumulated under G2, while in TFET the hole accumulation is based on forward biasing the p^+/i region [41-43] and in our work through reverse biasing the n^+/i region. The presence of both types of carriers avails better option. The current during write '1' in FED is in the range of milli-Amperes, while that in reverse-biased write '1' in TFET is in the micro-Amperes. Thus, the tunneling based write mechanism in TFET is more power efficient than that utilized in FED [37].

The other characteristic that distinguishes the FED and the twin gate TFET as DRAM is the distinction between the two states. The carrier profile for FED after write '1' (W1) and '0' (W0) demonstrate that the interpretation of physical memory as the accumulation (state '1') and depletion (state '0') is inappropriate [37]. This is evident from Fig. 4.4 that shows the hole concentration (n_h) in the storage region after write '1' and '0' are almost similar. In reference [37], for memory state '1' $n_h = \sim 3\times10^{19}$ cm^{-3} and '0' $n_h = \sim 10^{19}$ cm^{-3} in the region under

G2 that serves as storage area of FED structure. The states are distinguished based on presence (state '1') and absence (state '0') of deeply depleted regions (as defined in [37]) associated with two *pn*-junctions on the sides of the virtually induced *p*-type region (under G2) as illustrated in schematic representation of FED in Fig. 4.4.

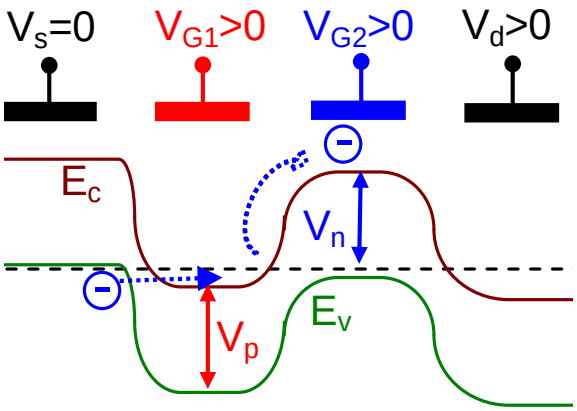

Fig. 4.3 Schematic representation showing read mechanism based on Band-to-Band Tunneling (BTBT) in TFET based DRAM.

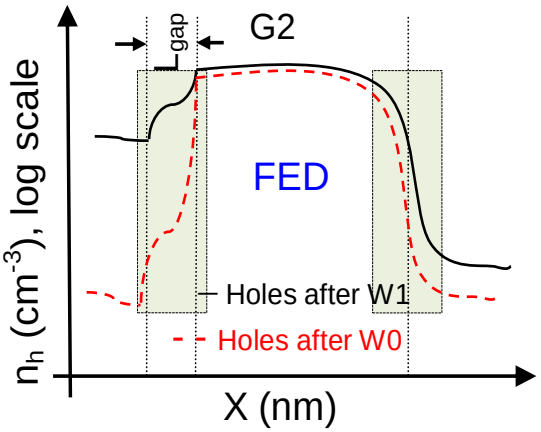

Fig. 4.4 Schematic illustration of hole concentration (n_h) in the FED structure after write '1' (W1) and '0' (W0) showing the states are distinguished based on the deeply depleted (state '1') and depleted (state '0') regions at the *pn*-junctions on either side of G2 (highlighted regions).

This is observed due to dynamic forward biasing of anode with the falling edge of G2 in write '0' operation that accumulates holes in the film region under G2 [37]. For TFET, as shown in Fig. 4.5, after write operations, the states are distinguished based on the presence of excess holes that is, accumulation (state '1') and

shortage of holes that is, depletion region (state '0') in the storage region, under G2. The analysis provides a qualitative comparison.

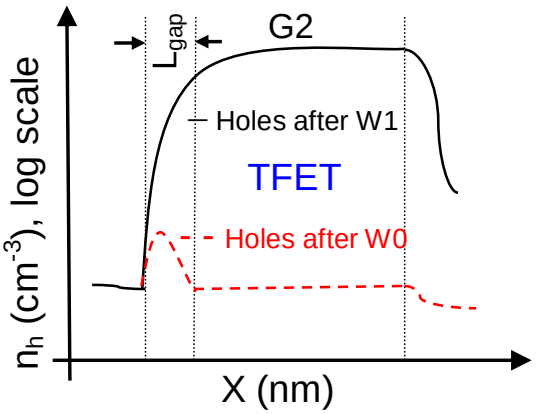

Fig. 4.5 Schematic illustration of hole concentration in the TFET structure after write '1' and '0', showing the states are distinguished based on the accumulated (state '1') and depleted (state '0') regions under G2.

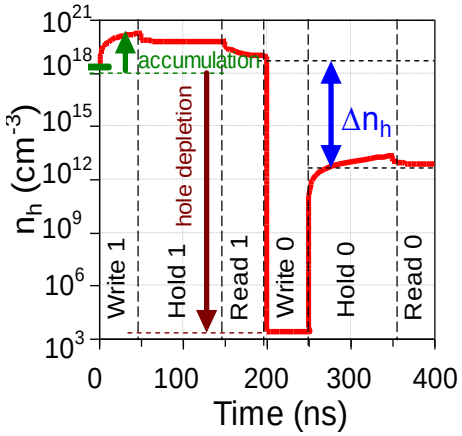

Fig. 4.6 Variation in hole concentration (n_h) in the potential well showing DRAM functionality in the sequence of operation.

4.3 Twin gate as Dynamic Memory

Fig. 4.1(a) illustrates tunneling based twin gate p^+-i-n^+ TFET device that employs two gates with different functionality for each to operate as DRAM. As discussed in chapter 2 and 3, the n^+ poly first front gate (G1) aligned to the source junction governs the read based on BTBT. The p^+ poly second front gate (G2) with a gap (L_{gap}) of 50 nm between the gates creates and preserves the physical well for charge storage (Fig. 4.1(b)).

Table 4.1 Operating mechanism

Operation	***Action under G2 (Fig. 4.6)***	***Bias***	***Mechanism***
W1	Hole accumulation (~10^{19} cm^{-3}) Fig. 4.7(a)	V_{g2} = -2 V → potential at G2 lowers, Fig. 4.7(b)	BTBT, Fig. 4.7(c)
W0	Hole depletion (~10^{6} cm^{-3}) Fig. 4.7(a)	V_{g2} = 2 V → potential raise, Fig. 4.7(b)	Forward bias, Fig. 4.7(d)
H1	Holes recombine with time (~10^{19} cm^{-3} @ 100 ns)	V_{g2} = -0.2 V → barrier lowering with time, Fig. 4.8 (a)	Thermal REC + diffusion of holes [13]
H0	Holes generate with time (~10^{13} cm^{-3} @ 100 ns)	V_{g2} = -0.2 V → barrier rise with time, Fig. 4.8 (b)	Thermal GEN + BTBT [13]

4.3.1. Physical insights into operating mechanism

The functionality as dynamic memory is based on redistribution of the generated holes to perform distinctly as write, read and hold operation for state '1' and '0' along with the optimized set of bias. Thus, regulating the hole concentration in the storage region defines the different operating regions of DRAM as demonstrated in Fig. 4.6 for twin gate. The work aims at describing different mechanism through energy band diagrams. Table 4.1 illustrates the operating principle, the phenomenon occurring and its consequence. The storage of excess holes define write '1', which is based on the generation of holes ($dQ_{G2}/dt > 0$) in the region under G2 [58,59] Q_{G2} is the stored charge in the potential well, given by $Q_{G2} = Q_0 + \Delta Q$, where Q_0 is the induced hole charge at zero bias condition due to p^+ poly gate (G2) and ΔQ is the excess hole charge accumulated in the potential well, during write '1'. Storage of bit '0' or write '0', is based on the recombination ($dQ_{G2}/dt < 0$) that removes the holes from the physical well. The charge sustenance is evaluated through hold time, classified as *RT*. A more negative bias shows more hole generation than recombination, confirmed through the difference in barrier heights of the hold '0' (~0.3 eV) and hold '1' (~0.1 eV) for hold time from 1 ms to 500 ms for a hold bias of -0.2 V at G2. Thus, maintaining the well-defined barrier for each state shows longer charge retention.

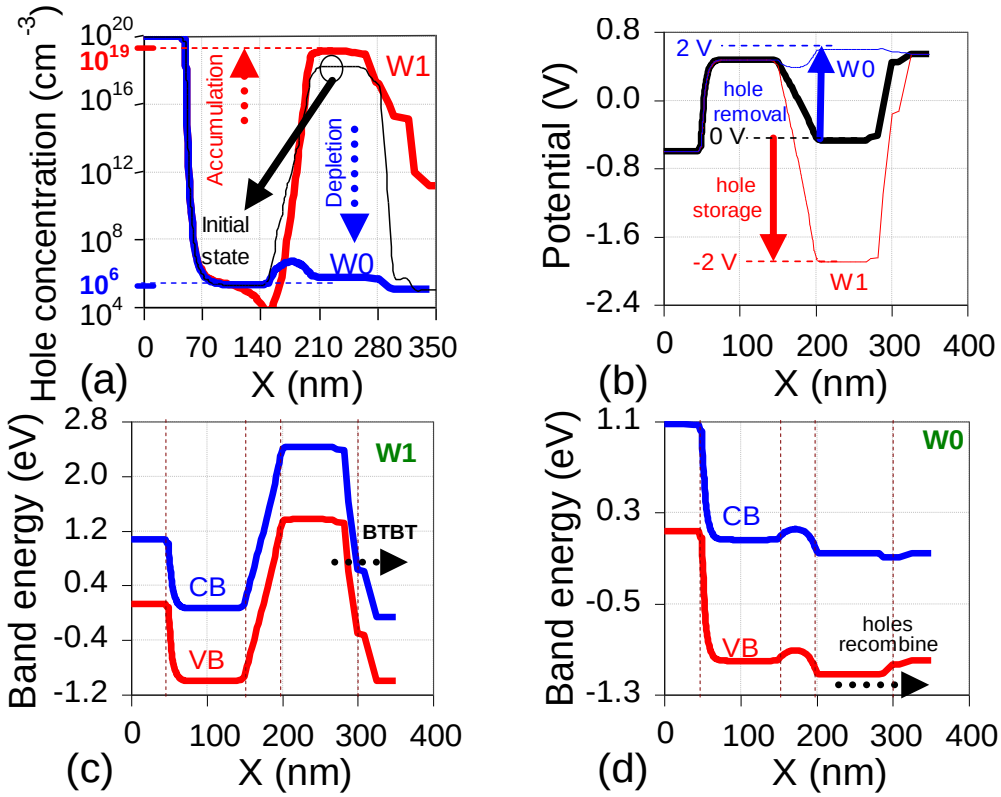

Fig. 4.7 Variation in (a) hole concentration, and (b) electrostatic potential for write '1' (W1) and write '0' (W0) compared to initial case at zero bias, along *X*. Energy band profile showing (c) W1 and (d) W0 operation in TFET along *X*. The cutlines are taken 1 nm below the front surface.

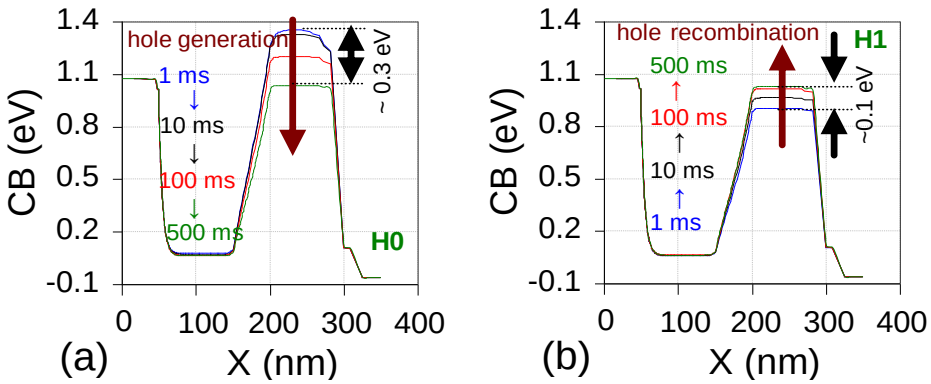

Fig. 4.8 Dependence of band energy during (a) hold '0' (H0) and (b) hold '1' (H1) on the hold time.

An optimized bias of -0.2 V at G2 is applied to balance the generation and recombination that shows maximum retention (Fig. 4.9). Fig. 4.10 reflects the variation in hole concentration with time, where subsequently with increasing hold time, hole generation increases, while hole concentration is almost maintained for state '1' in the storage region due to application of a negative bias

of -0.2 V at G2. Presence of excess holes or application of a positive bias results into barrier lowering that reduces hole concentration, while deficiency of holes or a negative bias at G2 lead to hole generation in the storage region.

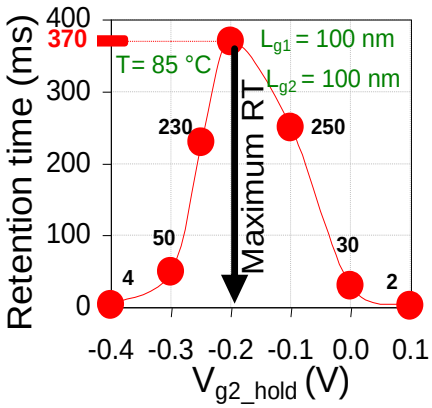

Fig. 4.9 Variation in *RT* with bias applied at G2 during hold operation.

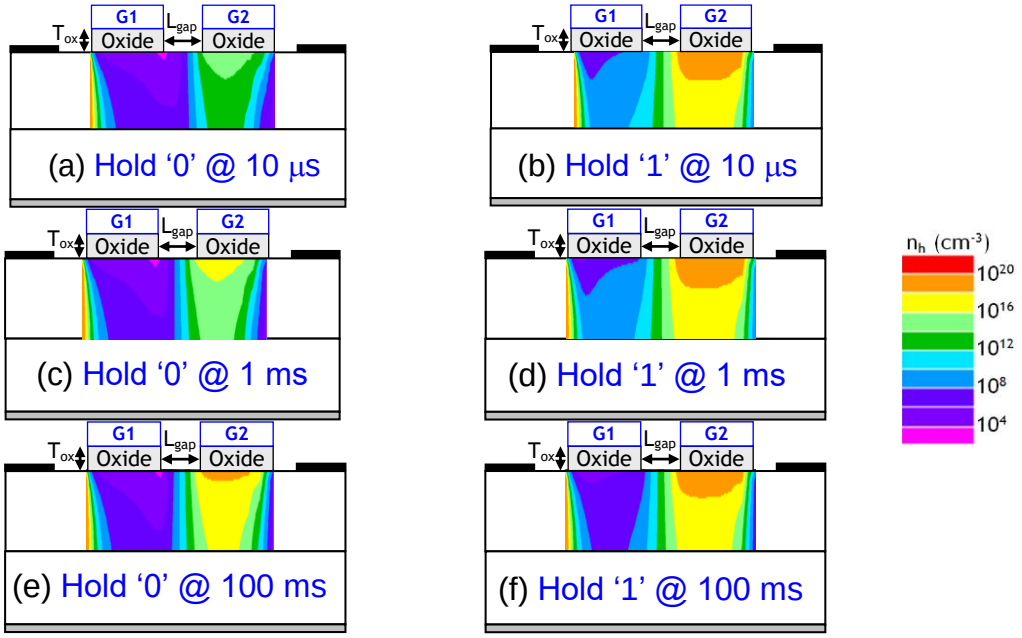

Fig. 4.10 Contour plots of TFET based DRAM illustrating hole concentration (n_h) for hold times of (a)-(b) 10 μs, (c)-(d) 1 ms, and (e)-(f) 100 ms for hold '0' and '1', respectively for L_{g1} = L_{g2} = 100 nm. An optimized negative bias of -0.2 V maintains state '1', while state '0' retention is gradually reduced with hole generation.

Further, as illustrated in Fig. 4.11(a) during read a positive potential of V_{g1} = 2 V results into tunneling of electrons from heavily doped source region towards channel. The virtually induced *p*-type region serves as a barrier for electron flow

which is lowered through a bias of V_{g2} = 1.2 V and finally, the electrons are drifted towards drain with V_d = 1.0 V. Presence of excess holes in the potential well for state '1' is evident from maintenance of higher electrostatic potential during read '1', compared to read '0' (Fig. 4.11(b)).

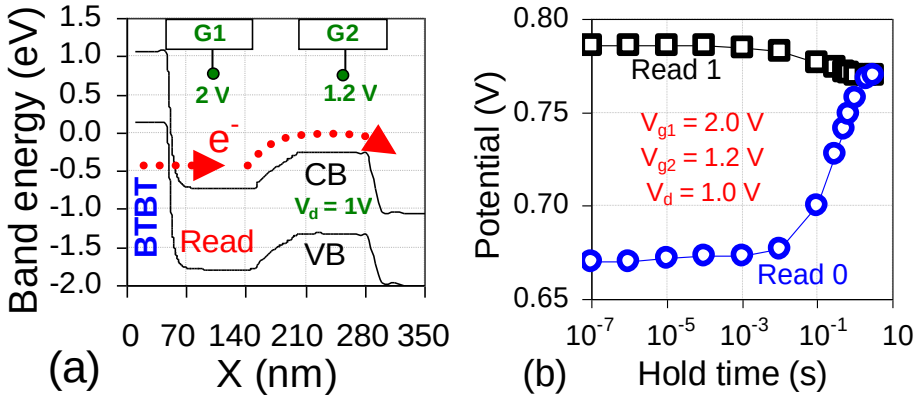

Fig. 4.11 (a) Conduction Band (CB) and Valence Band (VB) energy along *X* during read operation, where electron tunneled from source is drifted towards drain through biases: V_{g1} = 2 V, V_{g2} = 1.2 V, V_d = 1 V (b) Variation in potential for state '1' and '0' during read operation with time.

The lowering of hole barrier at film under G2 from write '1' to read '1', as illustrated in Fig. 4.12(a) result into hole diffusion from storage region and therefore, an optimal bias at G2 should be adopted to sense the difference between the two states. The channel corresponding to G2 during read '1' is formed at back gate, demonstrated through Fig. 4.12(b) that shows a higher barrier for electrons at the front surface. The same has been verified in Fig. 4.13(a)-(b) through electron current density that shows an inclined channel formation during read '1' (Fig. 4.13(a)). A higher potential under the silicon film at G2 region due to excess positive charges, result into a lower barrier of for electrons in read '1' as compared to read '0'. The reduced barrier allows more electron diffusion from G1, which are further drifted towards drain. This can also be concluded from Fig. 4.13(a) and (b) showing higher current density for state '1', compared to state '0', and thus, difference in read currents. Thus, the elevation and reduction in the energy barrier underneath G2 distinguishes the states. The sustenance of each state is determined as retention time, shown in Fig. 4.14(b) where a *RT* = ~370 ms is observed for gate lengths of 100 nm each with a *SM* of ~30 nA at 85℃.

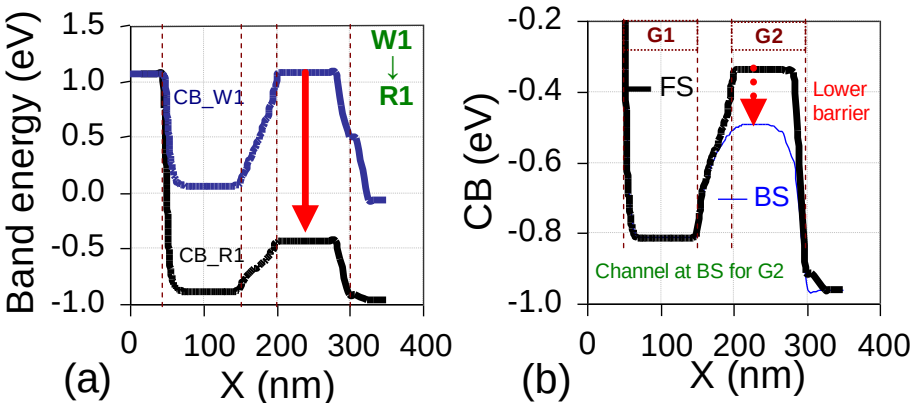

Fig. 4.12 (a) Dependence of band energy from W1 to read '1' (R1) along the channel direction at a cross-section 1 nm below front surface (Y = *0*). (b) Variation in CB profile for R1 at 1 nm below front surface (FS, Y = 0 nm) and 1 nm above back surface (BS, $Y = T_{si}$) at the film under G2.

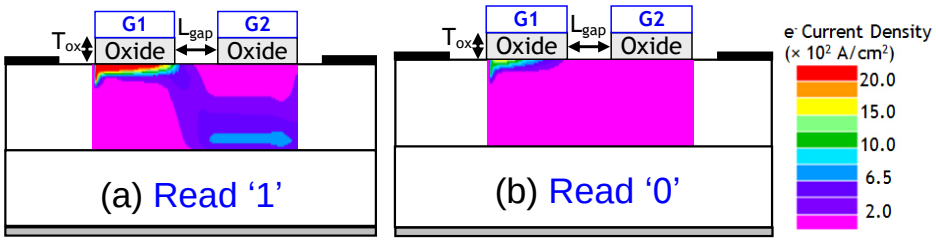

Fig. 4.13 Contour plots illustrating electron current density in the channel for (a) read '1' and (b) read '0'.

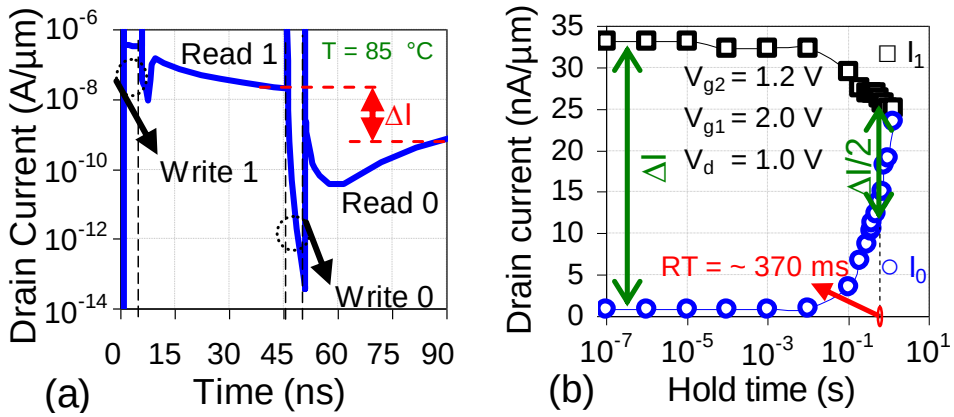

Fig. 4.14 Transient currents with *SM* (ΔI) of ~ 30 nA/μm and write time of 5 ns. Read currents at 85 °C with *RT* = ~ 370 ms, for $L_{g1} = L_{g2}$ = 100 nm.

4.4 Other DRAM metrics

4.4.1 Write speed and power

The other two important DRAM metric are speed and power, quantified through write operation [60-61]. The analysis is significant for application as embedded

memory. The write '1' mechanism basically speeds up the hole generation, while the excess charges that quantifies the read current is based on hold mechanism [62]. The state '1' and '0' are distinguished based on their currents determined during read operation, followed after the hold operation. Figs. 4.15(a) and (b) show the variation in the hole concentration from write '1' to hold '1' for write voltages of -3 V and -2 V, respectively. Most of the excess charges generated during Write '1' operation with bias voltage as -3 V are lost due to Gate to Source/Drain leakage and finally, during hold operation the hole concentration for both the bias voltages are almost same. The reason for hole loss during hold is explained in chapter 3, illustrated through Fig. 3.5. Thus, we obtain similar retention characteristics as well as sense margin for a write voltage of -2 V and -3 V. Therefore, the hole concentration represented as state '1' doesn't depend on the charges accumulated in the potential well during write '1' operation, but on the holes present during hold operation.

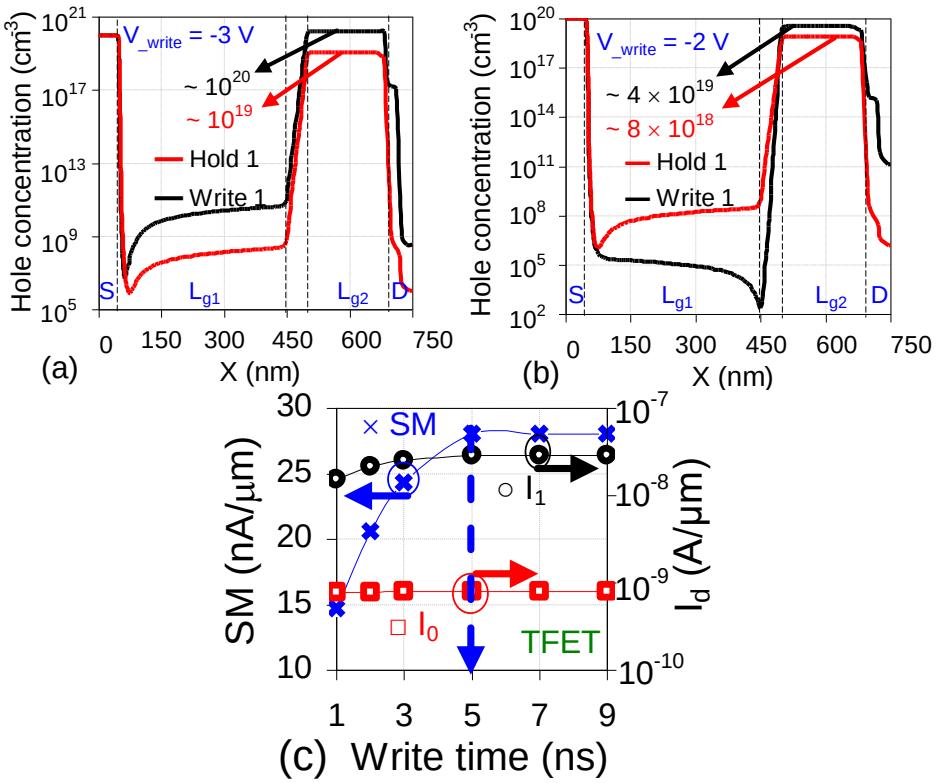

Fig. 4.15 Variation in the hole concentration after write '1' and hold '1' operation with a write bias of (a) -3 V and (b) -2 V at G2 and hold bias of -0.2 V for storage region of 200 nm. (c) Variation of read currents and *SM* with write time.

However, as we increase the write '1' bias to -1 V with the same write time, the hole generation decreases significantly resulting into lower hole concentration. The same hole concentration with a bias of 1 V can be obtained through increase in write time. Write time for a particular set of bias value is evaluated when *SM* attained is constant (Fig. 4.15(c)). Thus, either the write time or bias can be optimized for hole generation. While write time determine the speed of the DRAM operation, write voltage estimate the power consumed. Thus, power and speed associated with write operation of DRAM are another crucial DRAM metrics and must be optimized to reduce trade-off between them.

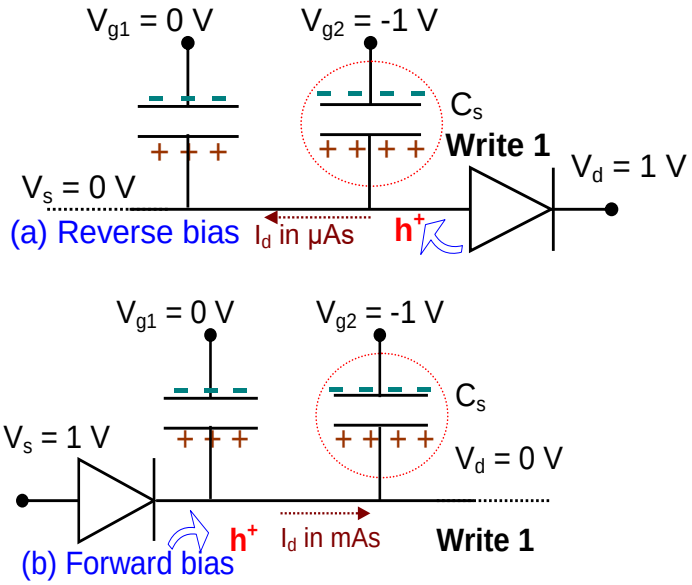

Fig. 4.16 Schematic illustration of charge storage during write (W) for state '1' through (a) reverse bias and (b) forward bias mechanism with the drain currents during the operation.

The two write mechanisms that can be utilized for hole accumulation is based on forward and reverse bias action. The p^+/i junction on being forward biased with a positive bias at source and negative bias at G2 accumulate holes under G2, while n^+/i junction is reversed with a positive bias at drain and a negative bias at G2 to perform as in Fig 4.7(c). To demonstrate their impact, the bias difference between G2 and drain/source is maintained at -2 V as shown in Fig. 4.16(a) for reverse bias and Fig. 4.16(b) in forward bias. Although both mechanisms provide better reliability in comparison to W1 based on impact ionization [63], and also, improves the speed with write time reduced to 5 ns for storage region of 100 nm.

However, reverse biased based methodology is power efficient (drain current, I_{ds} in μA) than forward bias mechanism (I_{ds} in mA). Considering the I_{ds} in μAs as compared to previous method [44] with comparable write time (Fig. 4.17) reverse bias based operation is useful for low power operation. Thus, write mechanism based on BTBT is a power efficient methodology with low write time and reliable compared to that based on impact ionization. The write time with power consumed is compared with previous architectures in Table 4.2, according to methodology suggested in [60]. Thus, the device shows feasibility for eDRAM.

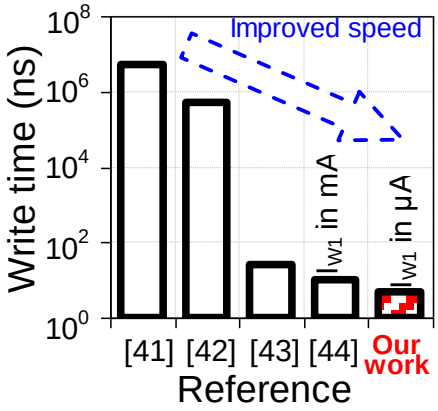

Fig. 4.17 Variation in write time for various TFET based DRAM. I_{w1} is the current during write '1'. Our work is for twin gate with L_{g1} = L_{g2} = 100 nm with bias voltage described in previous section.

Table 4.2 Comparison of write time and power during write '1' for twin gate, compared to previous architectures.

Device architecture [Reference]	***L_s (nm)***	***$\lvert V_d \rvert$ (V)***	***I_d (μA)***	***Power (μW)***	***Write time (ns)***
FDSOI (n^+-p^+-n^+) [60]	65	1.1	314	345	5
DG finFET (n^+-p^+-n^+) [61]	60	2.0	400	800	1
FED (p^+-i-n^+) [36]	400	1.2	~100	120	4
Z^2-FET (p^+-i-n^+) [38]	400	1.3	500	650	1
Our work (twin gate)	100	1.0	0.3	0.3	5

4.4.2 Temperature dependent retention characteristics

Temperature analysis is essential due to application specific requirements. The testing for DRAM cells is performed at 85 °C, specifically for stand-alone applications. However, in real time applications the temperature raises to ~100 °C

due to heating up in the computing systems, and thus, verification of DRAM cells is required at higher temperature [13]. Usually, DRAM performance is evaluated till 125 °C. On the other hand, mobile based embedded memory applications [8] operate reliably at even lower temperatures can be well-suited for eDRAMs [9,13], but with high speed operating capability. Analysing the retention characteristics for wide range of temperature shows the device performance degrades at 125 °C due to increased recombination/generation rate [65], the retention time for L_{g1} = 100 nm is > 64 ms up till 105 °C (Fig. 4.18). The *RT* attained at room temperature is ~1.3 s that shows a significant improvement over previous results [42-44].

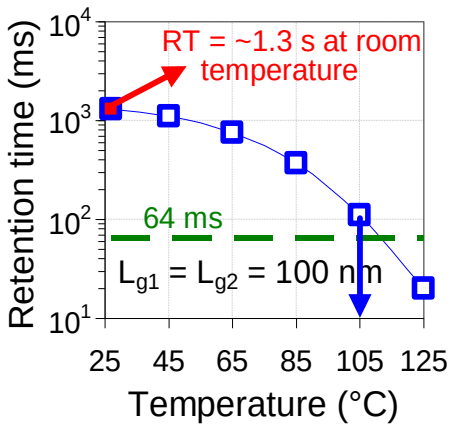

Fig. 4.18 Variation in *RT* with temperature for $L_{g1} = L_{g2}$ = 100 nm.

4.4.3 Results at longer gate lengths

The results shown above are for shorter gate lengths of 100 nm. However, the concept has been first verified at longer gate lengths with G1 as 400 nm and G2 as 200 nm, as utilized in previous topologies [42] and chapter 2. The results for longer gate lengths have been summarized in Fig. 4.19, where a negative bias at the second gate (V_{g2} = -3 V) increases the hole concentration in the potential well due to BTBT between drain and G2, during write '1'. A positive bias at G2 (V_{g2} = 3V) applied during write '0' removes holes from the well which recombine at the drain. During hold operation a small negative bias is applied at Gate-2 (V_{g2} = -0.2 V) that retain holes (~10^{19} cm^{-3}) in the physical well after write '1' (~10^{20} cm^{-3}) whereas after write '0' (~10^{6} cm^{-3}) the potential well is depleted of holes (Fig. 4.19(a)). A deep dedicated volume created using p^+ poly G2 along with an optimum bias sustains holes for several seconds at 85 °C (Fig. 4.19(b)). Fig 4.19

(c) shows a *SM* of ~ 140 nA with a current ratio (> 10^2) exhibiting an acceptable read sensitivity. *RT* achieved is ~1.5 s (Fig. 4.19(d)) for the optimized bias values at 85 °C with L_{g1} = 400 nm and L_{g2} = 200 nm. Further analysis at these gate lengths indicates, *RT* > 64 ms, at a higher temperature of 125 °C (Fig. 4.19(e)).

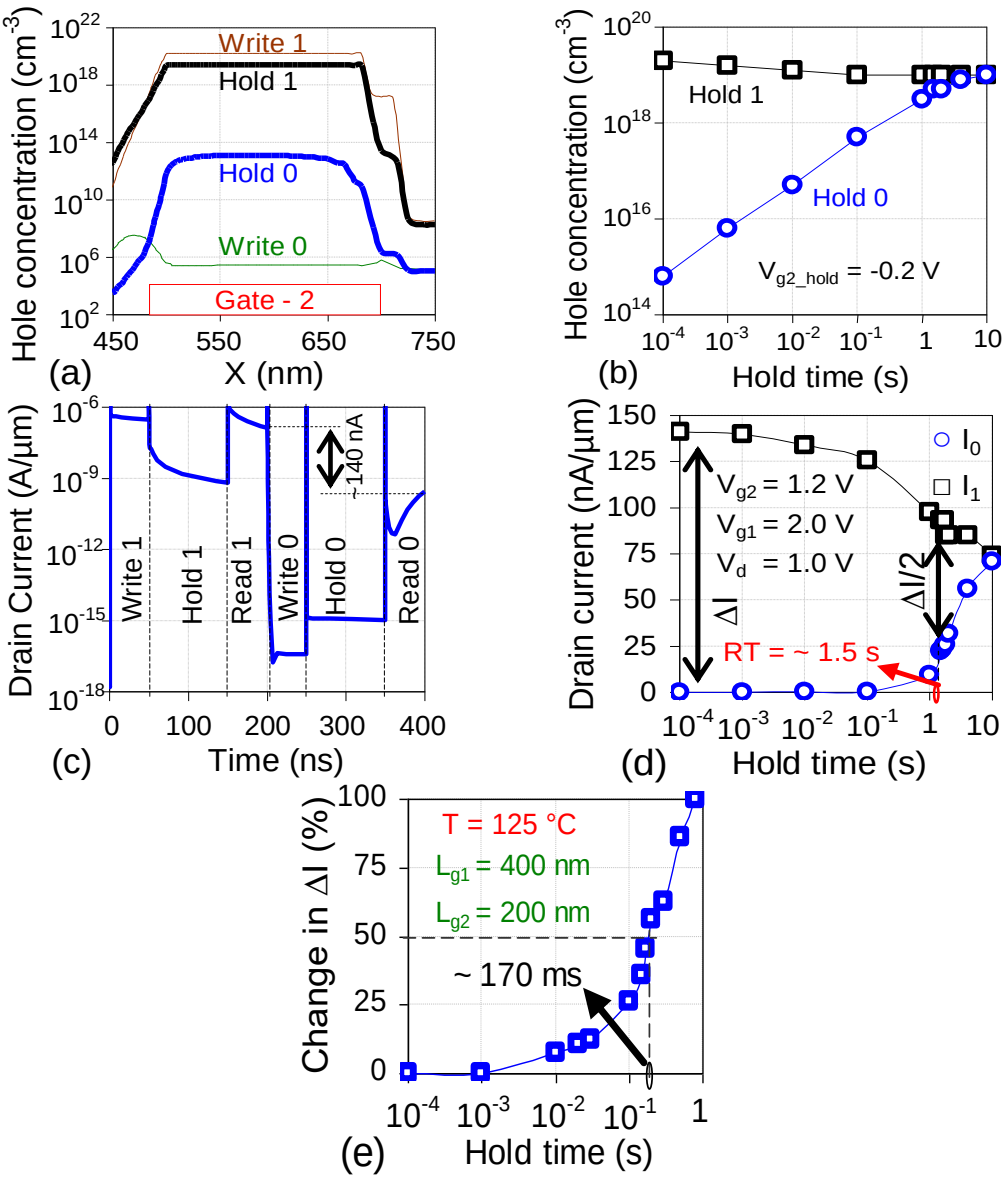

Fig. 4.19 Variation of hole concentration in the storage region (a) during write and hold (1 μs) and (b) in the hold state, with time. (c) Drain current transients in the sequence of operation with bias scheme: write '1': V_{g2} = -3 V, hold: V_{g2} = -0.2 V, read: V_{g1} = 2 V, V_{g2} = 1.2 V, V_d = 1 V, write '0': V_{g2} = 3 V, V_{g1} = V_d = 0 V. (d) Variation of read currents at optimized bias voltages with hold time. (e) Percentage variation in *SM*, where 50% change indicates *RT* of 170 ms at 125 °C.

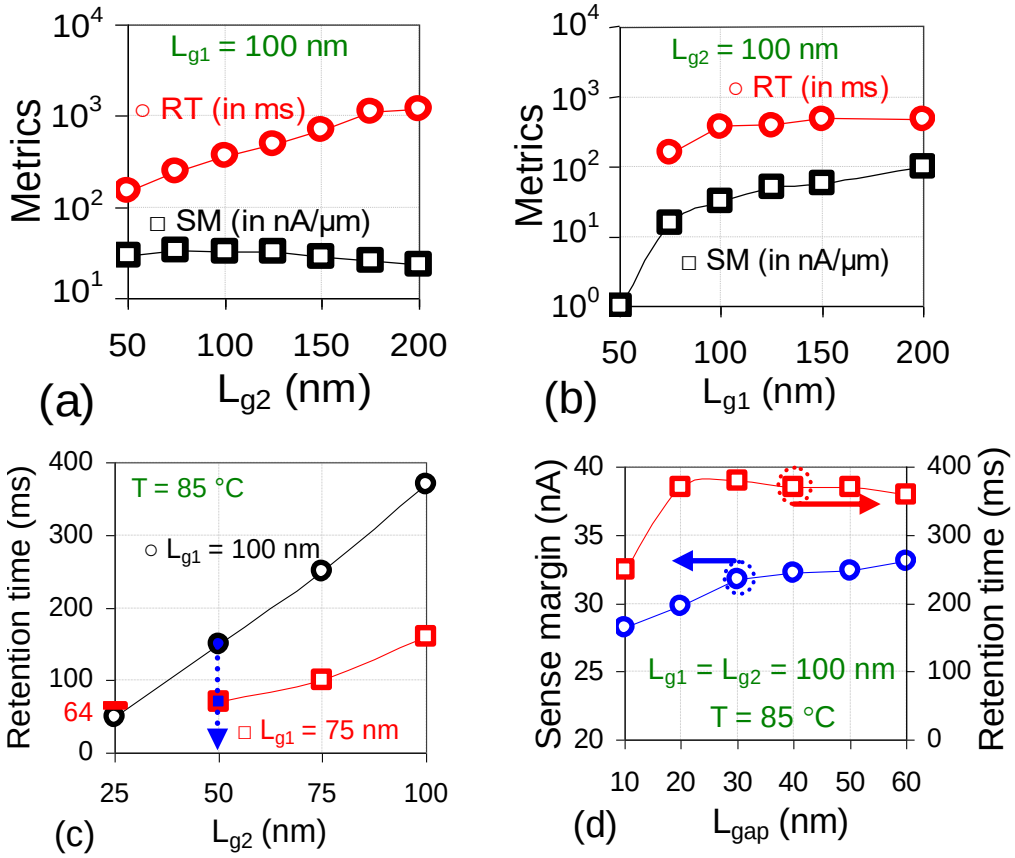

Fig. 4.20 Variation in DRAM metrics, *RT* and *SM* on length of (a) Gate-2 with L_{g1} = 100 nm, (b) Gate-1 with L_{g2} = 100 nm. (c) Variation in *RT* with L_{g2} and with L_{g1}. (d) Dependence of *RT* and *SM* on L_{gap}.

4.4.4 Gate length scalability

The key functionality of the gates is well illustrated in Figs. 4.20(a) and (b) that shows the *SM* is more influenced by G1 while *RT* by G2. The *RT* has significant effect of Gate-1 for L_{g1} < 100 nm. The scaling of L_{g1} results into reduced controllability of Gate-1 due to narrow barrier [55] that affects the read mechanism and degrades the *SM*. Further evaluation demonstrates negligible sense margin for L_{g1} beyond 75 nm. Scaling of Gate-2 beyond 50 nm reduces the storage region, thereby decreasing the *RT* (< 64 ms). The evaluation of device performance highlights its capability to scale L_{g1} till 75 nm with L_{g2} = 50 nm, still with *RT* > 64 ms at 85 °C (Fig. 4.20(c)). The device shows a good downscaling due to longer effective length of the storage region (L_{gap} + L_{g2}). Fig. 4.20(d) demonstrates the impact of L_{gap}, showing its scalability down to 20 nm without affecting *SM* and *RT*. While *RT* > 64 ms [18] is the criteria set for stand-alone applications, the retention time for embedded memory is relaxed [9]. eDRAM

cache memory are refreshed frequently (hundreds of μs to few ms) and therefore, can be scaled down to even lower values.

4.4.5 Improved metrics through underlap

As demonstrated in previous chapter, *RT* can be further enhanced by reducing the generation rate for state '0', through the use of an underlap between Gate2/Drain regions. This reduces the lateral electric field and thus, the tunneling rate. This effect becomes prominent for the shorter gate lengths, and therefore, we have analyzed the same on scaled devices (L_{g1} = 100 nm, L_{g2} = 50 nm) by using an optimal underlap (L_{un}) of 15 nm between Gate-2 (G2) and Drain (D) (Fig. 4.21).

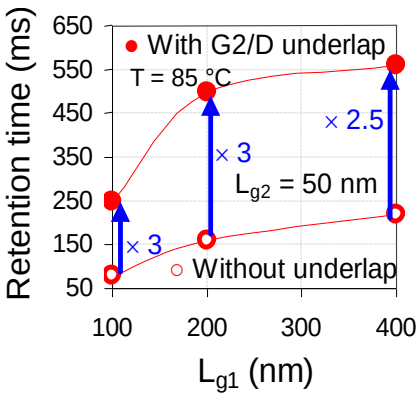

Fig. 4.21 Variation in *RT* with G2/D underlap for different L_{g1}

4.5 Comparison with Misaligned DG TFET

The configuration described in the previous chapters having misaligned topology work can also be realized as twin front gate TFET. The functionality of the devices as DRAM is based on the well-defined role of each gate, where the first gate in twin gate or the front gate in the misaligned DG TFET pre-dominantly regulates the read mechanism. The second gate or the back gate governs the creation and maintenance of potential well. However, the difference lies in location of second gate that affects the conduction channel, and the associated fabrication. The channel corresponding to p^+ poly gate is formed at the surface opposite to its location, therefore, for twin front gate, an inclined channel (from front interface for Gate-1 to back interface for Gate-2) is formed (Fig. 4.22(a)) while a horizontal channel at front surface is observed for misaligned topology (Fig. 4.22(b)). The inclined conduction channel has higher resistance (increases

with T_{si}) for twin gate topology as compared to top-down (misaligned) gate structure.

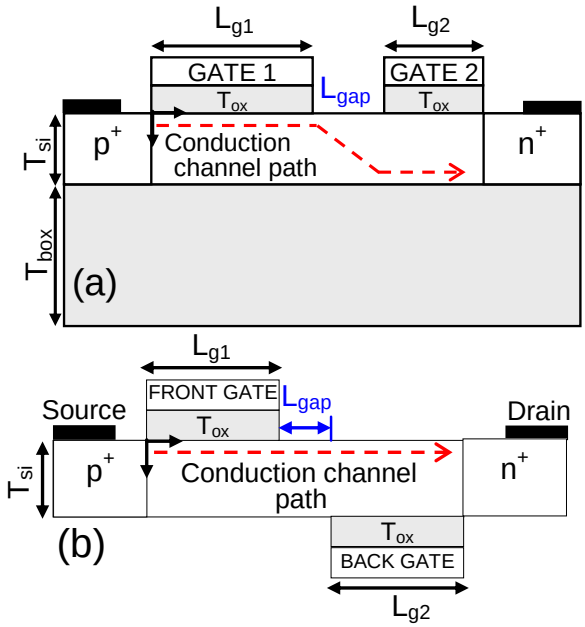

Fig. 4.22 Schematic illustration of (a) twin gate TFET and (b) misaligned DG showing an inclined and horizontal conduction channel, respectively.

The impact might be observable for current values higher in magnitude and in future, top-down architecture could be beneficial. However, at present tunneling based phenomenon limits the variation and also, for DRAM operation in sub-100 nm regime with a silicon film thickness of 20 nm as utilized in our work, both the architectures yield similar results with the first gate utilized for read mechanism based on band-to-band tunneling while the second gate responsible for creation and maintenance of potential well for charge storage. Although the performance is similar, chapter 3 demonstrates misalignment architecture to show systematic design optimization in three steps to show a progress over earlier work [18:T] at scaled gate lengths in sub-100 nm regime.

The planar process for fabricating misaligned DG may be difficult than twin front gate TFET [51-52], but is feasible through shifting of the electrical vernier [57]. The lateral spacing and underlap can be implemented using dual spacer process [51]. However, if we realize these topologies (twin gate and misaligned DG) based on the fabrication technique adopted for vertical finFET- like structures,

which is regarded as the mainstream for lower technology nodes and beyond [65,66], then both the topologies could be fabricated with similar processes. The requirement will be of different masks for the structures. The cross-section and top view of the vertical finFET is illustrated in Figs. 4.23(a) and (b). It should be also noted that the top gate oxide is much thicker than that of the sidewall gate oxide, which can be used as the hard-mask for the gate patterning. After the poly-gate deposition in a vertical finFET structure, the top gate is removed by the Chemical Mechanical Polishing (CMP) process as described in [66] to obtain a finFET-like double-gate structure. The double gate separation can also be achieved by a local etch-back gate process [66]. After that, the respective top-down gates structure and the twin-gate structure can be formed by same gate patterning process using different masks.

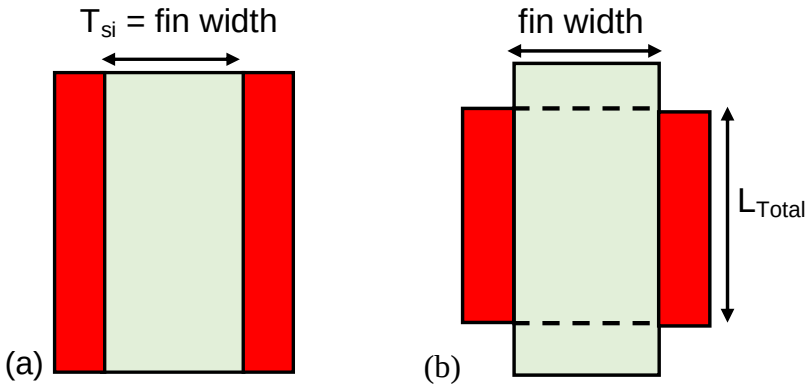

Fig. 4.23 The (a) cross-section and (b) top view of conventional finFET structure.

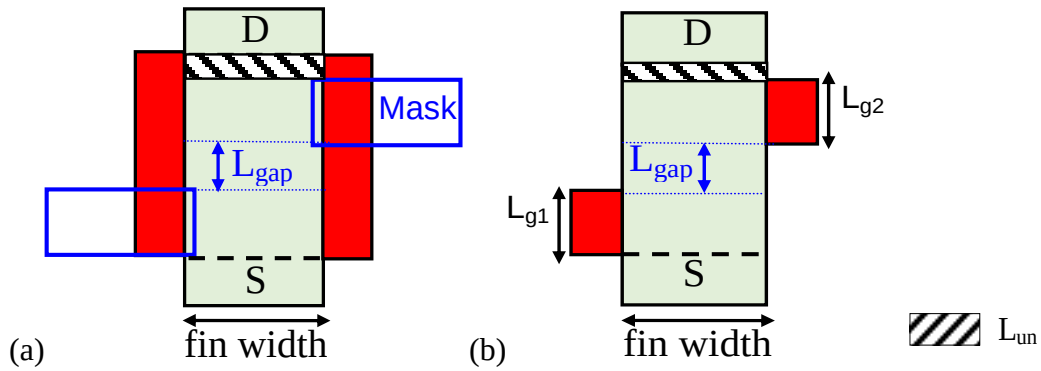

Fig. 4.24 (a) The top view of DG structure before gate patterning. The blue line indicates the gate-patterning masks for top-down gates structure. (b) The top view of misaligned DG after gate patterning.

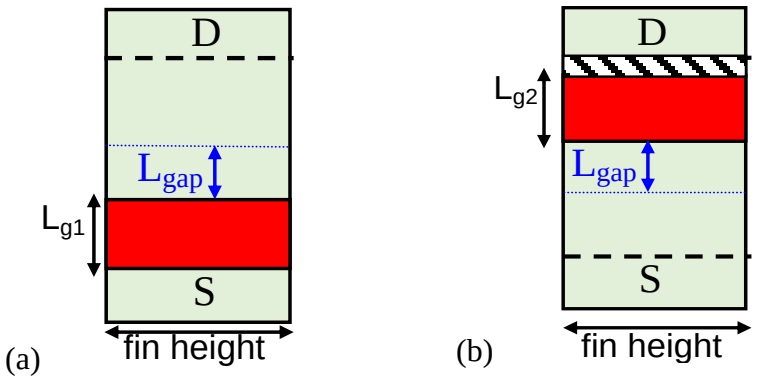

Fig. 4.25 The side view, (a) left and (b) right of misaligned DG structure.

Fig. 4.24(a) indicate the top view of double gate topology, which on gate patterning results into a misaligned double gate (Fig. 4.24(b)). Fig. 4.25 illustrates the side view of misaligned double gate after patterning. Further comparing twin gate topology with other similar architectures show remarkable improvement over the previous TFET based DRAM [41-43] and also, highlights the potential to compete with other steep switching devices [36-40]. The results show improved retention compared to similar architecture, FED with *RT* of ~1.0 second at room temperature for storage region of 400 nm.

Table 4.3 Various TFET structures utilized as DRAM

Device Topology [Reference]	***Length (nm)***		***RT***	***Temperature***
	Gate	***Storage***		
Planar TFET [42]	400	200	in ms	27 °C
Fin-TFET [43]	100	25	100 µs	85 °C
Present work (w/o L_{un})	400	200	1.5 s	
	100	50	80 ms	
Present work (with L_{un})	100	50	260 ms	
Present work (w/o L_{un})	100	50	250 ms	27 °C
Present work (w/o L_{un})	400	200	180 ms	125 °C
	100	200	35 ms	
		50	15 ms	
Present work (with L_{un})	100	50	30 ms	

4.6 Performance assessment and feasibility

Table 4.3 summarizes the results, featuring the enhanced retention time, better scalability, impact of G2/D underlap and improved performance at higher temperature as compared to the previous results [42-44]. A possible layout of twin gate topology is as shown in Fig. 4.26. The operation at circuit level gates will use gates (G1, G2) as wordline (WL) 1 and 2, while the drain as bitline (BL) as adopted in few circuits [67,68]. The unit cell area estimated as $7F^2$ for the present architecture (L_{g1} = 400 nm, L_{g2} = 200 nm) along with high retention time (~1.5 s) is remarkable improvement over the previous results [42,43].

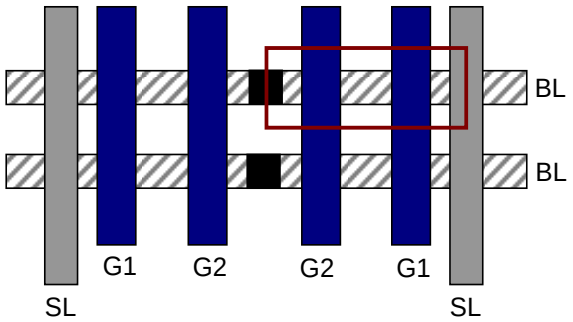

Fig. 4.26 A possible layout of TFET based DRAM. The block highlighted shows a single unit cell.

4.6 Conclusion

In summary, the chapter presents insights into optimized twin gate TFET based capacitorless dynamic memory with improved retention characteristics. The ideology, design and functionality as DRAM are explained through band diagrams by decoupling each operation and highlighting the underlying device physics. The operation is based on controlling the energy barriers with the optimized bias values and architecture that aid to attain a *RT* of ~ 1.5 sec at 85 °C exhibits an improved *RT* at higher temperature (125 °C) for L_{g1} = 400 nm and L_{g2} = 200 nm. Further evaluation at shorter gate lengths with each gate 100 nm long demonstrates,

(i) An improved *RT* (370 ms at 85 °C and ~ 1.3 s at 27 °C)

(ii) Speed (write time of 5 ns)

(iii) Reliability (Write '1' without impact ionization and operation at higher temperature)

(iv) Better scalability (L_{g1} = 75 nm and L_{g2} = 50 nm) and,

(v) Low power operation due to reduced refresh cycles and tunneling based W1 and read mechanism.

Other than reduced refresh cycles, emphasis on write bias and time reflect a low power, and high speed operation, respectively. The results show enhanced performance that could facilitate its applicability as stand-alone as well as embedded applications. However, with technological advancement another important DRAM metric that further needs improvement is the sense margin that will be discussed in the following chapter.

4.7 References

[1] Lee S. H., (2016) Technology scaling challenges and opportunities of memory devices, In Electron Devices Meeting, pp. 1-8.

[2] Itoh K., (2011) Embedded memories: Progress and a look into the future, IEEE Design & Test of Computers, 28, 10-13.

[3] Itoh K. (2008), The history of DRAM circuit designs–at the forefront of DRAM development, IEEE Solid-State Circuits Society Newsletter, 13, 27-31.

[4] Itoh K., Sasaki K., and Nakagome Y., (1995) Trends in low-power RAM circuit technologies, Proceedings of the IEEE, 83, 524-543.

[5] Kim K., and Lee S. Y., (2007) Memory technology in the future, Microelectronic Engineering, 84, 1976-1981.

[6] Kim K., (2005) Technology for sub-50nm DRAM and NAND flash manufacturing, IEEE Electron Devices Meeting, pp. 323-326.

[7] Mutlu O., (2013) Memory scaling: A systems architecture perspective. IEEE Memory Workshop, pp. 21-25.

[8] Micron Technology, Inc. 2015 Winter Analyst Conference

[9] E. Hannigan, McKinsey on Semiconductors, fifth edition, McKinsey & Company , (2015).

[10] Keeth, B., Baker, J. R., Jacob, B., Lin, F., (2007) DRAM circuit design: fundamental and high-speed topics 2nd ed. John Wiley & Sons Inc., New Jersey (ISBN- 978-0-4701-8475-2).

[11] Balestra, F., Kilchytska, V., and Flandre, D., (2014) Functional nanomaterials and devices for electronics, sensors and energy harvesting, Springer International Publishing, Switzerland (ISBN 978-3-319-08803-7).

[12] Zhang, K., (2009) Embedded memories for nano-scale VLSIs, Springer Science and Business Media, Oregon (ISBN 978-0-387-88496-7).

[13] Moore G. E., (1975) Progress in digital integrated electronics, IEEE Electron Devices Meeting, 21, 11-13.

[14] Schaller R. R., (1997) Moore's law: past, present and future, IEEE spectrum, 34, 52-59.

[15] Mollick E., (2006) Establishing Moore's law, IEEE Annals of the History of Computing, 28, 62-75.

[16] Bohr M., (2007) A 30 year retrospective on Dennard's MOSFET scaling paper, IEEE Solid-State Circuits Society Newsletter, 12, 11-13.

[17] ITRS International Technology Working Groups. International Technology Roadmap for Semiconductors, http://www.itrs.net (2015).

[18] Vuong H., and Q. Technologies, (2013) Mobile Memory Technology Roadmap Mobile Devices, Today & Tomorrow.

[19] Al-Fuqaha A., Guizani M., Mohammadi M., Aledhari M., and Ayyash M., (2015) Internet of things: A survey on enabling technologies, protocols, and applications, IEEE Communications Surveys & Tutorials, 17, 2347-2376.

[20] Chang J., Chang L., Dupret A., Kim C., Hamzaoglu F., and Yoshikawa T., (2015) F2: Memory trends: From big data to wearable devices, IEEE International Solid-State Circuits Conference, pp. 498-499.

[21] Ousterhout J., Agrawal P., Erickson D., Kozyrakis C., Leverich J., Mazières D., Mitra S., Narayanan A., Parulkar G., Rosenblum M., and Rumble S. M., (2010) The case for RAMClouds: scalable high-performance storage entirely in DRAM, ACM SIGOPS Operating Systems Review, 43, 92-105.

[22] Zhang H., Chen G., Ooi B. C., Tan K. L. and Zhang M., (2015) In-memory big data management and processing: A survey, IEEE Transactions on Knowledge and Data Engineering, 27, 1920-1948.

[23] Greenberg A., Hamilton J., Maltz D. A., and Patel P., (2008) The cost of a cloud: research problems in data center networks, ACM SIGCOMM computer communication review, 39, 68-73.

[24] Hassan H., Vijaykumar N., Khan S., Ghose S., Chang K., Pekhimenko G., Lee D., Ergin O., and Mutlu O., (2017) SoftMC: A flexible and practical open-source infrastructure for enabling experimental DRAM studies, IEEE International Symposium on High Performance Computer Architecture, pp. 241-252.

[25] Kim K., (2000) Perspectives on giga-bit scaled DRAM technology generation, Microelectronics reliability, 40, 191-206.

[26] Bhati I., Chang M. T., Chishti Z., Lu S. L., and Jacob B., (2016) DRAM refresh mechanisms, penalties, and trade-offs, IEEE Transactions on Computers, 65, 108-121.

[27] Lee D. U., Kim K. W., Kim H., Kim J. Y., Park Y. J., Kim J. H., Kim D. S., Park H. B., Shin J. W., and Cho J. H., (2014) A 1.2 V 8Gb 8-channel 128GB/s high-bandwidth memory (HBM) stacked DRAM with effective micro-bump I/O test methods using 29nm process and TSV, IEEE Solid-State Circuits Conference Digest of Technical Papers, pp. 432-433.

[28] Gamiz F., (2016) Capacitor-less memory: Advances and challenges in Joint International Workshop and International Conference on Ultimate Integration on Silicon, pp. 68-71

[29] Cristolovean S., Lee K.H., Parihar M. S., El Dirani H., Lacord J., Martinie S., Le Royer C., Barbe J. C., Mescot X., Fonteneau P., and Galy P., (2017) A review of the Z 2-FET 1T-DRAM memory: Operation mechanisms and key parameters, Solid-State Electronics, in press.

[30] De Marchi M., Sacchetto D., Frache S., Zhang J., Gaillardon P. E., Leblebici Y., and De Micheli G., (2012) Polarity control in double-gate, gate-all-around vertically stacked silicon nanowire FETs, IEEE Electron Devices Meeting, pp. 183-186.

[31] Ernst T., (2013) Controlling the polarity of silicon nanowire transistors, Science, 340,1414-1415.

[32] Sedighi B., Hu X. S., Liu H., Nahas J. J., and Niemier M., (2015) Analog circuit design using tunnel-FETs, IEEE Transactions on Circuits and Systems I: Regular Papers, 62, 39-48.

[33] Salman A. A., Beebe S. G., Emam M., Pelella M. M., and Ioannou D. E., (2006) Field Effect Diode (FED): A novel device for ESD protection in deep sub-micron SOI technologies, IEEE Electron Devices Meeting, pp. 1–4.

[34] Raissii F., (1996) A brief analysis of the field effect diode and breakdown transistor, IEEE Transaction on Electron Devices, 43, 362–365.

[35] Yang Y., Gangopadhyay A., Li Q., and Ioannou D. E., (2009) Scaling of the SOI field effect diode (FED) for memory application, IEEE Semiconductor Device Research, pp. 1-2.

[36] Badwan A. Z., Chbili Z., Yang Y., Salman A. A., Li Q., and Ioannou D. E., (2013) SOI field-effect diode DRAM cell: Design and operation, IEEE Electron Device Letters, 34,1002-1004.

[37] Badwan A. Z., Li Q., and Ioannou D. E., (2015) On the nature of the memory mechanism of gated-thyristor dynamic-RAM Cells, IEEE Journal of the Electron Devices Society, 3, 468-471.

[38] Wan J., Royer C. Le, Zaslavsky A., and Cristoloveanu S., (2013) Progress in Z^2-FET 1T-DRAM: retention time, writing modes, selective array operation, and dual bit storage, Solid-State Electronics, 84, 147-154.

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

[40] Wan J., Cristoloveanu S., Le Royer C., and Zaslavsky A., (2012) A feedback silicon-on-insulator steep switching device with gate-controlled carrier injection, Solid-State. Electron, 76, 109–111.

[41] Biswas A., Dagtekin N., Grabinski W., Bazigos A., Le Royer C., Hartmann J. M., Tabone C., Vinet M., and Ionescu A. M., (2014) Investigation of tunnel field-effect transistors as a capacitor-less memory cell, Applied Physics Letters, 104, 092108.

[42] Biswas A., and Ionescu A. M., (2015) 1T capacitor-less DRAM cell based on asymmetric tunnel FET design, IEEE Journal of the Electron Devices Society, 3, 217-222.

[43] Biswas A., and Ionescu A. M., (2014) Study of fin-tunnel FETs with doped pocket as capacitor-less 1T DRAM, IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference, pp. 1-2.

[44] Kim H., Park B.-G., (2016) A 1T-DRAM cell based on a tunnel field-effect transistor with highly-scalable pillar and surrounding gate structure, Journal of the Korean Physical Society, 69, 323-327.

[45] Vadizadeh M., (2016) Improving gate delay and ION/IOFF in nanoscale heterostructure field effect diode (H-FED) by using heavy doped layers in the channel, Applied Physics A, 122, 469.

[46] Manavizadeh N., Pourfath M., Raissi F., and Asl-Soleimani E., (2011) A comprehensive study of nanoscale Field Effect Diodes, IEEE International Conference on Thermal, Mechanical and Multi-Physics Simulation and Experiments in Microelectronics and Microsystems, pp. 1-4.

[47] Sheikhian I., and Raissi F., (2003) High-speed digital family using field effect diode, Electronics Letters, 39, 345-347.

[48] Chen W. C., Lue H. T., Hsiao Y. H., and Lu C. Y., (2017) A Novel SuperSteep Subthreshold Slope Dual-Channel FET Utilizing a Gate-Controlled Thyristor Mode-Induced Positive Feedback Current, IEEE Transactions on Electron Devices, 64, 1336-1342.

[49] Touchaei B. J., and Manavizadeh N., (2017) Design and Simulation of Low-Power Logic Gates Based on Nanoscale Side-Contacted FED, IEEE Transactions on Electron Devices, 64, 306-311.

[50] De Marchi M., Sacchetto D., Zhang J., Frache S., Gaillardon P. E., Leblebici Y., and De Micheli G., (2014) Top–Down Fabrication of Gate-All-Around Vertically Stacked Silicon Nanowire FETs With Controllable Polarity, IEEE Transactions on Nanotechnology, 13, 1029-1038.

[51] Park J. S., (2002) Dual Gate MOSFET Fabrication US Patent 6168998 B1.

[52] Anderson, B. A., Bryant, A., Clark Jr, W. F. and Nowak E. J., (2006) Low capacitance FET for operation at subthreshold voltages US patent 7009265 B2 2006.

[53] Chung T. M., Olbrechts B., Sodervall U., Bengtsson S, Flandre D., and Raskin J.-P., (2007) Planar double-gate SOI MOS devices: fabrication by wafer bonding over pre-patterned cavities and electrical characterization, Solid-State Electronics, 51, 231-238.

[54] Jeon Y., Kim M., Lim D., and Kim S., (2015) Steep subthreshold swing n- and p-channel operation of bendable feedback field-effect transistors with p^+–i–n^+ nanowires by dual-top gate voltage modulation, Nano Lett., 15, 4905–4913.

[55] Dirani H. E., Solaro Y., Fonteneau P., Ferrari P., and Cristoloveanu S., (2016) Properties and mechanisms of Z^2-FET at variable temperature, Solid-State Electron, 115, 201–206.

[56] Chen W. C., Lue H. T., Hsiao Y. H., and Lu C. Y., (2017) A Novel SuperSteep Subthreshold Slope Dual-Channel FET Utilizing a Gate-Controlled Thyristor Mode-Induced Positive Feedback Current, IEEE Transactions on Electron Devices, 64, 1336-1342.

[57] Widiez J., Lolivier J., Vinet M., Poiroux T., Previtali B., Daugé F., Mouis M., and Deleonibus S., (2005) Experimental evaluation of gate architecture influence on DG SOI MOSFETs performance, IEEE Transactions on Electron Devices, 52, 1772-1779.

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

[59] Fossum J. G., Lu Z. and Trivedi V. P., (2007) New insights on "capacitorless" floating-body DRAM cells, IEEE Electron Device Letters, 28, 513-516.

[60] Puget S., Bossu G., Fenouiller-Beranger C., Perreau P., Masson P., Mazoyer P., Lorenzini P., Portal J. M., Bouchakour R., and Skotnicki T., (2009) FDSOI Floating Body Cell eDRAM Using Gate-Induced Drain-Leakage (GIDL) Write Current for High Speed and Low Power Applications, IEEE International Memory Workshop, pp. 1-2.

[61] Hou J., Shao Z. and Miao X., (2009) A High speed low power capacitorless SOI-DRAM cell using impact ionization and GIDL effect, Electron Devices and Solid-State Circuits, pp. 517-520.

[62] Giusi G., (2014) Physical insights of body effect and charge degradation in floating-body DRAMs, Solid-State Electronics, 95, 1-7.

[63] Aoulaiche M., Nicoletti T., Almeida L. M., Simoen E., Veloso A., Blomme P., Groeseneken G., and Jurczak M., (2012) Junction Field Effect on the Retention Time for One-Transistor Floating-Body RAM, IEEE Transactions on Electron Devices, 59, 2167-2172

[64] Sasaki K. R. A., Nicoletti T., Almeida L. M., dos Santos S. D., Nissimoff A., Aoulaiche M., Simoen E., Claeys C., and Martino J. A., (2014) Improved

retention times in UTBOX nMOSFETs for 1T-DRAM applications, Solid-State Electronics, 97, 30-37.

[65] Yu, B., Chang, L., Ahmed, S., Wang, H., Bell, S., Yang, C.Y., Tabery, C., Ho, C., Xiang, Q., King, T.J. and Bokor, J., (2002) FinFET scaling to 10 nm gate length, IEEE Electron Devices Meeting, pp. 251-254.

[66] Masahara, M., Endo, K., Liu, Y.X., O'uchi, S., Matsukawa, T., Surdeanu, R., Witters, L., Doornbos, G., Nguyen, V.H., Vrancken, C. and Jurczak, M., (2007), Four-terminal FinFET device technology, Integrated Circuit Design and Technology.

[67] Lu Z., Fossum J. G., and Zhou Z., (2011) A floating-body/gate DRAM cell upgraded for long retention time, IEEE Electron Device Letters, 32, 731-733.

Chapter 5

Planar Tri-gate Tunnel Field Effect Transistor for Enhanced Performance Metrics

5.1 Introduction

The single transistor capacitorless DRAM [1-10] has overcome scalability issue associated with traditional capacitor in the conventional 1T-1C configuration [11-17]. Further, the voltage scalability concerns of conventional MOS transistor can be circumvented using TFETs [18-22] which exhibit a sub-60 mV/decade off-to-on transition. Although, the tunneling mechanism limits the on-current (I_{on}) a high on-to-off ratio makes it a viable candidate for low power applications. TFET also benefits from reduced short channel effects, weak temperature dependency and enhanced scalability. With technological advancement, the focus has been shifting towards the memory applications which require energy efficient device exhibiting improved performance in terms of speed, retention time, read sensitivity, scalability, and reliability [23].

Previous chapters have underlined various architecture modifications in TFET based dynamic memories. Although, the previous works outlined an opportunity to utilize TFET as DRAM, showed a compromise between sense margin and retention time (Table 5.1). The first concept for TFET based DRAM utilized an oversized back gate where the storage region of 200 nm was created at the front un-gated region [24,25]. The result showed a Sense Margin (*SM*) of 10-20 nA/μm with a Retention Time (*RT*) in few milliseconds at room temperature. The architecture was modified to misaligned Double Gate (DG) TFET, as demonstrated in chapter 2, that resulted into improved retention time of 2 s at 85 °C achieved through back gate engineering. The sense margin was further enhanced to 140 nA by bias optimization in twin gate architecture (chapter 4) with *RT* = 1.5 s at 85 °C. Maximum *SM* of 500 nA could be attained through a fin-

based TFET with *p*-doped pocket as the storage region [26]. However, this topology resulted in a low retention time of 100 μs at 85 °C.

Table 5.1 Previous works on TFET based DRAM

Architecture	L_{total}[a]	*SM*	*RT*
DG FDSOI [25]	600 nm	20 nA	Few ms at 27 °C
Fin-TFET [26]	125 nm	500 nA	100 μs at 85 °C
Twin gate TFET	650 nm	140 nA	1.5 s at 85 °C
Misaligned DG TFET	600 nm	20 nA	2 s at 85 °C
Misaligned DG TFET	160 nm	20 nA	85 ms at 85 °C

[a]L_{total} is the length between source and drain.

A higher *SM* facilitates better read sensitivity, simplifying the sensing circuit, and also, improves the device scalability that extends its applicability as high density memory [8,10], while a high *RT* indicates low refresh rates, thereby, signifying low power operation [11]. Therefore, maintenance of both, high *RT* and *SM* is imperative for an operational DRAM. However, there exists a trade-off between *SM* and *RT*. The storage of excess holes for state '1', shows an improved *SM*, but its sustenance requires a more negative bias during hold that increases hole generation for state '0', and hence, results in its degradation that consequently lowers *RT*. The retention time can be improved by reducing the hole generation associated with BTBT between gate and drain by introducing an underlap (L_{un}). However, an underlap reduces the *SM* or requires higher voltage to uphold the same *SM* [9]. Therefore, a careful investigation and optimization is required to minimize the trade-off between *SM* and *RT*.

Although, previous architectures proposed in this thesis work have shown a noteworthy improvement in terms of *RT*, the limitation in lower *SM* demands insightful understanding and analysis of the associated constraints to overcome the technological obstruction for feasible DRAM operation. Also, insights into chapter 3 and 4 suggests, the first gate of misaligned DG and twin gate could be scaled down to 75 nm, but the *SM* attained was low (< 50 nA) thereby being a hindrance for further scalability. In the previous chapters, the scaling limit was

due to low sense margin, and therefore, further investigation as well as device optimization is required to improve the tunneling process. Thus, this chapter investigates a planar tri-gate TFET with two gates controlling the tunneling mechanism which leads to enhanced *SM*, scalability, while maintaining the retention characteristics and current ratio at lower bias values. The benefit of the device is ascertained through comparison with twin gate TFET.

5.2 Device design for operation as DRAM

The planar tri-gate p^+-i-n^+ structure, shown in Fig. 5.1(a) has a p^+ doped source and n^+ doped drain with a concentration of 10^{20} cm^{-3} and a lightly doped (10^{15} cm^{-3}) p-type channel region between them. The silicon film thickness (T_{si}) is 20 nm and physical oxide thickness (T_{ox}) of 3 nm is used with HfO_2 as the dielectric layer. The two gates (G1) are aligned symmetrically to the source junction at a partial portion of the intrinsic film and are electrically connected, while the second gate (G2) is located at the front interface, adjacent to first gate with a lateral spacing (L_{gap}). The proposed device can be fabricated by adopting the approach as demonstrated in [27-32].

The transmission of carriers in a TFET is based on inter-band tunneling, which is directly proportional to the energy difference ($\Delta\Phi$) between the valence band in the source and the conduction band in the channel, and has inverse dependence on screening tunneling length (λ) [21]. The tunneling length is regulated by the device dimensions, architecture, doping profile and gate capacitance. The tunneling barrier is strongly governed by the electric field associated with the gate, and therefore, increasing the gate controllability over the tunneling junction through use of multiple gate architecture can enhance the drain current [33]. The concept has been utilized in this work to improve the Sense Margin (*SM*) of TFET based dynamic memory, where the two gates are symmetrically aligned at a partial portion of the intrinsic film near the source to control the tunneling of electrons. The concept, design and operation of planar tri-gate architecture provide valuable viewpoints for TFET based DRAM.

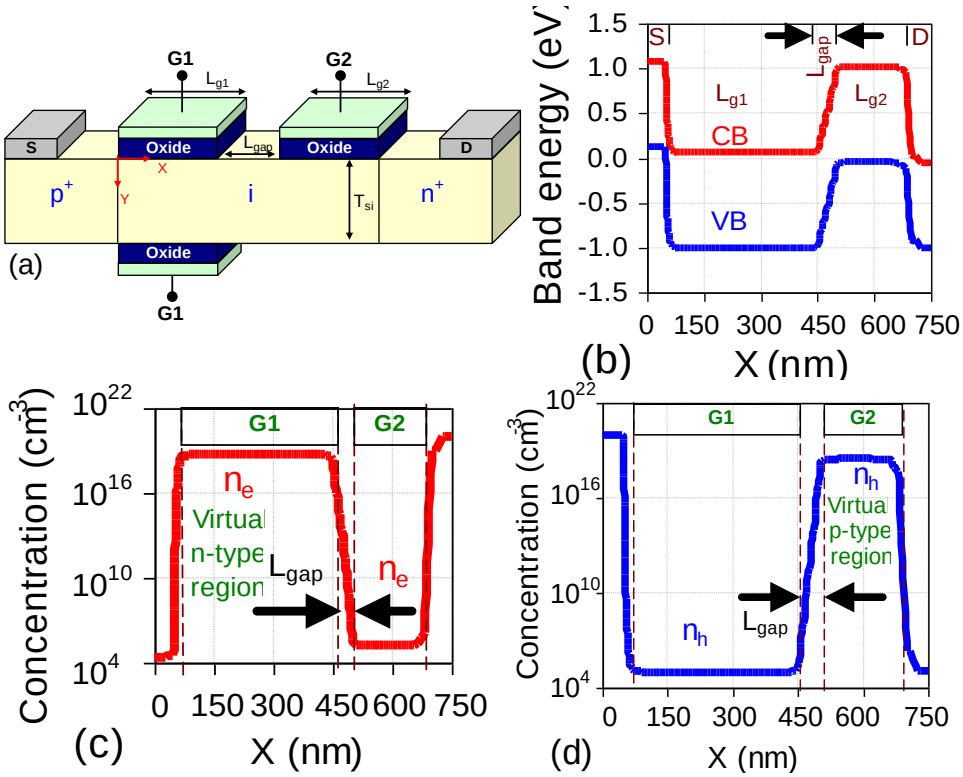

Fig. 5.1 (a) Schematic diagram of planar tri-gate TFET. Variation of (b) band energy, (c) electron concentration (n_e) and (d) hole concentration (n_h) with distance along the channel direction (X) at no bias condition. The cut-lines are taken at a cross-section 1 nm below $Y = 0$.

The key functionality as dynamic memory is based on controlling the barriers formed through the gates (G1 and G2). The injection barriers corresponding to the gates are demonstrated through the band energy in Fig. 5.1(b). As shown in Fig. 5.2(c) G1 with workfunction as 4.15 eV induces a virtual *n*-type region with electron concentration (n_e) of ~ 10^{18} cm^{-3} and primarily controls the read mechanism based on BTBT. The symmetric gate (G1) at the back interface, results into formation of a second conduction channel. Incorporation of gate (G1) at the back surface enhances the gate controllability of the semiconductor film [21] and increases the individual tunneling current for read '0' and '1' states along with the difference between them, and hence, a higher sense margin is expected to be achieved. The use of an additional back gate to regulate the tunneling, not only increases the driving capability of device, but reduces short channel effect with operation at lower bias values. The second gate (G2) with gate length L_{g2} and workfunction of ~5.25 eV accumulates a hole concentration (n_h) of ~ 10^{18} cm^{-3} at

surface that serves as an electrostatic potential well for charge storage (Fig. 5.1(d)). Although G1 controls BTBT, governing the electrons in the channel, further drifting of electrons towards drain is regulated by G2 and distinguished based on the positive charges stored in that region. Therefore, for a high sense margin, it is beneficial to have the front surface beneath silicon film at G2 with higher hole accumulation that increases the potential at the back surface of G2 region, and hence, read '1' current.

5.3 Insights into operation of planar tri-gate

Through an insightful analysis, the proposed tri-gate topology and its applicability as DRAM is evaluated to overcome the bottleneck of low *SM* without compromising charge retention. In order to achieve improved metrics, understanding the significance of each operation and contributing factors is indispensable, and the same is described in the following sub-sections.

5.3.1 Operating mechanisms

The functioning of TFET requires insights into various operations (write, read and hold) performed by dynamic memory. The analysis is based with L_{g1} = 400 nm, L_{g2} = 200 nm and L_{gap} = 25 nm and is further reduced to evaluate the scaling capability for memory operation. Tunneling based write mechanism through a negative bias at G2 (V_{g2} = -2 V) facilitates storage of excess holes beneath G2 (Fig. 5.2(a)) while holes are evacuated from storage region through a positive potential at G2 (V_{g2} = 2 V). While an optimized bias of -0.3 V at G2 has been adopted for hold operation for maximum *RT*, the biases for read operation also need to be focused for *SM*. Before assessing the importance of read biases, the role of each gate for estimation of *SM* is discussed. After write '1', during read '1' (Fig. 5.2(a)) the hole concentration (n_h) reduces beneath G2, influenced by thermal recombination and more positive bias applied at G2 [10].

A positive bias applied at G2 (V_{g2} = 1.2 V) reduces the potential barrier between G1 and G2 that facilitates the flow of holes towards the source. This increases n_h (~ 10^{15} cm^{-3}) at the centre of film under G1 region. The presence of excess carriers for state '1' increases the electrostatic potential of intrinsic film region covered by

both the gates, G1 and G2 (Figs. 5.2(b) and (c)) and the same is higher as compared to state '0'. This results into a lower barrier for electrons (Fig. 5.2(d)) and also, a higher tunneling window for state '1' as compared to state '0' and consequently, an increased current for read '1' is observed as compared to read '0'. Fig. 5.2(c) also illustrates a higher potential at the back surface of the silicon film underneath G2, justifying the channel at the back surface. Thus, the work demonstrates the significance of read operation in determining state '1' and '0', contributed not only by the excess charge carriers, but gate biases and gate coupling. Thus, bias optimization is very crucial in DRAM, and therefore, it is essential to comprehend the role of read biases for DRAM performance metrics.

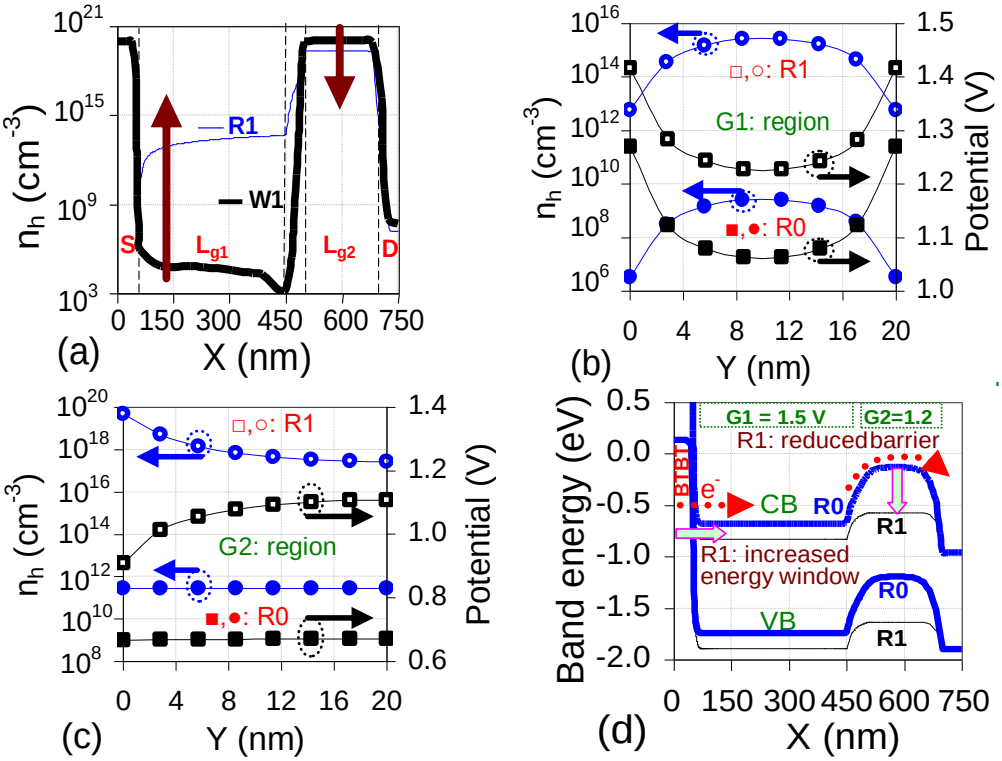

Fig. 5.2 (a) Variation in n_h in write (W) and read (R) operation for state '1', along the channel at 1 nm below the front interface (Y = 0). Dependence of n_h and potential in the semiconductor film beneath (b) G1 and (c) G2 during read operation along Y with X at the centre position of G1, and G2, respectively. (d) Variation in band energies for read '1' and '0' at a cross-section of 1 nm above the back interface ($Y = T_{si}$). Bias during read: V_{g1} = 1.5 V, V_{g2} = 1.2 V, V_d = 0.9 V.

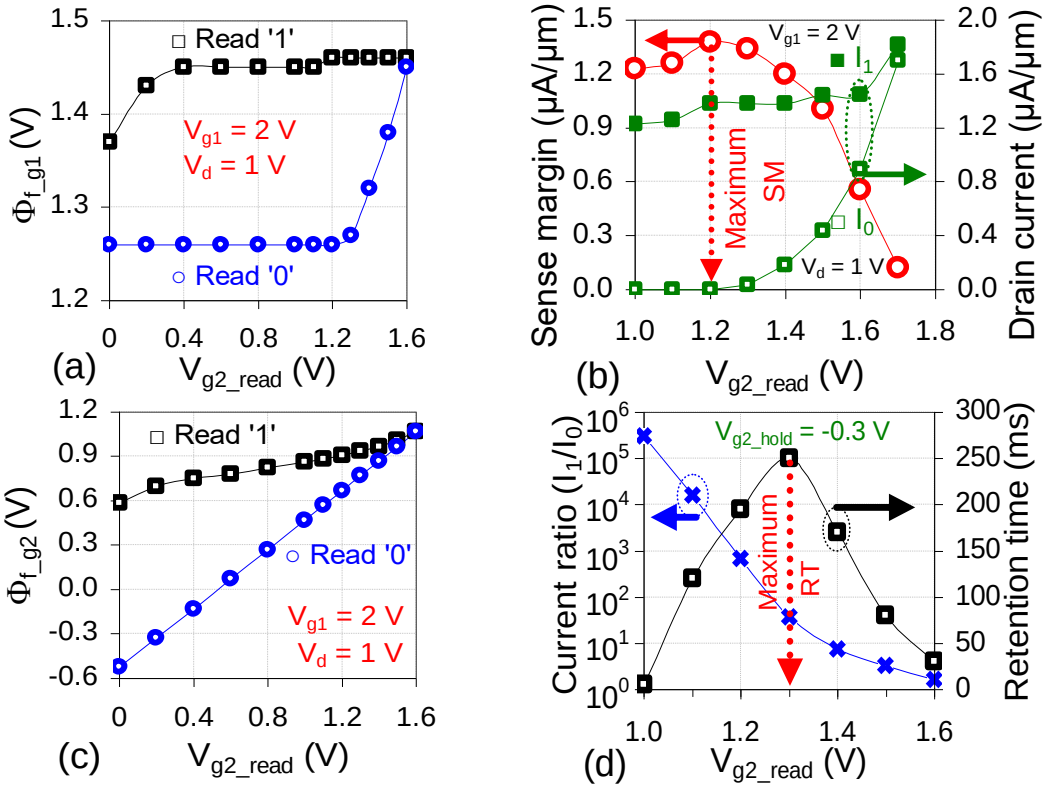

Fig. 5.3 Variation in (a) surface potential of G1 (Φ_{f_g1}) (b) sense margin, (c) surface potential of G2 (Φ_{f_g2}) (d) retention time and current ratio, with the bias applied at G2 (V_{g2}) during read operation, with V_{g1}= 2 V and V_d = 1 V. Bias applied at G2 during hold is -0.3 V.

5.3.2 Significance of read biases

As discussed before, the bias applied at G2 influences the energy barrier between G1 and G2 that affects the hole concentration under G1, and thus, impacts the electrostatic potential of the semiconductor film under G1. This is evident from Fig. 5.3(a) that shows variation in surface potential under G1 with bias at G2. For a positive lower bias at G2 (< 0.4 V) a lower surface potential for read '1' is observed which is almost constant for a wider range of (0.4 - 1.3 V) for both the states ('1' and '0') and increases significantly for read '0' for V_{g2} > 1.3 V. Similar trend is observed in Fig. 5.3(b) with the read currents. As observed in Fig. 5.3(c) the surface potential under G2 during read '0' is more sensitive to bias applied at G2 than read '1'. State '0' is depleted of holes during erase operation and thus, a lower positive bias applied at G2 during read from 2 V (for write '0') results into hole generation, which increases the surface potential of G2. For state '1', the accumulated excess holes already maintains the film under G2 at a higher potential after write '1'. Further, a positive potential applied at G2 lowers the

energy barrier between G1 and G2, thereby reducing the concentration of excess charge carriers contributing the positive potential. Therefore, a linear change in surface potential under G2 is not observed for read '1'. Thus, a higher increment in surface potential for read '0' compared to read '1', influencing the read currents, results into decrease in *CR* with increasing bias at G2 (Fig. 5.3 (d)). The analysis shows a programming window between 0.4 V - 1.6 V for *SM* > 500 nA. However, *RT* > 64 ms is achievable for a bias range of 1.1 - 1.5 V (Fig. 5.3(d)). It is also apparent from the figure, for bias > 1.2 V, a *CR* < 10^2 is attained, and thus, it eliminates the choice of bias between 1.3 V- 1.5 V applied at G2. Based on the investigation, an optimized bias of 1.2 V at G2 during read has been selected that shows maximum *SM* of ~1.4 µA/µm with *CR* > 10^2 and *RT* of ~ 200 ms for V_{g1} = 2 V and V_d = 1 V at 85 °C.

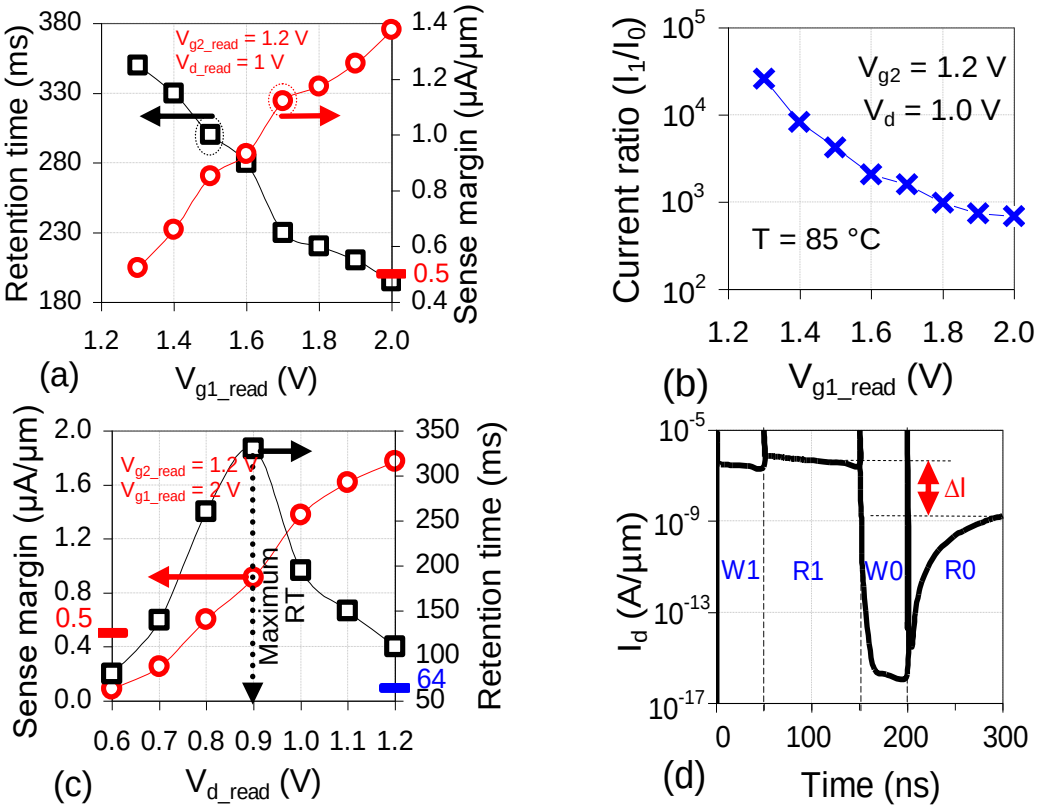

Fig. 5.4 Dependence of (a) *SM*, *RT* and (b) *CR* on the bias applied at G1 (V_{g1}) during read operation, for V_{g2} = 1.2 V and V_d = 1 V. (c) Variation in *SM* and *RT* with drain bias during read operation with V_{g2} = 1.2 V and V_{g1} = 2 V. (d) Transient current for the optimized bias set: V_{g2} = 1.2 V and V_{g1} = 1.5 V, V_d = 0.9 V. Δ*I* indicates the *SM*.

Figs. 5.3(b) and (d) also, highlights device capability to operate with a *CR* of ~10^4, *SM* of ~ 900 nA/μm and *RT* > 64 ms with V_{g2} = 1.1 V. This is an appreciable improvement over the previous TFET based DRAM [24-25], misaligned DG, twin gate, with L_{g1} = 400 nm and L_{g2} = 200 nm. The operation is further analysed through the variation in bias at G1 (Fig. 5.4(a) and (b)) and drain (Fig. 5.4(c)) that illustrates a gradual variation in DRAM characteristics with an improved *CR* and *RT* for lower bias at G1. Additionally, with V_{g1} = 2 V and V_{g2} = 1.2 V, it is observed that the DRAM operation is feasible at a drain bias of 0.8 V with *SM* > 500 nA, *CR* >10^2 and *RT* > 64 ms. Along with focus on improved *SM*, the biases are opted to obtain a high *RT* as well as *CR*. Therefore, further analysis is performed with V_{g1_read} = 1.5 V, V_{g2_read} = 1.2 V and V_{d_read} = 0.9 V that shows a *SM* = 575 nA, *RT* = 430 ms and *CR* > 10^3 for L_{g1} = 400 nm, L_{g2} = 200 nm and L_{gap} = 50 nm at 85 °C (Fig. 5.4(d)).

5.4 Comparative analysis with twin gate architecture

The benefit of integrating a bottom gate to the previous reported twin gate architecture can be analysed through potential profile of the two architectures, with (planar tri-gate) and without the back gate (twin gate). The comparison presents the physical insights that distinguish the performance metrics.

5.4.1 Physical insights and operating principle

As illustrated in Fig. 5.5(a) the presence of back gate raises the potential by ~160 mV at zero bias condition. Fig. 5.5(b) shows a higher steady state drain current for planar tri-gate as compared to twin gate, contributed by the two channels, at top and bottom surface. Also, a higher potential at 3 nm below the front surface in planar tri-gate TFET reflects on the impact of gate coupling and higher tunneling. At front surface, both the architectures will show same tunneling rate. Also, at back surface and sub-front surface a higher tunneling probability is observed for planar tri-gate, as demonstrated in Figs. 5.5(c) and 5.5(d). Fig. 5.5(c) highlights increased tunneling window ($\Delta\Phi$) and decreased tunneling barrier (λ) at surface below the front interface for planar tri-gate. Therefore, an increased gate control over the tunneling junction shows higher tunneling probability and an increased tunneling area that result into improved current drive. Before evaluating the

impact of potential and band modulation on the absolute values of drain current, physical insights into device operation as dynamic memory is essential.

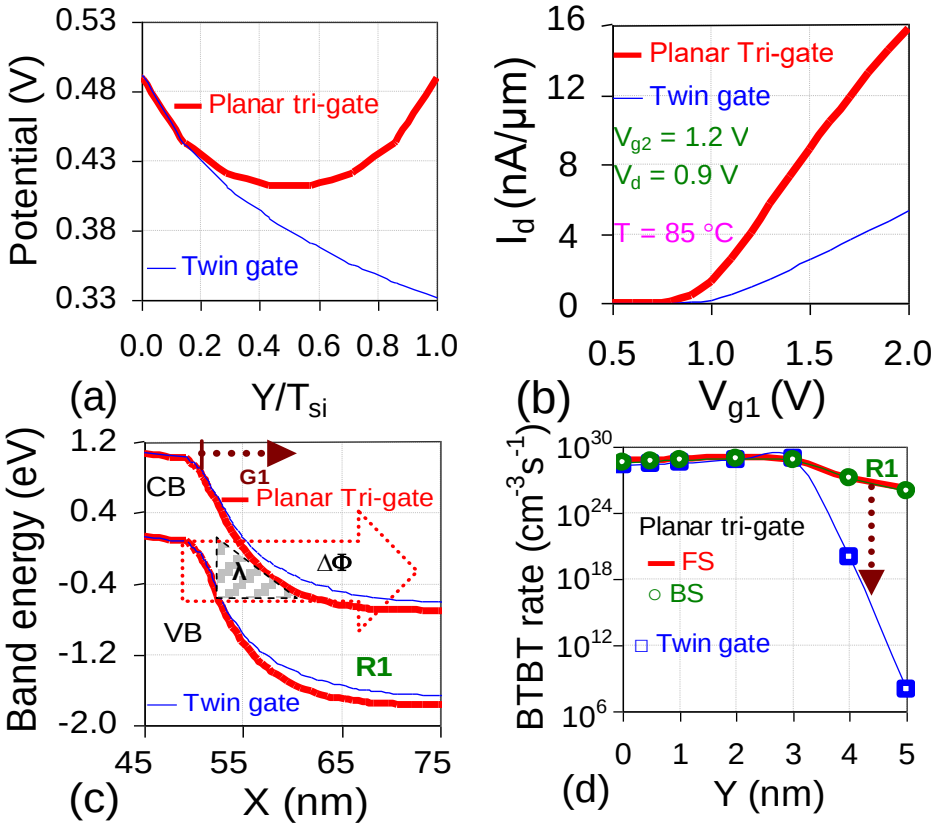

Fig. 5.5 Comparison of a planar tri-gate TFET with twin gate architecture, illustrating their (a) electrostatic potential along Y at $X = L_{g1}/2$, and (b) drain current characteristics with gate voltage (G1) with V_{g2} = 1.2 V and V_d = 0.9 V at 85 °C, (c) band energy during R1 along X at a depth of 6 nm below the front surface, and (d) BTBT rate along Y at the source/G1 tunneling junction.

Fig. 5.6 illustrates the variation in hole concentration in the storage region (under G2) as a function of time with each subsequent operation. The variation in hole concentration is reflective for read '1' operation, which is due to variation in potential under G1 that regulates the barrier between G1 and G2. A negative bias at W1 raises the barrier under G2 region, but positive bias from write '1' to read '1' (R1) lowers the barrier (Fig. 5.7(a)) that decreases the hole concentration underneath G2. A higher potential under the silicon film at G2 region due to excess positive charges, result into a lower barrier of ~0.3 eV for electrons (Fig. 5.7(a)) in read '1' as compared to read '0' (Fig. 5.7(b)). The reduced barrier allows more electron diffusion from G1, which are further drifted towards drain.

As shown in Fig. 5.5(a) for the twin gate, the lowest potential under G1 is observable at the back surface with a value ~330 mV at no bias condition, while that for planar tri-gate at the centre of film (~400 mV at zero bias). The planar tri-gate exhibits a higher potential in the film under G1, maintaining a raised barrier between G1 and G2, and thus, permits an increased sustenance of holes during read in the potential well. The barrier difference affects the read mechanism where the bias applied at G1 exhibit a higher potential at film under G1 for planar tri-gate (due to incorporation of dual gates) compared to twin gate. This is evident from Fig. 5.6 with hole concentration of ~ 2×10^{19} cm^{-3} for planar tri-gate and ~ 6×10^{18} cm^{-3} for twin gate. The variation in the presence of excess holes in the potential well for planar tri-gate and twin gate is reflective in the current for read '1' and an increased tunneling area that shows higher currents.

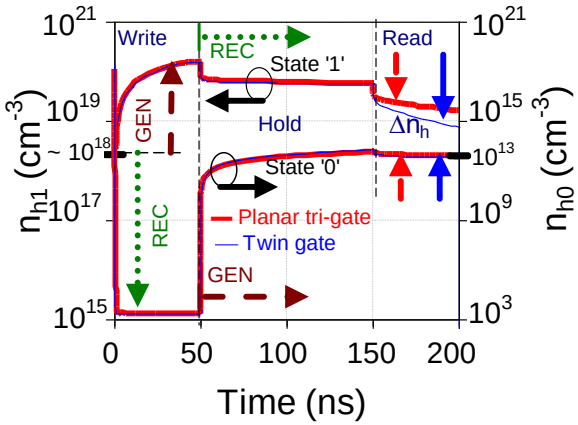

Fig. 5.6 Variation of hole concentration (n_h) in the storage region for each operation in twin gate and planar tri-gate TFET with probe positioned at mid of G2, 1 nm below front surface.

Fig. 5.7(c) shows a high *SM* of ~200 nA which is a significant improvement in comparison to twin gate (Fig. 5.7(d)) with a sense margin of ~ 20 nA for $L_{g1} = L_{g2}$ = 100 nm and L_{gap} = 50 nm. For same set of bias values (Table 5.2) the *RT* observed is lower by ~ ×1.4, but a remarkable improvement in sense margin (~ ×10) is achieved. Thus, the planar tri-gate show enhanced sense margin with similar retention characteristics and also, maintains the *CR*. As discussed earlier, the incorporation of an additional bottom gate to the twin gate, (a) creates a second conduction channel, and (b) results into enhanced tunneling due to coupling with the top gate [20-22]. Therefore, the symmetric double gate

positioned near source at a partial region of intrinsic film to regulate the read mechanism, which results into a higher tunneling probability.

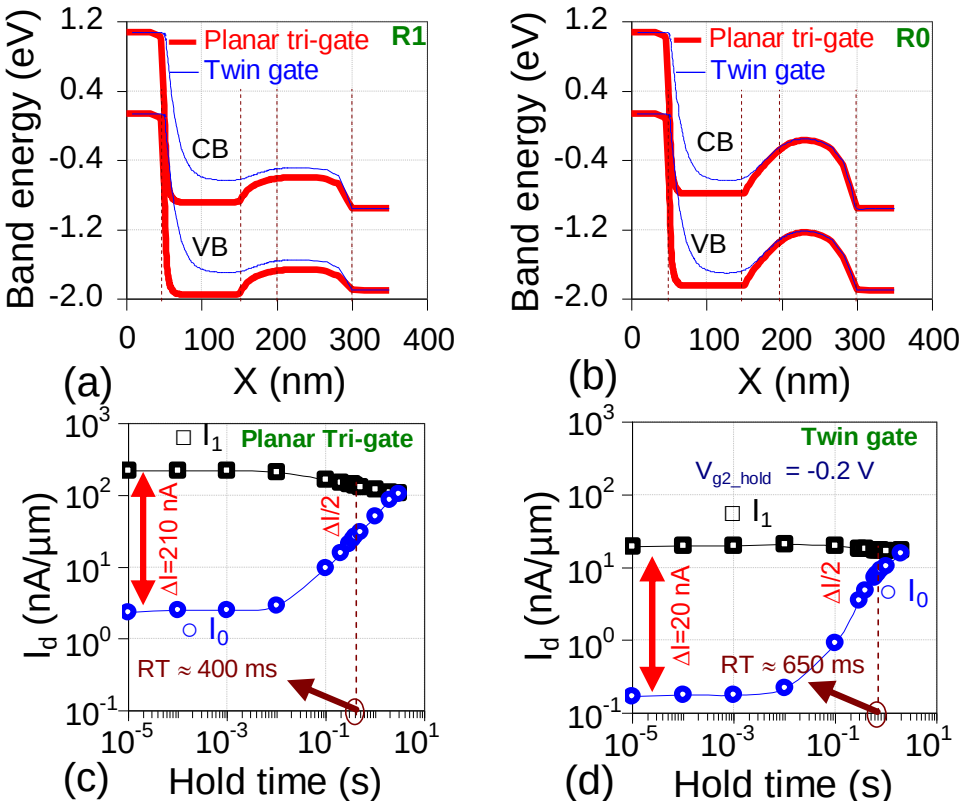

Fig. 5.7 Energy band profile for (a) R1 and (b) R0 operation for planar tri-gate and twin gate, evaluated at a cross-section 1 nm above the back surface ($Y = T_{si}$). Dependence of read currents for state '1' (I_1) and state '0' (I_0) on hold time for (c) planar tri-gate TFET and (d) twin gate TFET.

Table 5.2 Programming Scheme for Planar Tri-gate TFET

Operation	***V_d***	***V_{g1}***	***V_{g2}***	***Time***
Write 1	0 V	0 V	-2.0 V	50 ns
Write 0	0 V	0 V	2.0 V	50 ns
Read	0.9 V	1.5 V	1.2 V	50 ns
Hold	0 V	0 V	-0.2 V	-

5.4.2 Scalability

5.4.2.1 Evaluation through composite metrics

A high *SM* and *RT* paves a way toward better scalability at lower bias values, indicating applicability as low power dynamic memory. As discussed in section

2.4.5, trade-offs exist between *SM* and *RT* as well as between *SM* and *CR*. Focusing on each metric, two composite metrics (*M1* and *M2*) namely, product of (i) Sense Margin (*SM*) and Retention Time (*RT*) i.e. $M1 = SM \times RT$, and (ii) Sense Margin and Current Ratio (CR) i.e. $M2 = SM \times CR$ are introduced in this section. This is evident from Fig. 5.8(a) where improved performance is achieved for planar tri-gate TFET compared to twin gate and the same is demonstrated through *M1*, factor governing scalability, and *M2*, parameter which determines the read sensitivity. Higher values of composite metrics reflect on the suppression of trade-off between various constraints for an optimal DRAM design.

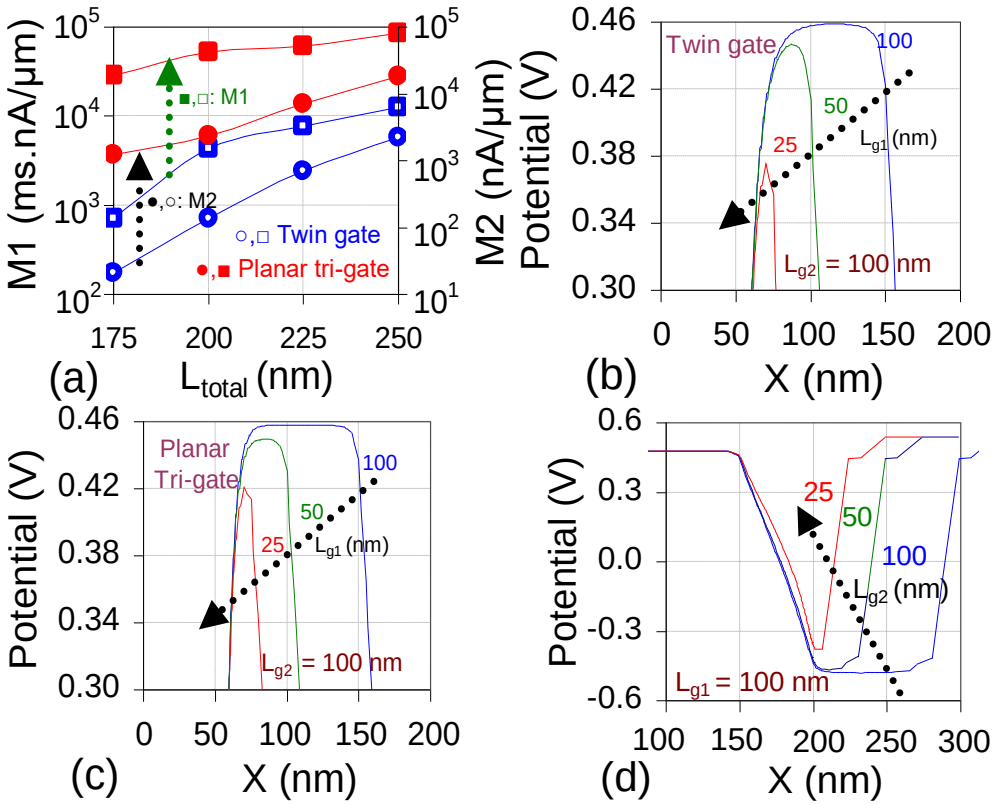

Fig. 5.8 (a) Variation in DRAM metrics (*M1* and *M2*) with variation in total length ($L_{g1} + L_{g2} + L_{gap}$) where *M1* indicates $RT \times SM$ and *M2* is $CR \times SM$, L_{gap} is fixed at 50 nm. Dependence of electrostatic potential on length of G1 (L_{g1}) for (b) twin gate and (c) planar tri-gate along *X* at a cross-section of 2 nm below the front surface ($Y = 0$). (d) Variation in electrostatic potential with length of G2 (L_{g2}) for planar tri-gate along *X* at 1 nm below the front surface ($Y = 0$).

Enhanced values of *M1* and *M2* at shorter gate lengths signify enhanced scalability. Figure shows that *M1* and *M2* decrease with downscaling gate lengths.

The performance is better for planar tri-gate with higher values of both the metrics as compared to twin gate at lower gate lengths. Thus, planar tri-gate is more suitable at reduced gate lengths.The decrease in performance metrics (*M1* and *M2*) with gate length scaling can be analysed through Figs. 5.8(b)-(d). The reduction in the length of G2 reduces the storage capacitance (Fig. 5.8(d)) that decreases the retention time. Figs. 5.8(b) and (c) show a reduced potential with lower L_{g1}, for both, planar tri-gate and twin gate, thereby lowering the barrier for electrons to drift towards G2. This results into significant increase in current for state '0', thereby reducing the *SM*. However, a prominent change for L_{g1} = 25 nm is observed between the architectures with a difference of ~50 mV in the potential at zero bias condition. Thus, planar tri-gate TFET establishes as an improved architecture with reduced short channel effects due to improved electrostatic control over the region aligned, demonstrating enhanced sense margin and scalability.

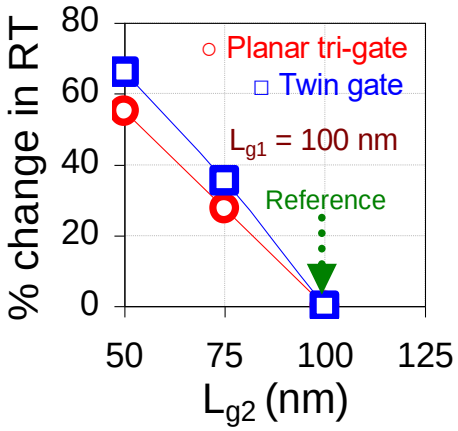

Fig. 5.9 Percentage change in *RT* for planar tri-gate and twin gate architecture with varying length of G2 for L_{g1} = 100 nm and L_{gap} = 50 nm.

The scaling of G2 region reduces the storage capacitance that decreases the retention capability. The architectures, twin gate as well as planar tri-gate utilize the region under G2 to create a dedicated volume for charge storage. Therefore, the trend of retention and the degradation in both the architectures with gate length scaling is similar (Fig. 5.9). The same trend is followed with L_{g1} = 75 nm. However, for shorter gate lengths (L_{g1} < 75 nm) twin gate fails to perform as dynamic memory. Additionally, planar tri-gate benefits with a higher sense margin and better scalability. The twin gate architecture can show retention for

L_{g1} = 75 nm, while planar tri-gate till L_{g1} = 25 nm with a retention > 64 ms for L_{g2} = 50 nm. This shows advancement in TFET based DRAM with the utilization of planar tri-gate.

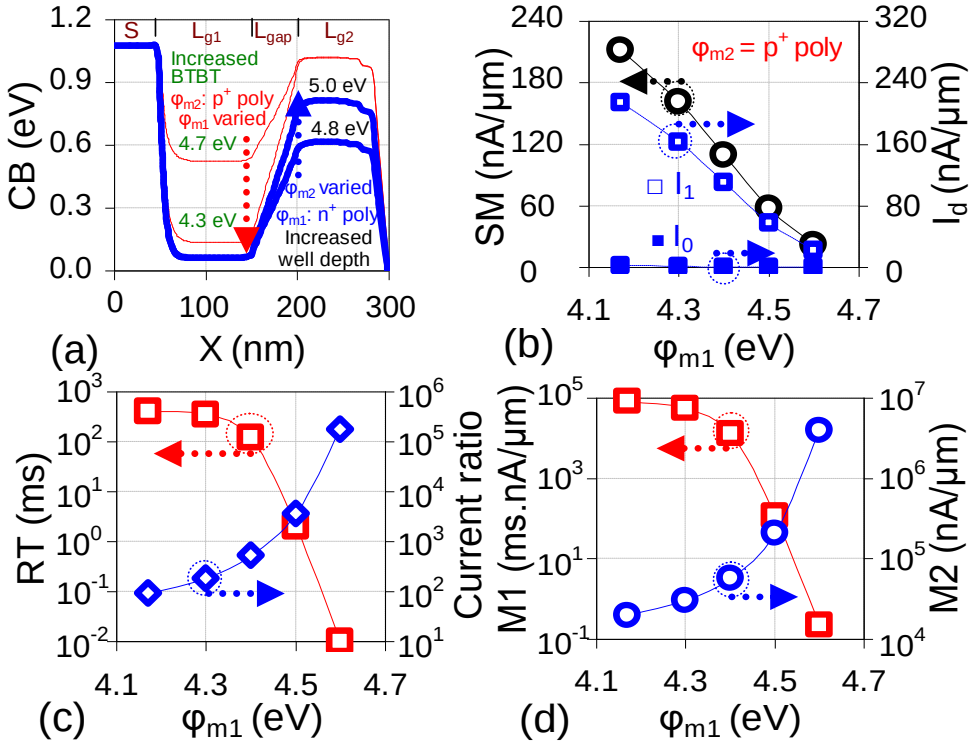

Fig. 5.10 (a) Dependence of conduction band on the workfunction of G1 (φ_{m1}) and G2 (φ_{m2}). Variation in various DRAM metrics, (b) *SM*, I_1, I_0, (c) *RT* and *CR*, and (d) *M1* and *M2* with φ_{m1} for φ_{m2} = 5.25 eV.

5.5 Factors affecting performance of planar tri-gate DRAM

DRAM performance metrics are influenced by recombination (thermal recombination and diffusion) and generation of holes (thermal generation and BTBT) that is regulated by device dimensions, architecture, bias applied and temperature. For an optimized set of bias values (Table 5.2) at 85 °C with gate lengths of 100 nm, the work shows the impact of device parameters that affects the gate controllability. Thus, use of optimal values maintain the energy barriers for respective states, induced by gates showing improved read sensitivity and *RT*.

5.5.1 Impact of workfunction

As discussed before in chapter 2, workfunction engineering is essential to create a potential well. The gate with a lower workfunction value results into a high

potential under the gate while the higher value into a low potential region. A lower workfunction value of G1 (φ_{m1}) increases the maximum electric field at the tunneling junction (p^+/i) that reduces the tunneling width [21]. This improves the tunneling between source and G1 (Fig. 5.10(a)) and thus, enhances read currents. Fig. 5.10(b) demonstrates increase in currents for both the states with lower φ_{m1} that improves the *SM*, but simultaneously reduces the current ratio (Fig. 5.10(c)). However, higher φ_{m1} reduces barrier between G1 and G2 that enhance hole recombination, thereby degrading state '1' and thus, the *RT*. Thus, decrease in *SM* and *RT* show reduction in *M1* with higher φ_{m1}. Although, *M2* increases with increase in φ_{m1}, value > 10^4 ($\varphi_{m1} \geq 4.15$ eV) along with high *M1* shows utilization of lower workfunction values for G1. The results demonstrate $\varphi_{m1} \leq 4.4$ eV as a preferable choice to achieve high DRAM metrics. Thus, the design proposed utilizes an *n*-type gate ($\varphi_{m1} = 4.15$ eV) for tunneling based read operation.

The impact of workfunction value of G2 (φ_{m2}) on DRAM performance metrics is illustrated in Fig. 5.11. A lower φ_{m2} reduces the energy barrier between G1 and G2. This enhances the current for both the states. However, state '0' being depleted of charges is influenced more. It results into a decreased *SM* and *CR* with lower workfunction values (Fig. 5.11(a)). Moreover, the reduced barrier between G2 and G1/drain increases recombination, thereby degrading state '1'. A higher φ_{m2} results into a deeper potential well (Fig. 5.10(a)) that retain holes for a longer duration, maintaining state '1', while it degrades state '0' due to increased tunneling between G2 and drain/G1. Thus, workfunction regulates the barrier heights that directly affect the generation/recombination mechanism. Fig. 5.11(b) demonstrates the potential change between Write and Hold operation for state '1' and '0' that indicate a higher variation in state '1' for lower φ_{m2} due to higher recombination. A higher potential change in state '0' is observed for higher φ_{m2} due to increased generation. *RT* along with *CR* are shown in Fig. 5.11(c). A *RT* > 64 ms is attained for $\varphi_{m2} \geq 4.9$ eV. However, a lower *CR* and *SM*, reflective in metric *M2* show $\varphi_{m2} \geq 5.1$ eV as the optimal value (Fig. 5.11(d)).

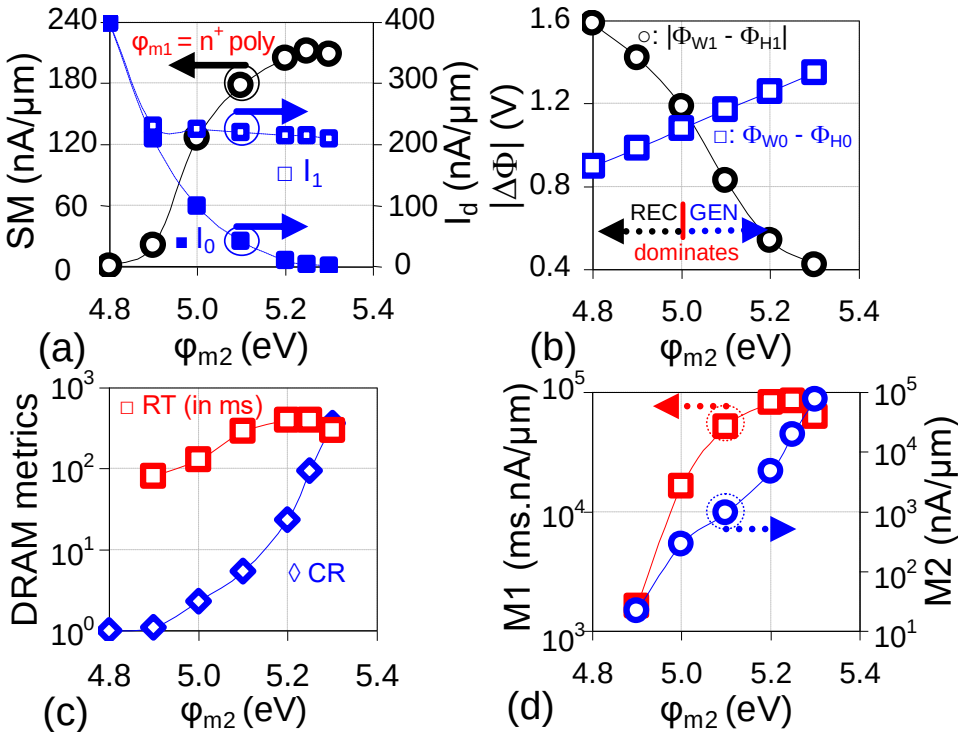

Fig. 5.11 Dependence of (a) *SM*, I_1 and I_0 with φ_{m2}. (b) Variation in potential ($\Delta\Phi$) from write to hold operation for state '1' and '0', respectively. Variation in DRAM metrics, (c) *RT* and *CR*, and (d) *M1* and *M2* with φ_{m2} for φ_{m1} = 4.15 eV.

5.5.2 Impact of silicon film thickness

Fig. 5.12(a) illustrates increase in *SM* with reduction in film thickness. Fig. 5.12(b) shows increase in maximum electric field at the source gate tunneling junction with decrease in film thickness that results into higher read currents. Decrease in silicon film thickness improves the electrostatic control of the gate over the channel that results in more band-bending at source/channel junction [20-22]. This results into improved gate coupling, reflective in tunneling rate at subsurface region (5 nm below front surface) in Fig. 5.12(c). The increase in maximum electric field at the source gate tunneling junction with decrease in film thickness results into higher read currents, and thus, higher *SM* but at the expense of a lower *CR* (Fig. 5.12(c)). The decrease in film thickness also increases the hole generation in the storage region during Hold '0'. This is due to increased BTBT at G2 (virtual *p* region) and drain (n^+ region) that degrades the retention characteristics (Fig. 5.13(a)). The analysis has been performed for $T_{si} \geq 10$ nm and therefore, no quantum effects have been considered [2]. A thicker film reduces gate control over the channel region, thereby increasing short channel effects that

increase the hole diffusion and thus, the recombination, which further decreases retention time [2]. However, for evaluation $T_{si} \leq 30$ nm is considered that shows $RT > 64$ ms. Results highlight an optimal range of T_{si} between 20-30 nm for high DRAM performance metrics (Fig. 5.13(b)).

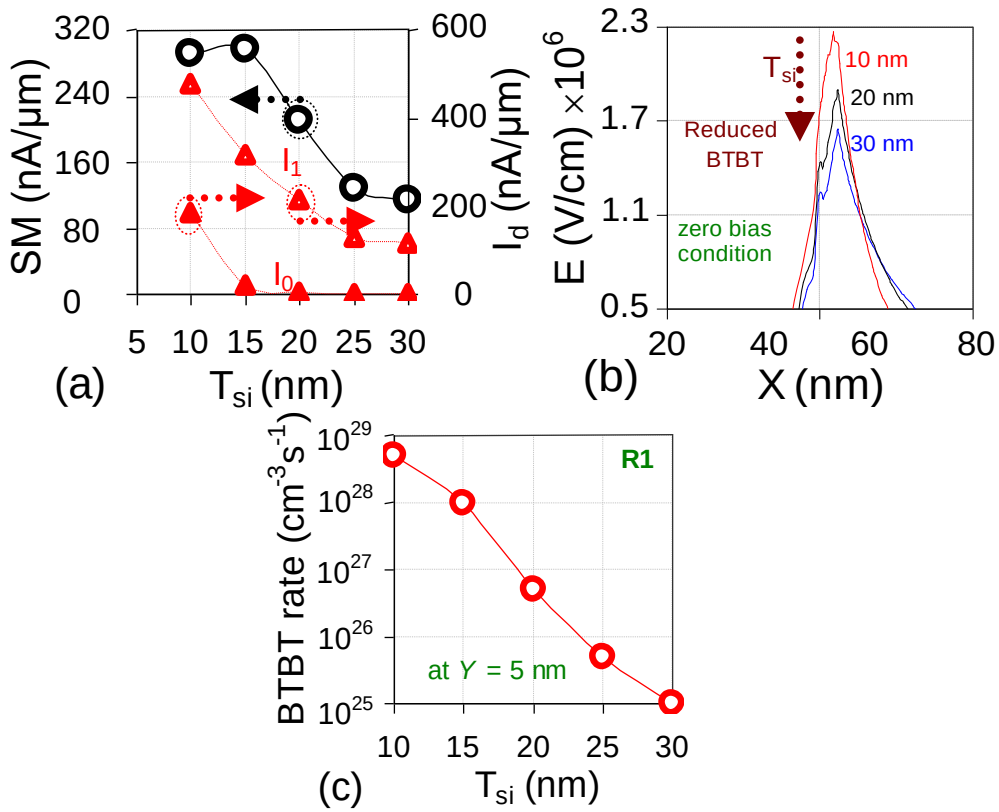

Fig. 5.12 Dependence of (a) *SM*, I_1 and I_0, (b) maximum electric field at the tunneling junction, 1 nm below the front surface along *X*, (c) BTBT rate along the channel direction (*X*) at 5 nm below the front surface, on T_{si}.

5.5.3 Impact of oxide thickness

The oxide thickness regulates the transverse electric field, where a thicker oxide shows reduced gate controllability, showing requirement of higher bias or time for achieving specific range of DRAM characteristics. The thicker oxide reduces gate capacitance responsible for creation and maintenance of potential well, and thus, leads to a shallower potential well. Secondly, it also reduces the current driving capability [34]. Write '1' mechanism is based on BTBT where the tunneling is regulated by the bias applied at G2. The reduced electric field with thicker oxide of G2 (T_{ox2}) requires longer write time. As demonstrated in Fig. 5.14(a) the write time for T_{ox2} = 6 nm increases to a few μsec. Although, DRAM performance

metrics in Fig. 5.14(b) show high values of *M1* and *M2* for T_{ox2} between 3 to 6 nm, a higher write time is required for thicker oxide.

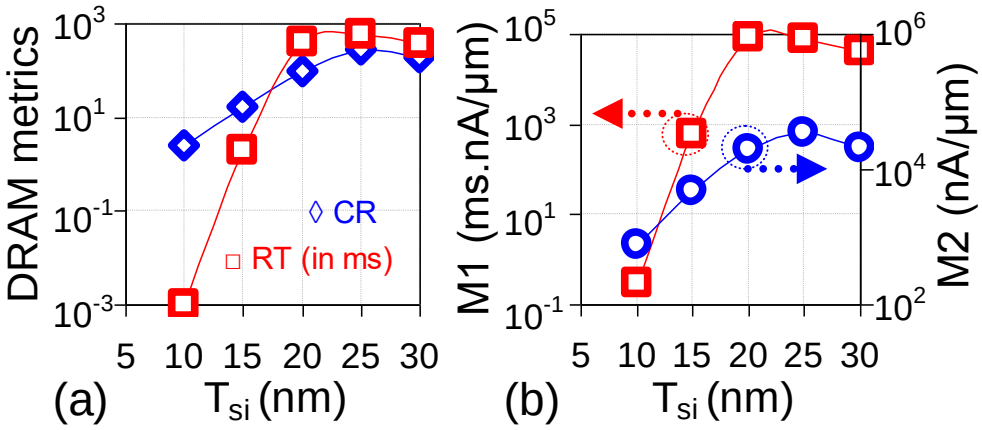

Fig. 5.13 Dependence of (a) *RT* and *CR*, and (b) *M1* and *M2*, on T_{si}

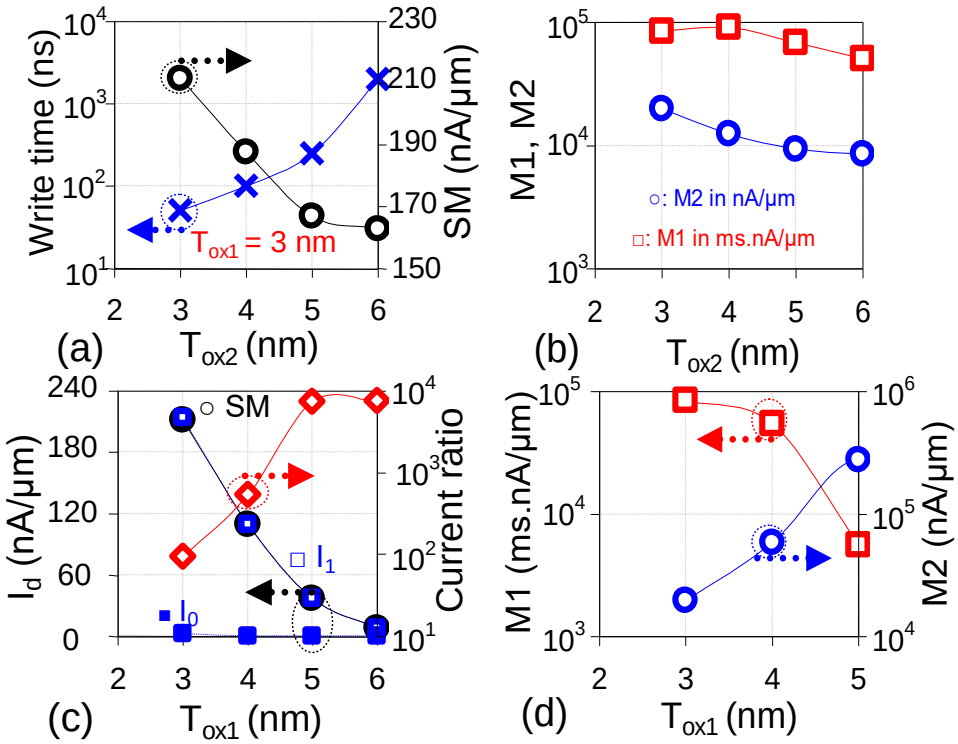

Fig. 5.14 Variation in (a) *SM* and write time, and (b) *M1* and *M2* with oxide thickness of G2 (T_{ox2}) for T_{ox1} = 3 nm (HfO_2). Dependence of (c) *SM*, I_1 and I_0, and (d) *M1* and *M2* on oxide thickness of G1 (T_{ox1}) for T_{ox2} = 3 nm (HfO_2).

G1 is responsible for read mechanism based on BTBT. An increased oxide thickness of G2 (T_{ox1}) reduces the gate capacitance, regulating the tunneling at source/channel junction, and hence, a decrease in *SM*. Device with T_{ox1} = 6 nm

shows *SM* < 20 nA and thus, has been eliminated for further evaluation (Fig. 5.14 (c)). *SM* can be improved with a higher bias for T_{ox1} = 6 nm. However, this will require high bias. The improved current ratio with increased oxide (T_{ox1}) shows an improved value of *M2*, but *M1* decreases (Figs. 5.14(c) and (d)). This is due to lower potential energy barrier between G1 and G2, showing higher recombination. Thus, based on the analysis 3-5 nm for T_{ox1} is acceptable and therefore, T_{ox1} = 3 nm is used in the analysis. Based on the analysis we have an optimal range of design parameters for an improved DRAM performance metrics (Table 5.3).

Table 5.3 Optimal design parameters for DRAM operation at 85 °C

Parameters	Optimal values	Selected values
φ_{m1}	≤ 4.4 eV	4.15 eV
φ_{m2}	≥ 5.1 eV	5.25 eV
T_{si}	20 - 30 nm	20 nm
T_{ox1}	3-5 nm	3 nm
T_{ox2}	3-6 nm	3 nm

5.5.4. Impact of temperature

The DRAM performance metrics is also, influenced by the temperature, as demonstrated in Fig. 5.15. Temperature impacts the generation and recombination that influences the *RT*, reducing it from 1.5 s at 27 °C to ~60 ms at 125 °C (Fig. 5.15(a)). TFET has weak dependency on temperature which is reflective in the read currents, and therefore, *SM* changes gradually (Fig. 5.15(b)). TFET is benefited with weak temperature dependency on the drain characteristics [21], that is reflective in *SM*, showing variation from ~180 nA at 27 °C to ~220 nA at 125 °C. The DRAM performance is degraded with temperature, which can be estimated through composite DRAM metrics, where *M1* and *M2* reduce by an order of two from 27 °C to 125 °C. Although, performance metrics degrade with temperature, an acceptable sense margin (~220 nA) and retention time (~60 ms) at 125 °C, indicates planar tri-gate TFET as a reliable dynamic memory. The proposed topology shows a high retention time of 1.5 s at room temperature, and

also, enhanced DRAM metrics, *M1* = ~10^5 ms.nA/µm and *M2* = ~10^4 nA/µm at 85 °C for L_{g1} = L_{g2} = 100 nm, which highlights the potential of device to operate efficiently and better than previous TFET based DRAM [24-26]

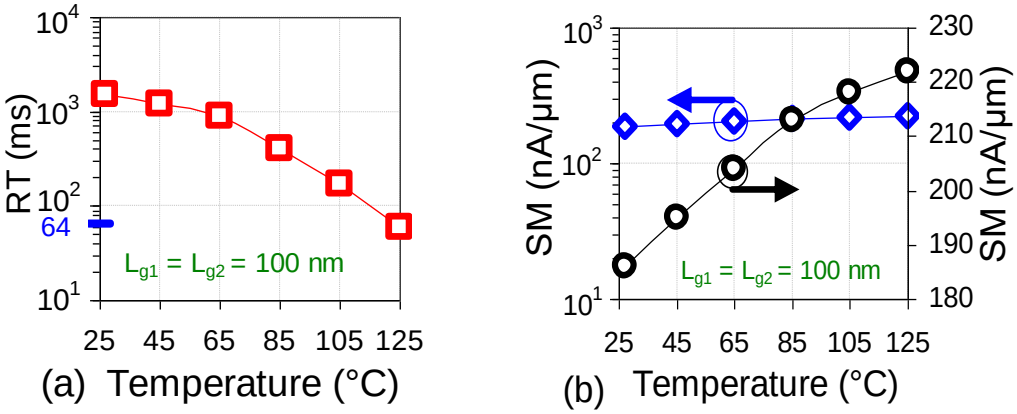

Fig. 5.15 Variation in (a) *RT* and (b) *SM* as a function of temperature.

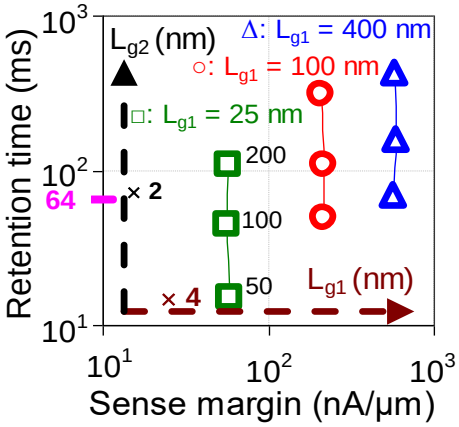

Fig. 5.16 Dependence of *RT* and *SM* on gate lengths (L_{g1} and L_{g2}) indicating minimum value of L_{g1} and L_{g2} for operation as DRAM with Gate-2 voltage of -0.3 V during hold.

5.5.5 Impact of gate lengths

In planar tri-gate TFET, the decrease in L_{g2} reduces the storage capacitance, thereby lowering *RT*, while reduction in L_{g1} lowers *SM* due to short channel effects. The same is observed in Figs. 5.16 and 5.17 with hold voltages of -0.3 V and -0.2 V, respectively. Also, figures highlight L_{g1} scalability down to 25 nm with *SM* of ~ 60 nA due to double gate (G1) architecture that has better electrostatic control over the film. As discussed before, the performance metrics (*M1* and *M2*) reduces with gate length scaling (Fig. 5.8(a)). Further analysis

highlights the device capability to scale L_{g2} down to 50 nm with $L_{g1,\ min} \cong L_{g2,\ min}/2$, depicting an acceptable read sensitivity with RT > 64 ms [23] at 85 °C (Fig. 5.17) which translates into $M1$ ~10^3 ms.nA/μm, $M2$ ~150 nA/μm. The scalability for twin gate architecture, demonstrated in chapter 4 can be approximated as $L_{g1,\ min} \cong 2L_{g2,\ min}$, and thus, the comparison shows an improved scalability of L_{g1} for planar tri-gate TFET. Thus, the minimum total length of planar tri-gate TFET based dynamic memory can be scaled down to 125 nm (L_{gap} = 50 nm) with high retention characteristics as well as improved SM compared to previous architectures [24-26], twin gate and misaligned DG TFET.

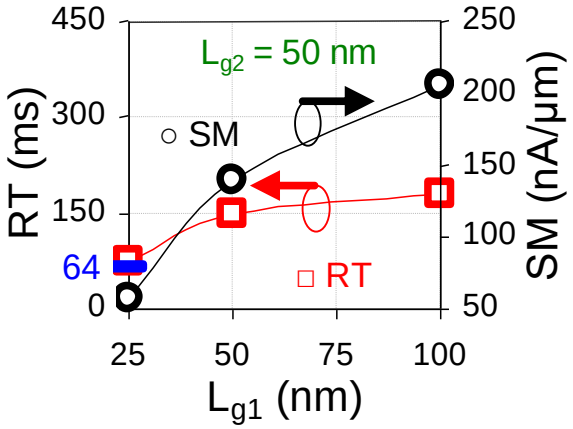

Fig. 5.17 Variation in SM and RT with length of G1 (L_{g1}) for L_{g2} = 50 nm with Gate-2 voltage of -0.3 V during hold.

The other significant advantage of using a planar tri-gate other then the dual gate for improving SM is, the well-defined G2 that dedicatedly serves as storage region, and thus, does not degrade retention characteristics. The retention trend observed in both the architectures (planar tri-gate and twin gate TFET) with scaling of G2 is similar. However, for shorter gate lengths (L_{g1} < 75 nm) twin gate fails to perform as dynamic memory. Additionally, planar tri-gate benefits with from a higher sense margin and better scalability. This shows advancement in TFET based DRAM with the utilization of planar tri-gate.

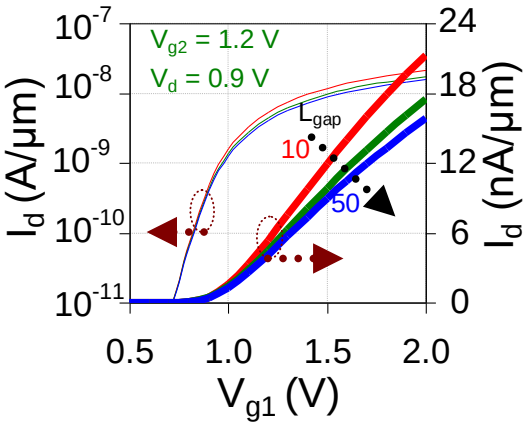

Fig. 5.18 Variation in drain characteristics with gate voltage for varying L_{gap} with $L_{g1} = L_{g2} = 100$ nm.

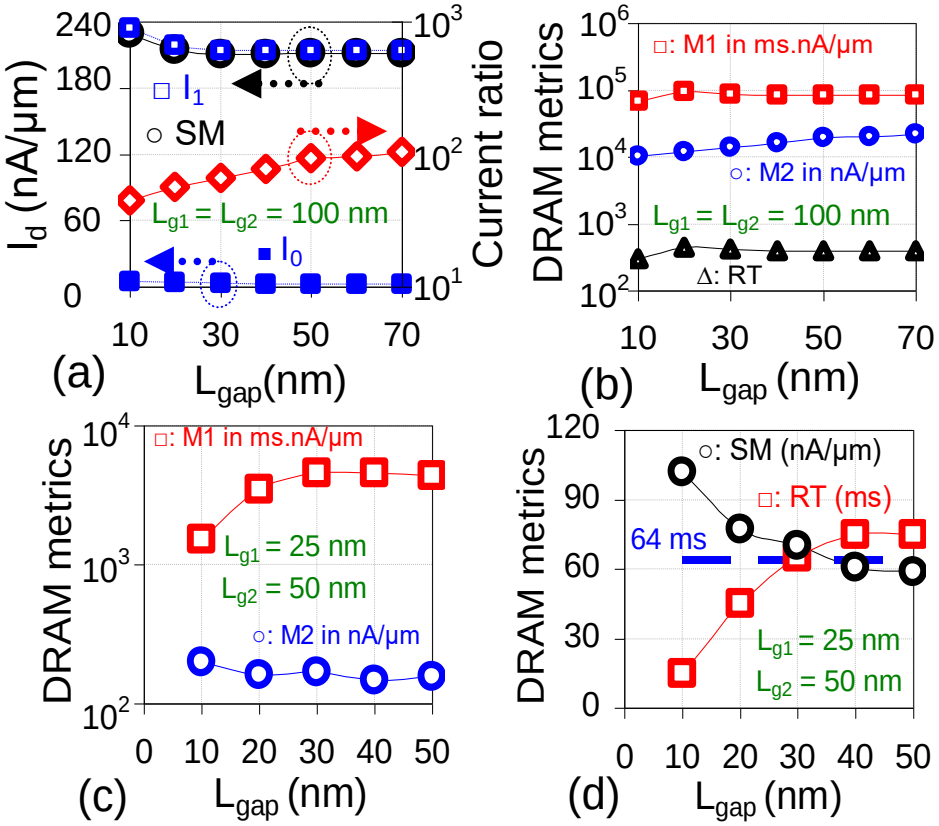

Fig. 5.19 Variation in (a) *SM* and read currents, and (b) *M1*, *M2* and *RT* with lateral space between the gates (L_{gap}) for $L_{g1} = L_{g2} = 100$ nm. Dependence of (c) M_1 and M_2, and (d) *SM* and *RT* on L_{gap} for minimum gate lengths to operate as DRAM ($L_{g1} = 25$ nm and $L_{g2} = 50$ nm).

The lateral spacing incorporated between the gates increases the effective gate lengths. More significantly, it reduces the tunneling of electrons between G1 and G2. These features result into improved retention characteristics. The increased

effective lengths due to incorporation of L_{gap} reduce the drain current. However, tunneling based current doesn't show a strong dependence on gate length (Fig. 5.18) [20]. Thus, read currents are almost maintained along with DRAM metrics, *SM*, *RT*, *CR*, *M1*, *M2* evident in Figs. 5.19(a)-(b) for L_{g1} = L_{g2} = 100 nm. For lower gate lengths, variation in L_{gap} is more critical, as increasing the effective gate lengths show improved retention, but also, reduced sense margin. Therefore, $L_{gap} \geq$ 30 nm is preferable to maintain the DRAM performance metrics (Figs. 5.19(c)-(d)). Although Fig. 5.17 demonstrates, *SM* is maintained for L_{g1} = 25 nm with $L_{g2} \geq$ 50 nm, design optimization is required to improve the retention characteristics for the same geometry. This can be achieved through the use of underlap between Gate-2 and drain, as discussed in the following section.

5.5.6 Impact of underlap/overlap

RT can further be improved by incorporating an underlap region (L_{un}) between drain and G2, adopted in conventional MOS devices [9]. However, inclusion of L_{un} results in reduced tunneling during write '1' that lowers *SM* due to lower hole accumulation, affecting read '1' appreciably. Thus, a different approach (Method B) based on forward biasing the p^+-*i* region with a source bias of 1 V along with V_{g2} = -0.5 V is adopted. It increases the charge storage during write '1' along with an underlap, sufficient to differentiate with state '0'. Therefore, *SM* degradation due to write mechanism using tunneling (Method A) can be recovered by Write '1' through Method B, as illustrated in Fig. 5.20(a). *SM*, although marginally degraded due to the increased resistance contributed by L_{un}, maintains acceptable values. The most distinguishing result of the tri-gate architecture is the remarkable improvement is observed in *RT* (× ~7) with the adopted methodology, while maintaining nearly same *SM*. The optimum choice of underlap is essential for retention which is reflective in Fig. 5.20(b) and thus, an underlap of 15 nm at drain/channel junction is considered for further investigation. The read currents can also be recovered when an underlap is incorporated, by increasing the write time. Fig. 5.20(c) demonstrates the same, where the write time is increased to 30 μs to attain the same sense margin as shown in Fig 5.20(a) for read bias V_{g1} = 1.5 V. Thus, design optimization with use of L_{un} and a different write mechanism

(Method B) in planar tri-gate topology reduces the trade-off between *SM* and *RT*, which is one of the key challenges for DRAM performance.

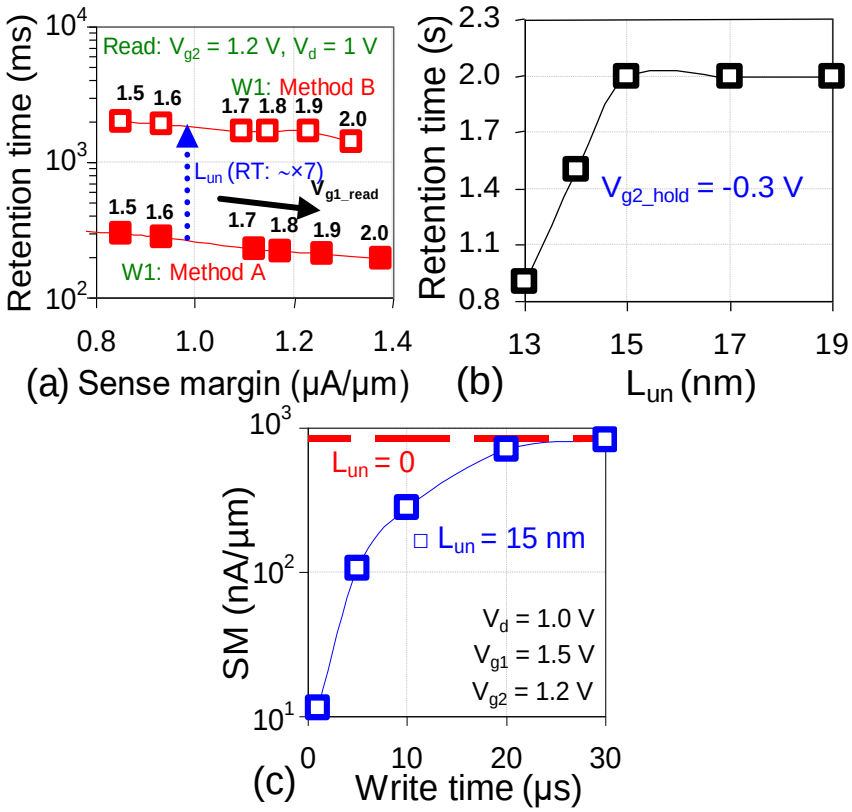

Fig. 5.20 (a) Variation in *RT* with an underlap (L_{un}) between second gate and drain as a function of V_{g1} with same *SM* achieved through two different Write mechanisms. Method A: V_{g2} = -2 V, Method B: V_{g2} = -0.5 V and V_s = 1 V. (b) Dependence of *RT* on the L_{un} at drain/channel interface. (c) Variation in *SM* with write time for L_{un} (15 nm) at drain/channel junction with write bias of -2 V at G2.

The underlap/overlap at source/channel region also influences the currents [35-38]. As demonstrated in previous works [35-38], overlap/underlap lengths require optimization for improved performance. The current in a TFET is indicative of the DRAM current in different states, and thus, influences the sense margin and current ratio of DRAM. Fig. 5.21(a) demonstrates that although overlap reduces both the currents, significant impact on read '0' influences current ratio more, while sense margin is almost maintained >10^2 nA/μm. An underlap at G1/source junction degrades both the read currents, and hence, the sense margin, but current ratio increases. The degradation in device performance can however, be improved by appropriate choice of bias values. For the case with underlap, increasing the

bias will enhance read '1', thereby increasing the sense margin (Fig. 5.21(b)). Thus, it is very crucial to optimize bias values with the device architecture adopted for DRAM. The advantage we achieve from underlap at source is a high *CR*, but *SM* is a compromise. Thus, the analysis highlights variation in device performance with gate to source/drain extensions, and the significance of bias optimization. In our work, we have a nominal overlap of 2 nm at source/channel region.

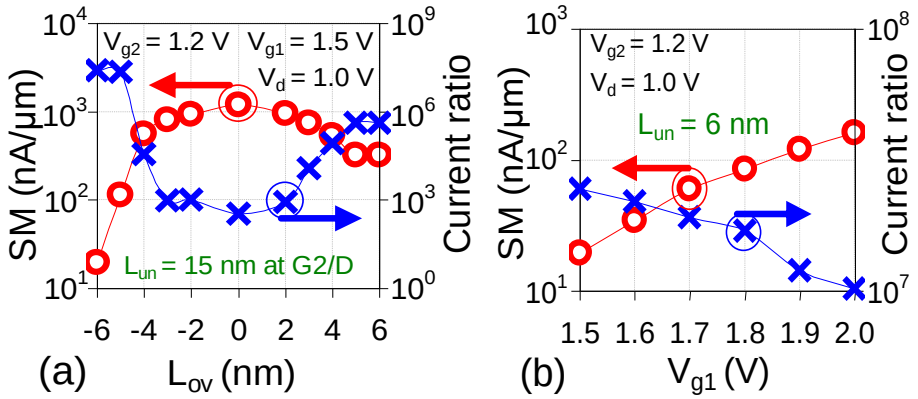

Fig. 5.21 (a) Variation in *SM* and *CR* with (a) source to channel underlap (indicated by negative values) and overlap, (b) gate voltage (G1) for source to channel underlap of 6 nm.

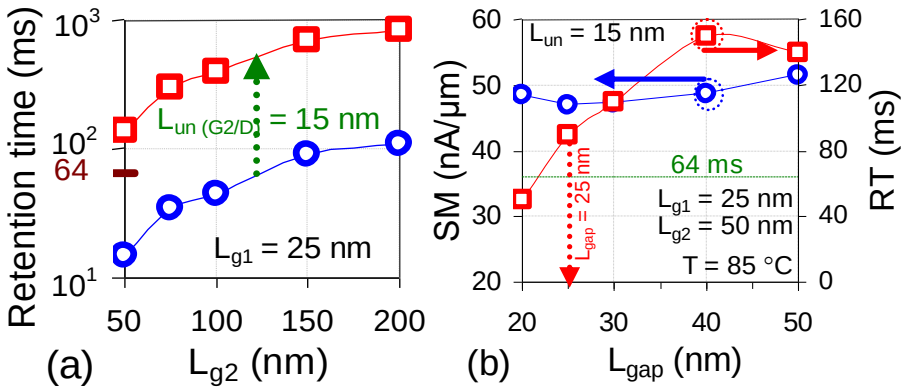

Fig. 5.22 (a) Variation in *RT* with L_{g2} for cases, with and without L_{un} (15 nm) for L_{g1} = 25 nm. (b) Dependence of *SM* and *RT* on the lateral spacing (L_{gap}) between the gates, highlighting scalability of L_{gap} up to 25 nm and total length scaling down to 115 nm with *RT* > 64 ms.

The optimization techniques studied for longer gate lengths are further analyzed at shorter gate lengths. Fig. 5.22(a) demonstrates the impact of the structural

modification through the incorporation of an underlap between G2 and drain for L_{g1} = 25 nm. This permits G2 scaling down to 50 nm (RT > 64 ms) with Method B as write mechanism that shows similar SM with an improved RT. The previous results utilized a L_{gap} = 50 nm. However, L_{gap} can be scaled down to 25 nm with RT > 64 ms for L_{g1} = 25 nm and L_{g2} = 50 nm, without degrading SM (Fig. 5.22(b)). L_{gap} not only reduces the state '0' degradation, associated with tunneling between G1 and G2, but also, increases effective lengths. Therefore, reducing L_{gap} reduces the storage region and also, increases the tunneling, thereby reducing RT. However, for L_{g1} = 25 nm and L_{g2} = 50 nm, the scalability of L_{gap} down to 25 nm with a 15 nm underlap shows device potential to scale the total length ($L_{g1} + L_{g2} + L_{gap} + L_{un}$) down to 115 nm.

5.5.7 Impact of interface charges

The device performance can also be affected by interface charges (Q_{if}). The oxide/silicon interface can consist of donor and acceptor states, considered as positive and negative charges, respectively [39]. Fig. 5.23(a) demonstrates the impact of interface charges that are similar to that shown in the previous works [39-41]. The positive charge on the interface result into more band bending at tunneling junction, and also, lowers the barrier for electrons at film under G2, thus, increases the current. The negative charge interface, on the other hand, decreases the electric field at tunneling junction, thereby reducing the current (Fig. 5.23(b)). Thus, the positive interface charges increase both the read currents, thereby enhancing the sense margin (Fig. 5.24(a)).

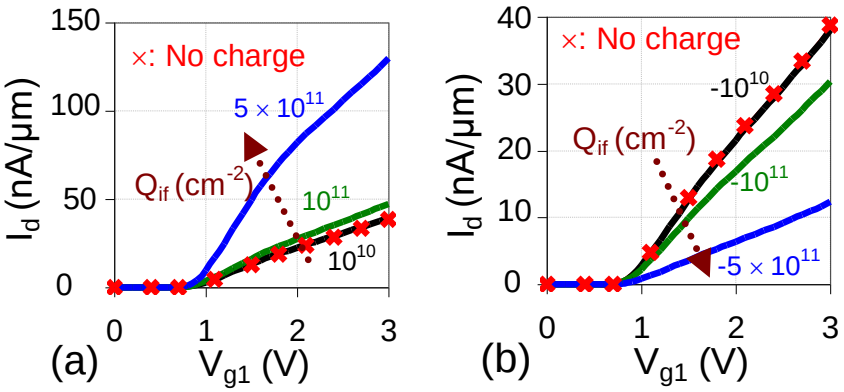

Fig. 5.23 Variation in drain current characteristics with V_{g1} as a function of charge interface (Q_{if}) with (a) donor states (positive) and (b) acceptor states (negative).

Fig 5.24(b) show improvement in retention characteristics with positive interface charges, but degrade the current ratio (Fig. 5.24(c)). However, beyond $Q_{if} \geq 10^{12}$ cm^{-2}, Read '0' being influenced more, decreases the sense margin. The trend is similar to that observed in Fig. 5.3(b) where the read characteristics are modified due to the bias applied. The negative interface charges, on the other hand, decreases current, and thus, the *SM*. However, as explained before, Read '0' is more sensitive as compared to Read '1' that increases *CR* with increase in interface charges, associated with acceptor states. The DRAM performance with interface charges has been analysed for the same set of bias values for all the operations.

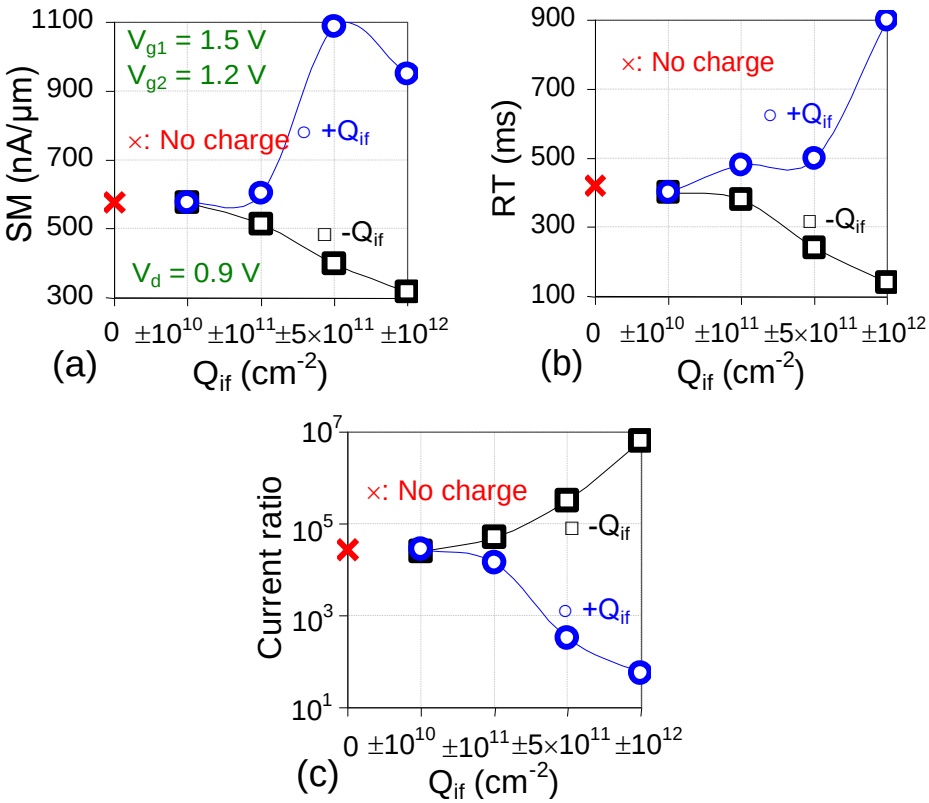

Fig. 5.24 Variation in (a) *SM*, (b) *RT*, and (c) *CR* with interface charges.

The interface charges affect the generation and recombination. The presence of negative charges increases the tunneling between G2 and G1/drain regions and thus, reduces the retention characteristics (Fig. 5.6(c)). The intrinsic region between the gates, designated as L_{gap} is influenced by presence/absence of charges and thus, impacts the tunneling between G1 and G2. While a negative charges increase the tunneling between gates, the positive interface lowers the tunneling

phenomenon. The positive interface charges lower the band energy, thereby decreasing the tunneling probability between G2 and G1/drain. The variation in DRAM characteristics (*SM*, *CR* and *RT*) are significantly observable for $Q_{if} \geq 5 \times 10^{11}$ cm^{-2}.

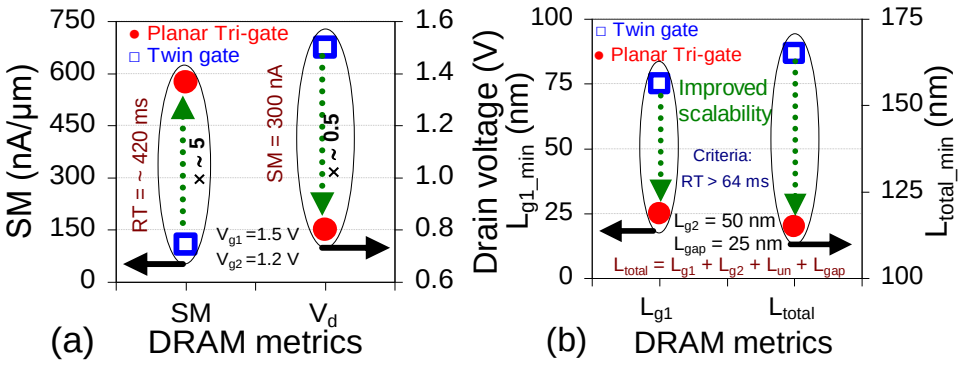

Fig. 5.25 Comparison of DRAM metrics for twin gate and planar tri-gate structure showing better performance of planar tri-gate TFET with improved (a) *SM* for same *RT* = ~ 420 ms (read: V_{g1} = 1.5 V, V_{g2} = 1.2 V, V_d = 0.9 V) voltage scaling for *SM* = 300 nA, V_{g1} = 1.5 V and V_{g2} = 1.2 V, and (b) length scalability for L_{g1} and L_{total} (total length) with *RT* > 64 ms.

5.6 Comparative Analysis

In this section, we summarize the comparison between two architectures, planar tri-gate and twin gate. For the same set of bias values (Read: V_{g1} = 1.5 V, V_{g2} = 1.2 V, V_d = 0.9 V) Fig. 5.25(a) illustrates a considerable improvement in *SM* (~ × 5) for a planar tri-gate TFET with similar retention characteristics (~ 420 ms). Secondly, for V_{g1} = 1.5 V and V_{g2} = 1.2 V, a sense margin of ~ 300 nA is obtained with V_d = 0.8 V for planar tri-gate while the same value is attained at V_d =1.5 V for twin gate architecture. Thus, it highlights better voltage scalability. Focusing on length scaling, G1 for twin gate could be scaled down to 75 nm while for planar tri-gate, downscaling G1 up to 25 nm is possible (Fig. 5.22(b)). The reduction in the barrier between G1 and G2 lead to a decrease in the stored charges that differentiates the states, thereby reducing the *SM*, and thus, limiting G1 to be scaled up to 75 nm for twin gate. Thus, a planar tri-gate architecture is adopted with the two gates controlling L_{g1} region, that reduces the short channel effects, and thus, show L_{g1} scaling down to 25 nm. Scalability analysis highlights the total

length scalability (L_{g1} + L_{gap} + L_{g2} + L_{un}) for the planar tri-gate down to 115 nm, while twin gate is scalable up to 165 nm at 85 °C (Fig. 5.25(b))

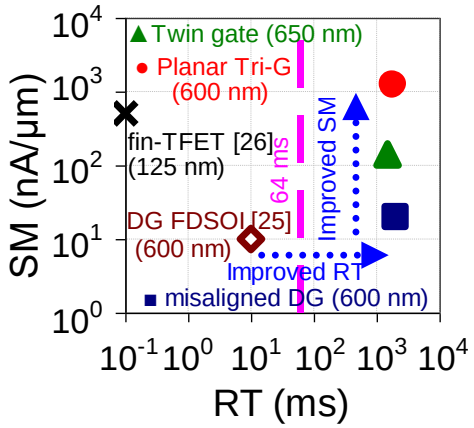

Fig. 5.26 Comparison of various TFET based DRAM in terms of *RT* and *SM* with total length (L) = L_{g1} + L_{g2} + L_{gap} + L_{un} with L_{gap}= 25 nm and L_{un} = 15 nm. Filled symbol (•) indicate L = 600 nm with L_{g1}= 400 nm, L_{g2} = 160 nm, and open symbol (o) represent a device of L = 115 nm with L_{g1}= 25 nm, L_{g2} = 75 nm. The results are for T = 85 °C, except for DG FDSOI [25] with operation at room temperature. Programming set in planar tri-gate TFET: write '1': V_{g2} = -2 V, write '0': V_{g2} = 2 V, hold: V_{g2} = -0.2 V, read: V_{g1} = 2 V, V_{g2} = 1.2 V, V_d = 1 V.

The previous results with a total length of 600 nm for a DG TFET with back gate at complete intrinsic region [25] showed a *SM* = 10 nA with *RT* in few milliseconds at room temperature while incorporation of intentionally misaligned back gate at front un-gated region resulted in a *SM* = 20 nA with *RT* = 2 sec at 85 °C. The twin gate design exhibited a *SM* = 140 nA with *RT* = 1.5 sec at 85 °C. Although, the improvement in *RT* from few milliseconds to seconds is commendable, the *SM* of TFET with L_{g1} = 400 nm and L_{g2} = 200 nm was a shortcoming (< 150 nA). The planar tri-gate TFET has succeeded in overcoming the limitation associated with *SM*, improving it to few micro-Amperes (Read biases: V_{g1} = 2 V, V_{g2} = 1.2 V, V_d = 1 V) for a total length of 600 nm with *RT* in seconds at 85 °C. It is important to note that the systematic methodology and design resulted in a tenfold improvement in *SM* along with similar retention characteristics. In addition, the total length can be scaled down to 115 nm with *RT* > 64 ms. As shown in Fig. 5.26, for shorter gate lengths, *SM* achieved is lower in

comparison to *p*-doped fin based TFET [26], but an improvement in *RT* (× ~10^3) provides an opportunity to adopt planar tri-gate TFET topology as low power DRAM.

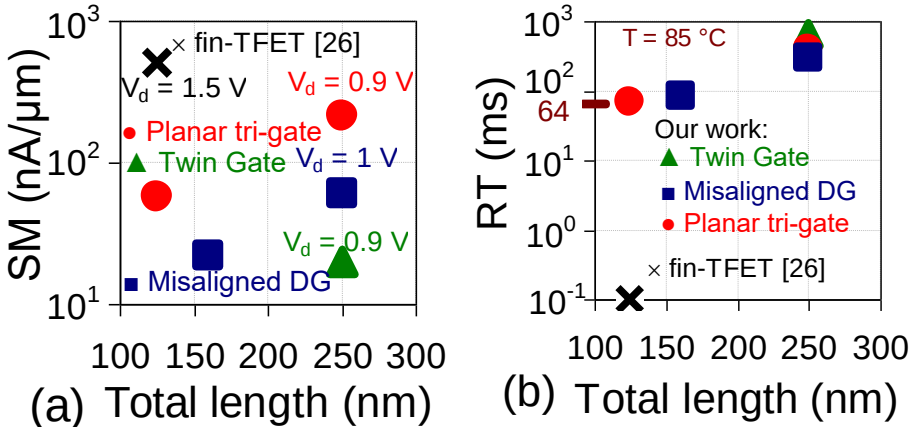

Fig. 5.27 Comparison of DRAM metrics: (a) *SM* and (b) *RT* with total length for various Tunnel FET devices based dynamic memory at 85 °C.

Figs. 5.27(a) and (b) compares various TFET based DRAM with their Sense Margin (*SM*) and Retention Time (*RT*) along with scaling perspective. Fin-based TFET [26] show maximum sense margin, but the low retention characteristics limit its utility. On the other hand, twin gate TFET with maximum retention, shows reduced sense margin that indicates trade-off and thus, lower values of performance metrics. Thus, planar tri-gate balances the trade-off between *SM* and *RT*, highlighting better performance compared to previous architectures at lower bias. This can be verified from the value of DRAM metrics (*M1*) which is ~10^2 higher than fin-based TFET for L_{total} = 125 nm, while ~6 times higher than twin gate for L_{total} = 250 nm. The improvement in metrics is more significant for shorter gate lengths, as observed in Fig. 5.22(a). Also, planar tri-gate shows better scalability with total length scaled down to 105 nm. Misaligned Double Gate (DG) topology have also succeeded in attaining *RT* > 64 ms for a total length of 160 nm, but the *SM* achieved is also low (~20 nA). Thus, results showcase the opportunities of utilizing a planar tri-gate as an energy efficient DRAM with reduced bias, gate lengths, improved *SM* and long charge sustenance at 85 °C.

5.7 Conclusion

In this chapter, an in-depth analysis of the planar tri-gate TFET as DRAM is presented with physical insights into device operation. The functionality as DRAM is based on distinctly defining the role of the gates and controlling the barriers induced. The use of an additional gate at the bottom surface, symmetric to the first front gate enhances the extent of tunneling and the transmission probability that result into improved performance metrics for planar tri-gate in comparison to twin gate architecture. This is achieved due to creation of a second conduction channel, associated with the back gate, and its coupling with the front gate. The utility of three gates, where the two gates control the Read mechanism, results in *SM* of ~1.2 μA/μm, and the third gate efficiently controls the charge sustenance with a *RT* in seconds at 85 °C, for a total length of 600 nm.

The improved gate controllability allows the device operation at lower biases with an enhanced value of ($RT \times SM$) and ($SM \times CR$) thereby permitting memory functionality at total length of ~115 nm. The analysis highlights planar tri-gate TFET as an efficient low power memory that outperforms the previous TFET based DRAM while balancing the two important metrics, sense margin and retention time. Thus, the work overcomes critical bottleneck in terms of scalability and performance metrics exhibited by previous TFET DRAM architectures. The drain bias can be scaled down to 0.8 V for the proposed tri-gate TFET which is nearly half of that required for twin gate architecture for achieving nearly the same *SM*. The planar tri-gate TFET has the potential to outperform the previous TFET based DRAM with better sense margin and scalability (length and voltage) while maintaining the retention characteristics.

5.8 References

[1] Bawedin M., Cristoloveanu S., and Flandre D., (2007) Innovating SOI memory devices based on floating-body effects, Solid-State Electronics, 51, 1252–1262.

[2] Butt N. Z., and Alam M. A., (2007) Scaling limits of double-gate and surround-gate Z-RAM cells. IEEE Transactions on Electron Devices, 5, 2255-2262.

[3] Badwan A. Z., Chbili Z., Yang Y., Salman A. A., Li Q., and Ioannou D. E., (2013) SOI field-effect diode DRAM cell: Design and operation, IEEE Electron Device Letters, 34, 1002-1004.

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

[5] Wan J., Zaslavsky A., and Cristoloveanu S., (2015) Comment on Investigation of tunnel field-effect transistors as a capacitor-less memory cell, Applied Physics Letters, 106, 016101.

[6] Wan J., Le Royer C., Zaslavsky A., and Cristoloveanu S., (2013) Progress in Z^2-FET 1T-DRAM: retention time, writing modes, selective array operation, and dual bit storage, Solid-State Electronics, 84, 147-154.

[7] Almeida L. M., Sasaki K. R. A., Caillat C., Aoulaiche M., Collaert N., Jurczak M., Simoen E., Claeys C., and Martino J. A., (2013) Optimizing the front and back biases for the best sense margin and retention time in UTBOX FBRAM, Solid-State Electronics, 90, 149-154.

[8] Giusi G., Alam M. A., Crupi F., and Pierro S., (2010) Bipolar mode operation and scalability of double gate capacitorless 1T-DRAM cells, IEEE Transactions on Electron Devices, 57, 1743-1750.

[9] Aoulaiche M., Nicoletti T., Almeida L. M., Simoen E., Veloso A., Blomme P., Groeseneken G., and Jurczak M., (2012) Junction Field Effect on the Retention Time for One-Transistor Floating-Body RAM, IEEE Transactions on Electron Devices, 59, 2167-2172.

[10] Giusi G., (2014) Physical insights of body effect and charge degradation in floating-body DRAMs, Solid-State Electronics, 95, 1-7.

[11] Kim K., and Jeong G., (2007) Memory technology for sub-40 nm node in IEEE Electron Devices Meeting, 27-30.

[12] Luryi S., Xu J., and Zaslavsky A., (2013) Future trends in microelectronics: Frontiers and Innovations, John Wiley & Sons, pp. 59-70.

[13] Kim K., (2000) Perspectives on giga-bit scaled DRAM technology generation, Microelectronics reliability, 40, 191-206.

[14] Kim C. H., and Chang L., (2011) Guest editors' introduction: Nanoscale Memories Pose Unique Challenges, IEEE Design & Test of Computers, 28, 6-8.

[15] Itoh K., (2011) Embedded memories: Progress and a look into the future, IEEE Design & Test of Computers, 28, 10-13.

[16] Gamiz F., (2016) Capacitor-less memory: Advances and challenges, Joint IEEE International Workshop and International Conference on Ultimate Integration on Silicon, pp. 68-71.

[17] Wan J., Royer C. L., Zaslavsky A., and Cristoloveanu S., (2011) Tunneling FETs on SOI: Suppression of ambipolar leakage, low-frequency noise behavior, and modeling, Solid-State Electronics, 65-66, 226-33.

[18] Biswas A., Dan S. S., Royer C. L., Grabinski W., and Ionescu A. M., (2012) TCAD simulation of SOI TFETs and calibration of non-local band-to-band tunneling model, Microelectronic Engineering, 98, 334-337.

[19] Fukuda K., Mori T., Mizubayashi W., Morita Y., Tanabe A., Masahara M., Yasuda T., Migita S., and Ota H., (2013) A compact model for tunnel field-effect transistors incorporating nonlocal band-to-band tunneling, Journal of Applied Physics, 114, 144512.

[20] Boucart K., and Ionescu A. M., (2007) Double-Gate Tunnel FET with High-κ Gate Dielectric, IEEE Transactions on Electron Devices, 54, 1725-1732.

[21] Lu H., and Seabaugh A., (2014) Tunnel field-effect transistors: state-of-the-art, IEEE Journal of the Electron Devices Society, 2, 44-49.

[22] Knoch J., Mantl S., and Appenzeller J., (2007) Impact of the dimensionality on the performance of tunneling FETs: Bulk versus one-dimensional devices, Solid-State Electronics, 51, 572-578.

[23] ITRS International Technology Working Groups. International Technology Roadmap for Semiconductors, http://www.itrs.net (2015).

[24] Biswas A., Dagtekin N., Grabinski W., Bazigos A., Royer C. L., Hartmann J.-M., Tabone C., Vinet M., and Ionescu A. M., (2014) Investigation of tunnel field-effect transistors as a capacitor-less memory, Applied Physics Letters, 104, 092108.

[25] Biswas A., and Ionescu A. M., (2015) 1T capacitor-less DRAM cell based on asymmetric tunnel FET design, IEEE Journal of the Electron Devices Society, 3, 217-222.

[26] Biswas A., and Ionescu A. M., (2014) Study of fin-tunnel FETs with doped pocket as capacitor-less 1T DRAM, IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference, pp.1-2.

[27] Park J. S., (2002) Dual Gate MOSFET Fabrication US Patent 6168998 B1.

[28] Anderson, B. A., Bryant, A., Clark Jr, W. F. and Nowak E. J., (2006) Low capacitance FET for operation at subthreshold voltages US patent 7009265 B2 2006.

[29] Chung T. M., Olbrechts B., Sodervall U., Bengtsson S, Flandre D., and Raskin J.-P., (2007) Planar double-gate SOI MOS devices: fabrication by wafer bonding over pre-patterned cavities and electrical characterization, Solid-State Electronics, 51, 231-238.

[30] Jeon Y., Kim M., Lim D., and Kim S., (2015) Steep subthreshold swing n- and p-channel operation of bendable feedback field-effect transistors with p^+–i–n^+ nanowires by dual-top gate voltage modulation, Nano Lett., 15, 4905–4913.

[31] Dirani H. E., Solaro Y., Fonteneau P., Ferrari P., and Cristoloveanu S., (2016) Properties and mechanisms of Z^2-FET at variable temperature, Solid-State Electron, 115, 201–206.

[32] Chen W. C., Lue H. T., Hsiao Y. H., and Lu C. Y., (2017) A Novel SuperSteep Subthreshold Slope Dual-Channel FET Utilizing a Gate-Controlled Thyristor Mode-Induced Positive Feedback Current, IEEE Transactions on Electron Devices, 64, 1336-1342.

[33] Colinge J.-P., (2004) Multiple-gate SOI MOSFETs, Solid-State Electronics, 48, 897–905.

[34] Zhang L., Chan M., and He F., (2010) The impact of device parameter variation on double gate tunneling FET and double gate MOSFET, IEEE conference on Electron Devices and Solid-State Circuits, pp. 1-4.

[35] Toh E. H., Wang G.H., Chan L., Samudra G., and Yeo Y. C., (2007) Device physics and guiding principles for the design of double-gate tunneling field effect transistor with silicon-germanium source heterojunction, Applied Physics Letters, 91, 243505.

[36] Chen S., Huang Q., and Huang R., (2014) Source doping profile design for Si and Ge tunnel FET, ECS Transactions, 60, 91-96.

[37] Abdi D. B., and Kumar M. J., (2014) Controlling ambipolar current in tunneling FETs using overlapping gate-on-drain, IEEE Journal of the Electron Devices Society 2 187-90.

[38] Chattopadhyay A., and Mallik A., (2011) Impact of a spacer dielectric and a gate overlap/underlap on the device performance of a tunnel field-effect transistor, IEEE Transactions on Electron Devices, 58, 677-683.

[39] Qiu Y., Wang R., Huang Q., and Huang R., (2014) A comparative study on the impacts of interface traps on tunneling FET and MOSFET IEEE Transaction on Electron Devices, 61, 1284–1291.

[40] Madan J., Chaujar R., (2016) Interfacial charge analysis of heterogeneous gate dielectric-gate all around-tunnel FET for improved device reliability, IEEE Transactions on Device and Materials Reliability, 16, 227-234.

[41] Venkatesh P., Nigam K., Pandey S, Sharma D., Kondekar P. N., (2017) Impact of Interface Trap Charges on Performance of Electrically Doped Tunnel FET with Heterogeneous Gate Dielectric, IEEE Transactions on Device and Materials Reliability, 17, 245-252.

Chapter 6

Conclusion and Scope for Future Work

6.1 Conclusion

This chapter presents the summary of different aspects related to the functioning of TFETs as DRAM. TFET devices were introduced to replace drift-diffusion based conventional Metal Oxide Semiconductor (MOS) that sets a fundamental limit to subthreshold slope (60 mV/decade at room temperature) [1]. The device is benefited with low off-current (I_{off}) a high I_{on}/I_{off}, reduced short channel effects, weak temperature dependency and enhanced scalability [1]. Being a promising candidate for low power applications, TFET has recently been explored as capacitorless DRAM [2-4]. Although, previously published work [2-4] led towards tunneling based device as DRAM, the *RT* which is the most crucial metric of DRAM showed values lower than the target of 64 ms, specified by ITRS [5]. Therefore, a careful reinvestigation was required to enhance the charge retention of TFET based DRAM. The research work presented in the thesis provides insights into the understanding of the performance and behaviour of TFET devices for memory applications through comprehensive physical device simulations [6].

The key contribution of this research is to provide insights into the physical phenomenon occurring in the device, which influences the operation of TFET as dynamic memory, with a focus on improving retention and scaling capability. The thesis work demonstrates device perspective, where various DRAM metrics are regulated by device architecture (misaligned, twin, and planar tri-gate TFET) geometry (L_{g1}, L_{g2}, L_{gap}, L_{un}, T_{si}) parameters (T_{ox}, gate workfunctions) biases and temperature. These parameters govern hole generation and recombination in the storage region that defines distinct operations (write, read and hold) of DRAM. The three architectures present a systematic advancement in TFET based DRAM,

while analysing various DRAM performance metrics. The key conclusions of the work are as follows:

I. Design Perspective of Tunnel FET based DRAM

The functionality of architectures as DRAM is based on the distinct role of each gate, where the first gate (G1) is an n^+ poly gate (~4.15 eV) and is primarily used to control the read mechanism based on BTBT. The second gate (G2) is a p^+ poly gate (~5.25 eV) and is used to create an electrically induced potential well for charge storage. While the first gate is aligned to source at top surface, the second gate is aligned to drain, positioned at the bottom at the front un-gated region in misaligned DG, and at the front, adjacent to the first gate in twin gate and planar tri-gate topology. The non-overlapping of G1 and G2 along with a higher difference in the workfunctions of the two gates result into a profound well formation. The regulation of barriers created by the gates (G1 and G2) through the optimal use of device parameters and bias lead to better performance, with significant improvement at scaled lengths as compared to other tunneling based dynamic memory architectures [2-4].

II. Misaligned Tunnel FET based DRAM

The work shows innovative viewpoints of transforming gate misalignment, traditionally considered detrimental into a unique opportunity, coupled with appropriate selection of back gate workfunction and bias to significantly enhance the charge sustenance in capacitorless DRAM. The back gate is engineered by positioning it intentionally at the region uncovered by the front gate and using a gate with high workfunction value (p^+ poly G2). This creates a more profound potential well that aids sustain charges for longer duration. However, along with formation of a deeper physical well, optimal bias values are also essential to preserve charges for longer duration, and hence, improve *RT*. During hold, a more negative back gate bias results into significant hole generation, while a more positive bias into hole recombination from the storage region, and thus, optimal bias is required to balance hole generation and recombination [7-9]. Optimal back gate workfunction, alignment and bias result into high *RT* of ~ 170 ms at 85 °C for an UTBOX TFET with BOX thickness of 10 nm. This can be further enhanced

by utilizing a Double Gate (DG) TFET with *RT* of ~ 2 seconds at 85 °C for L_{g1} = 400 nm and L_{g2} = 200 nm, which is a significant improvement over previous works [2-4].

In addition, the thesis showcases a methodological assessment at 85 °C that highlights optimization of device design through three steps: misalignment, lateral spacing between the gates (L_{gap}) and an underlap between the G2 and drain region (L_{un}) that improves,

(i) the total length scalability from 220 nm (without L_{gap} and L_{un}) to 160 nm for L_{g1} = 75 nm, (considering *RT* > 64 ms [5] as the criteria)

(ii) *RT* from ~ 80 ms to ~ 300 ms with the use of L_{gap}, which can be further improved to ~ 600 ms with the introduction of L_{un} for gate length (L_{g1}) and storage region (L_{g2}) of 100 nm,

(iii) G2 scalability down to 25 nm with L_{g1} = 100 nm, (considering *RT* > 64 ms as the criteria)

(iv) *RT* by a factor of 3, compared to case without L_{gap} and L_{un} for a constant total length (200 nm).

The improvement in device performance is attributed the increase in effective length of storage region (L_{g2}+L_{gap}+L_{un}). More significantly, L_{gap} and $L_{un,}$ reduces the generation of holes associated with tunneling during hold state '0' between G1-G2 and G2-drain regions, respectively, that enhances state '0' sustenance, and hence, the *RT*.

III. Twin gate Tunnel FET based DRAM

The applicability of TFET based dynamic memory is further extended through use of twin gate device that demonstrates *RT* = 1.5 s with *SM* = 140 nA for L_{g1} = 400 nm and L_{g2} = 200 nm. The device can be scaled down to L_{g1} = 75 nm and L_{g2} = 50 nm with use of an underlap at 85 °C. Besides, temperature assessment shows the decrease in retention with increasing temperature due to higher generation and recombination. However, twin gate can be operated efficiently up till 105 °C with L_{g1} = L_{g2} = 100 nm with *RT* > 64 ms. Additionally, its applicability for embedded memory is explored due to compatibility with conventional MOS devices and process [1,10,11]. Improved *RT* reduces the refresh rates, which along with

tunneling based write mechanism demonstrate low power operation. The low write time of 5 ns, indicates its use for high speed applications, which further nominates it as eDRAM. Thus, improved retention, scalability, speed, and operation at high temperature, demonstrates device applicability for both, standalone as well as embedded applications. Although, low on-current is a limitation in TFET device that results into low Sense Margin (*SM*) in TFET based DRAMs, the high current ratio achieved makes it a viable option with acceptable read sensitivity. However, further analysis reflects the drawback of low *SM* can be overcome through incorporation of a symmetric G1 as in planar tri-gate TFET.

IV. Planar Tri-gate Tunnel FET as Dynamic Memory

The use of an additional gate at the bottom surface, symmetric to the first front gate enhances the extent of tunneling and the transmission probability that result into improved *SM* and scalability. G2 is dedicated for creating a deep potential well that enables charge sustenance for longer duration, and thus, benefits *RT*. These results in,

(i) *SM* of ~1.2 μA/μm, with *RT* in seconds at 85 °C, for a total length of 600 nm.

(ii) drain bias scaling down to 0.8 V, which is nearly half of that required for twin gate architecture for achieving nearly the same *SM*,

(iii) scalability of G1 down to 25 nm and that of G2 down to 50 nm with *RT* > 64 ms and acceptable *SM*,

(iv) total length scalability down to ~115 nm, which is better than twin gate and misaligned DG TFET showing scalability down to ~160 nm,

Thus, while twin gate and misaligned DG TFET outperforms the previous TFET based DRAM architectures [2-4] in terms of retention and scalability, planar tri-gate showcase improved DRAM metrics than twin gate as well as misaligned DG TFET. The thesis showcases the opportunities of utilizing a planar tri-gate as an energy efficient dynamic memory with reduced bias, gate lengths, improved *SM* and long charge sustenance at 85 °C. Thus, the work presented in the thesis presents insights and new viewpoints for TFET to function efficiently as DRAM.

The physical insights and analysis of different attributes with optimal utilization lead to improved performance metrics as well as suppressed trade-offs. The

systematic analysis through innovative approaches leads to capacity, retention, at low energy with operation at reduced size. The use of an energy optimized DRAM memory with RT > 64 ms is well-suited for standalone applications, and as well as, for the integrated circuits embedded with logic devices. Although device analysis highlights its applicability for memory applications, the fabrication is an overhead due to requirement of additional procession steps, as described in chapter 4. Also, the low sense margin due to tunneling based transport mechanism as compared to MOS based architectures [7-9], and other p^+-i-n^+ architecture [12,13] needs further investigation, and also, development of alternative device topologies.

6.2 Scope for future work

6.2.1. Applicability as eDRAMs

The scalability of TFET has improved in comparison to other similar structures [2-4], but still need exploitation to compete with conventional DRAM. Conventional 1T-1C DRAM [14,15] has been investigated since several decades and has been continuously scaled. The major issue associated with conventional DRAM include capacitor scaling, which has been resolved in recent years with the use of 3D cell storage capacitor and vertical array transistors. However, the fabrication cost and uniformity issues are challenging [15]. Tunneling based DRAMs or even other 1T DRAMs could be a substitute to conventional DRAMs, but are not yet competitive to replace them commercially. However, the utility of TFETs as cheap embedded memory with CMOS process compatibility is also expected to be of interest.

eDRAMs are being developed to replace SRAM to achieve cheaper and lower voltage embedded chips [11,16,17]. Also, the advancement of various technologies with IoT, cloud computing and big data, requires evolution in the computing landscape [11,16,17]. The fabrication cost associated with capacitor in 1T-1C for embedded applications can be reduced through use of TFETs, which have been demonstrated as reliable, low power and high speed DRAM cells. Although, TFET shows a possibility as embedded applications, further investigation and validation of its functionality at circuit level, considering

different bitline disturbance could serve as guideline for device and circuit designers. It would be useful for the semiconductor industry in the developing of dynamic memories.

6.2.2 Study of other factors affecting DRAM metrics

Dynamic memories are sensitive to Random Dopant Fluctuations (RDFs) ionizing and radiation effects, and traps [7,8]. These factors affect the retention characteristics of the device, which is significantly controlled by carrier lifetime. The carrier lifetime depends upon film material, doping and the traps.

(a) From retention perspective, the requirement is of material with high carrier lifetime [9,10]. Germanium can be utilized due to higher carrier lifetime than Silicon. However, lower energy band gap could lead to increased BTBT, affecting retention characteristics. Therefore, requirement is of hetero-structures and strained Silicon material that performs optimally through band gap as well as carrier lifetime engineering. The material properties on integration with devices having ease of fabrication can be well-suited for real time applications.

(b) While TFETs require additional fabrication steps, Junctionless transistors having same type of dopant throughout the film show simpler fabrication process [18]. Although Junctionless devices have a heavy channel doping that will degrade carrier lifetime, and thus, retention, the availability of more charge carriers could lead to low write time, and thus, better speed. Such characteristics would show applicability as eDRAMs, where speed is prime concern with retention in few milliseconds [14,15]. Thus, exploitation of an optimal material-device co-design would be a choice for memory in future.

(c) Other than these factors, the effect of traps in tunneling based devices is under investigation [19,20]. The impact of traps is based on its distribution in the energy band gap and thus, needs insightful exploration. The traps affect the carrier lifetime as well as the sub-threshold swing [20]. The degraded subthreshold swing (being quantified through ratio of I_{on}/I_{off}) indicates either decrease in I_{on} that decreases sense margin, or enhanced I_{off} that decreases charge retention. Although our work focussed on design optimization, understanding the impact of carrier lifetime reduction (chapter 2) along with calibration with degraded S (> 60 mV/decade) [21], reflects on the impact of

traps. The effect of traps can be added after understanding their type (donor and acceptor) and position in the energy band gap. Thus, exploiting their impact could further throw light on the viability of TFET as DRAM with better physical insights.

6.3.3. Improved performance through hybrid memories

Recently, the hybrid memories are evolving that combines the benefits of different memory technologies into a single device to show efficient performance [22]. Hybrid memories have been introduced to reduce the energy consumption of a system that consumes ~40% of the total and thus, the device integrates non-volatile memory and DRAM in one chip [11,22]. The thesis work presents the advantages as well as shortcomings of TFET for memory applications. The limitations of these devices can be overcome through a hybrid structure, which could be of interest in near future.

6.3 References

[1] Ionescu A. M., and Riel H., (2011) Tunnel field-effect transistors as energy-efficient electronic switches, Nature, 479, 329-337.

[2] Biswas A., Dagtekin N., Grabinski W., Bazigos A., Le Royer C., Hartmann J. M., Tabone C., Vinet M., and Ionescu A. M., (2014) Investigation of tunnel field-effect transistors as a capacitor-less memory cell, Applied Physics Letters, 104, 092108.

[3] Biswas A., and Ionescu A. M., (2015) 1T capacitor-less DRAM cell based on asymmetric tunnel FET design, IEEE Journal of the Electron Devices Society, 3, 217-222.

[4] Biswas A., and Ionescu A. M., (2014) Study of fin-tunnel FETs with doped pocket as capacitor-less 1T DRAM, IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference, pp. 1-2.

[5] ITRS International Technology Working Groups. International Technology Roadmap for Semiconductors, http://www.itrs.net (2015).

[6] Atlas User's Manual, Silvaco Int., Santa Clara, CA, Jan. 18, 2015.

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

[8] Cho M. H., (2012) Thin-Body SOI Capacitorless DRAM Cell Design Optimization and Scaling. University of California, Berkeley, Technical Report No. UCB/EECS-2012-176.

[9] Moon D. I., Choi S. J., Han J. W., Kim S., and Choi Y. K., (2010) Fin-width dependence of BJT-based 1T-DRAM implemented on FinFET, IEEE Electron Device Letters, 31, 909-911.

[10] Gamiz F., (2016) Capacitor-less memory: Advances and challenges, Joint International EUROSOI Workshop and International Conference on Ultimate Integration on Silicon, pp. 68-71.

[11] Lee S. H., (2016) Technology scaling challenges and opportunities of memory devices, In Electron Devices Meeting, pp. 1-8.

[12] Badwan A. Z., Chbili Z., Yang Y., Salman A. A., Li Q., and Ioannou D. E., (2013) SOI field-effect diode DRAM cell: Design and operation, IEEE Electron Device Letters, 34,1002-1004.

[13] Wan J., Royer C. Le, Zaslavsky A., and Cristoloveanu S., (2013) Progress in Z^2-FET 1T-DRAM: retention time, writing modes, selective array operation, and dual bit storage, Solid-State Electronics, 84, 147-154.

[14] Keeth, B., Baker, J. R., Jacob, B., Lin, F., (2007) DRAM circuit design: fundamental and high-speed topics 2nd ed. John Wiley & Sons Inc., New Jersey (ISBN- 978-0-4701-8475-2).

[15] Joodaki, M., (2012) Selected advances in nanoelectronic devices: logic, memory and RF, Springer Science and Business Media, Heidelberg (ISBN-978-3-642-31349-3).

[16] Micron Technology, Inc. 2015 Winter Analyst Conference

[17] E. Hannigan, McKinsey on Semiconductors, fifth edition, McKinsey & Company , (2015).

[18] Colinge J.P., Lee C.W., Afzalian A., Akhavan N.D., Yan R., Ferain I., Razavi P., O'neill B., Blake A., White M., and Kelleher A.M., (2010) Nanowire transistors without junctions, Nature nanotechnology, 5, 225-229.

[19] Qiu Y., Wang R., Huang Q., and Huang R., (2014) A comparative study on the impacts of interface traps on tunneling FET and MOSFET IEEE Transaction on Electron Devices, 61, 1284–1291.

[20] Sant S., Schenk A., Moselund K. and Riel, H., (2016) Impact of trap-assisted tunneling and channel quantization on InAs/Si hetero tunnel FETs, IEEE Device Research Conference (DRC), (pp. 1-2).

[21] Wan J., Royer C. L., Zaslavsky A., and Cristoloveanu S., (2011) Tunneling FETs on SOI: Suppression of ambipolar leakage, low-frequency noise behavior, and modeling, Solid-State Electronics, 65-66, 226-33.

[22] Jing W., Yang K., Lin Y., Lee B., Yoon S., Ye Y., Du Y. and Chen B., (2017) Retention-Aware Hybrid Main Memory (RAHMM): Big DRAM and Little SCM, IEEE Transactions on Computers, 66, 912-918.