

Charge Trap Transistors (CTT): Turning Logic Transistors into Embedded Non-Volatile Memory for Advanced High-k/Metal Gate CMOS Technologies

A seminar report (EE539) submitted in partial fulfillment of
the requirements for the degree of

Master of Technology (VLSI)

by

Risabh Saxena
(Entry No. 2023EEM1024)

Under the guidance of
Prof./Dr. Brajesh Rawat



DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLOGY ROPAR

2023

Declaration

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

Rishabh Saxena

Entry No. 2023EEM1024

Date:11 Nov 2023

Abstract

The requirement for embedded non-volatile memory (eNVM) in contemporary computer systems is increasing at a rapid rate; nevertheless, there aren't as many possibilities as there once were because of issues with integration, scalability, and operating voltage compatibility. This study presents a novel (MTPM) solution for advanced high-k/metal-gate (HKMG) CMOS technologies. This approach converts ordinary logic transistors that are as-fabricated into eNVM components without requiring extra masks or process adders. When used as eNVM components, these logic transistors are known as "Charge Trap Transistors" (CTTs). This study successfully demonstrates the implementation of CTT eNVM in several manufacturing technologies, including 32 nm, 22 nm, 14 nm, and 7 nm. This study investigates the current state of the memory technology landscape and specifically focuses on the position of CTT technology within this domain. In addition to the wide array of digital applications, it is worth noting that CTTs may also serve as an analog memory for various purposes, such as neuromorphic computing in the fields of machine learning (ML) and artificial intelligence (AI).

Contents

Abstract	i
List of Figures	v
1 Introduction	1
1.1 Objective	1
2 An Introduction to Charge Trap Transistors (CCTs) and an Overview	5
3 Improved Self-Heating Charge Trapping and Optimized CTT Design	11
3.1 The manipulation of self heating enhanced charge trapping by bias modulation.	12
3.2 Charge Injection Mechanism and the Charge Trapping Profile	13
3.3 DATA RETENTION	14
3.4 Fundamental Understanding	17
4 Modeling and Reliability Consideration	19
5 Conclusion	21

List of Figures

1.1	SEM voltage contrasting vs. unblown and unblow anti-fuses	2
1.2	CTT vs. alternative eNVM resolutions	2
1.3	Mapping of Emerging eNVM	3
2.1	Oxygen deficiency in HfO_2 molecule	6
2.2	TEM cross-sections of FinFET CTT on coordinates	6
2.3	A schematic of the CTT programming operation.	7
2.4	Measured ΔV_T from a CTT programmed with and without device self-heating.	8
2.5	Measured activation energies (E_a) for CTTs	8
2.6	The operation of CTT memory	9
3.1	ΔV_T function of PVRS and fixed V_D values	12
3.2	Reverse vs. forward-mode division for saturation and linear (a) VT and (b) channel current in FinFET CTTs.	14
3.3	Normalizing delta between forward- and reverse-mode reads in FinFET CTTs.	14
3.4	% charge loss vs. bake time 85 °C	15
3.5	16
3.6	High-temperature data retention for FinFET CTTs	16
4.1	Off-state gate leakage current for CTT threshold voltage shift (ΔV_T).	20

Chapter 1

Introduction

1.1 Objective

- The advancement high-K/metal gate's (HKMG) on-chip non-volatile memory availability. One of the leading contenders for the next generation of complementary MOS (CMOS devices) is high K/metal gate (HKMG).
- The high-K dielectric used in this innovative technology lowers leakage and raises the dielectric constant.
- HKMG technology lowers gate leakage by combining low K-dielectric and metal gates. This increases transistor capacitance and enables chips to operate with less power.
- The two patterns that flow in common are gate-first and gate-last. However, integration problems and process complexity are major reasons for the scalability of eFLASH into FinFET technology. Emerging memory technologies like PCM, ReRAM, and MRAM contrast with widely operable and multi-programmable logic capacitor voltages. This called for more intricate procedures and masks.
- The primary goal of the HKMG method is to create an embedded non-volatile memory (eNVM) technology that is multi-time programmable (MTP) and compatible with logic compatible voltages (-2V) for process freedom.
- Charge Trap Transistor (CTT) eNVM provides a safe solution for data and hardware security. It is possible to reverse engineer the eFUSE and anti-fuse memory in the data store seen in figure 1.1.

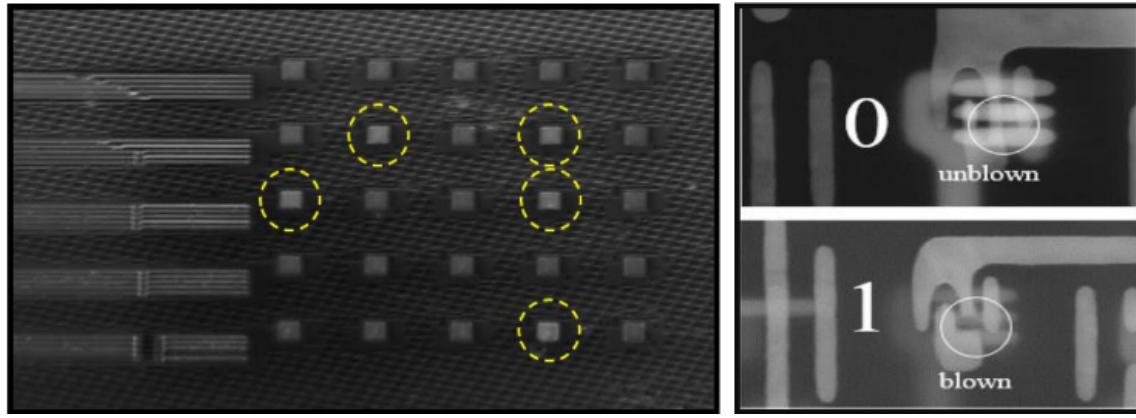


Figure 1.1: SEM voltage contrasting vs. unblown and blown anti-fuses

- Figure 1.2 displays the comparison between the different eNVM and the CTT. [?] There

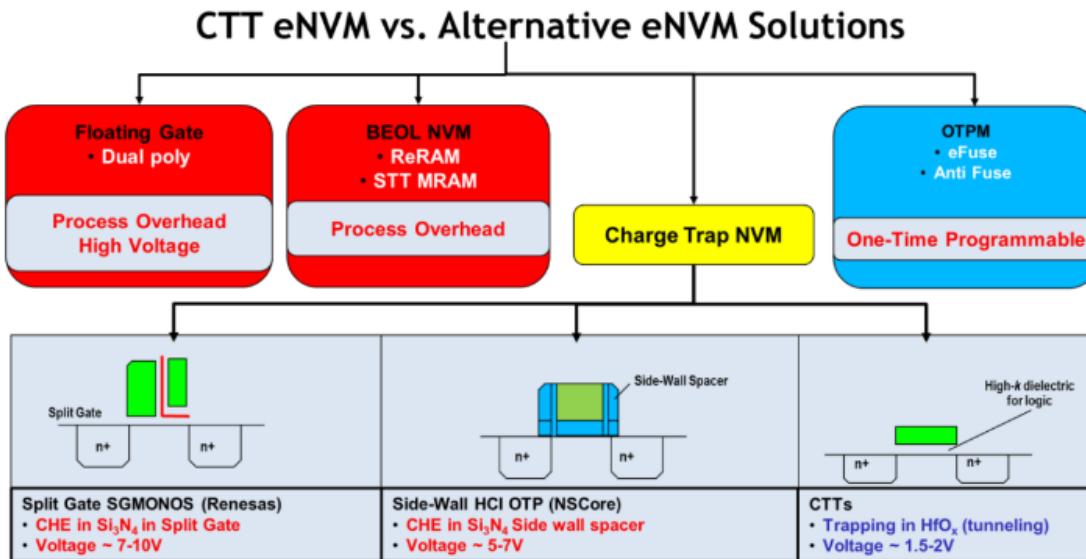


Figure 1.2: CTT vs. alternative eNVM resolutions

is a demand for eFLASH replacement technologies due to the complexity and scaling issues with eFLASH and the lack of a clear roadmap to sub-28 nm nodes. Fig. 1.3 shows an overview of the eNVM ecosystem, the developing technologies that might replace SRAM/eDRAM for working memory and eFLASH for code/data storage at 14nm technology nodes .

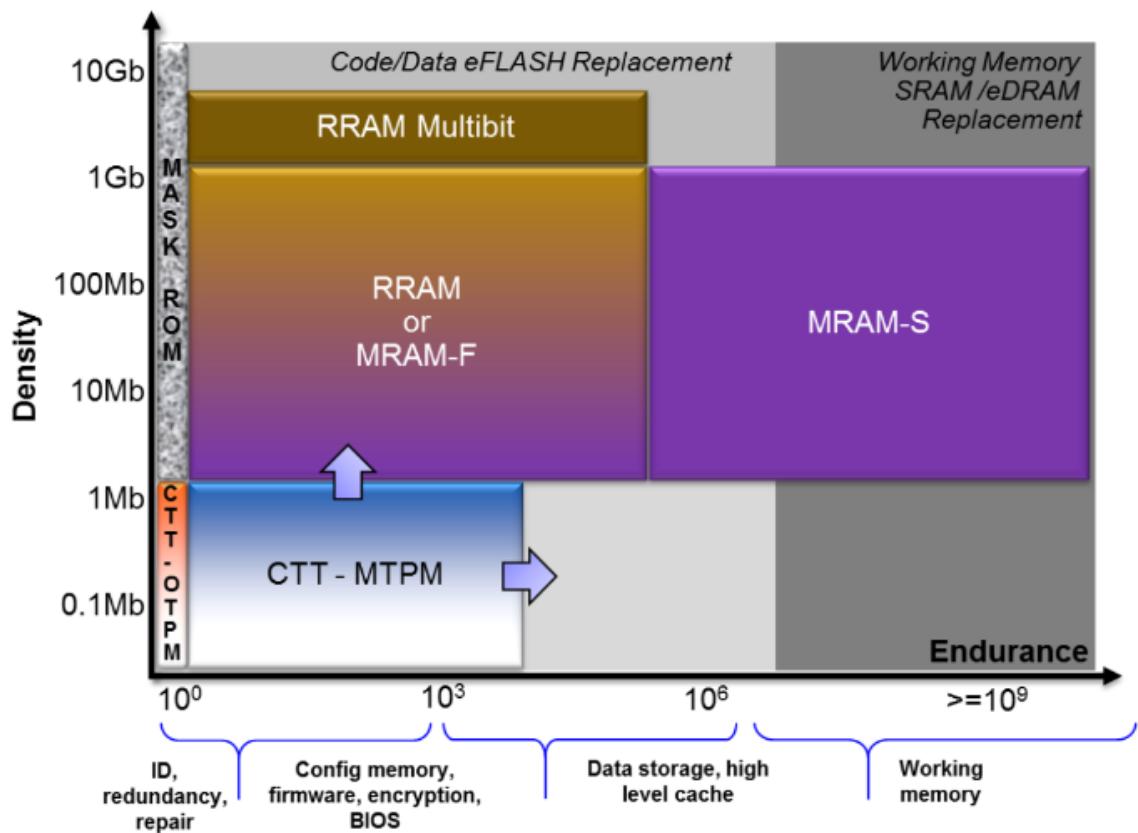


Figure 1.3: Mapping of Emerging eNVM

Objective The primary goal is to present the basic physics and operating principles of CTTs. On the other hand, it also showcases the HKMG CMOS advanced technology nodes' CTT eNVM technology.

The 32nm, 22nm, 17nm, and 14nm manufacturing technologies that are accomplished all the implementation of CTT eNVM.

Chapter 2

An Introduction to Charge Trap

Transistors (CCTs) and an Overview

- An oxygen vacancy, which is produced when one oxygen atom separates from the HfO_2 molecule to form HfO_2 , is shown in Figure (2.1). An oxygen vacancy, a positively charged vacancy defect and thermodynamic point defect, is left behind by the diffusion of oxygen from HfO_2 .
- In addition, it is widely recognized that temperature plays a crucial role in expediting the occurrence of defects and charge trapping in HfO_2 resulting from bias stress. This study demonstrates that the phenomenon of charge trapping in HfO_2 may be leveraged as a beneficial characteristic for incorporating non-volatile memory (eNVM) in high-k metal gate (HKMG) CMOS technologies. Despite being commonly seen as an undesirable factor due to its tendency to introduce device unpredictability and therefore impact circuit performance, this research highlights the potential use of charge trapping in eNVM applications.
- Numerous studies have provided evidence that the utilization of suitable voltages, which align with logical processes and above the standard threshold of roughly 0.9V, can result in an increase in charge trapping inside the high-k gate dielectric material of HKMG logic transistors. This phenomena has the capability to produce variations in threshold voltage, which are represented as ΔV_T , that are significant in size and consistent in stability. Therefore, these transitions can be efficiently employed as a method of non-volatile data storage. [?]

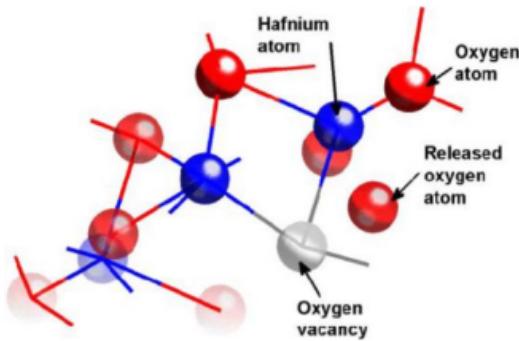


Figure 2.1: Oxygen deficiency in HfO_2 molecule

- Temperature significantly accelerates HfO_2 the bias stress-Produced charge trapping and defect development. Applications for non-volatile memory (NVM) have already been suggested in "H.K. dielectric". When eNVM elements are used, the typical logic transistors are replaced with the (CTT) charge trap transistors.
- Put very simply, the HKMG CMOS standard logic transistors depicted in Fig2.2 functioned in an enhanced charge trapping state. [?]

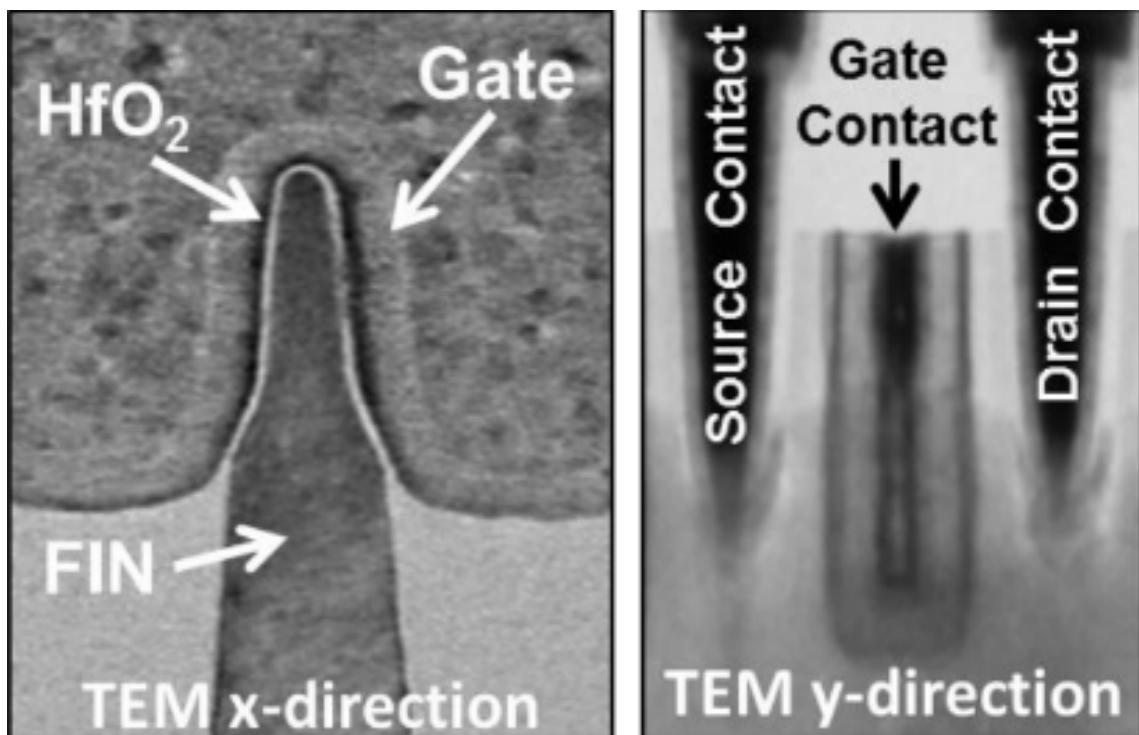


Figure 2.2: TEM cross-sections of FinFET CTT on coordinates

- The 14nm FinFET CTT's TEM cross-section is parallel to the FIN direction in one direction and perpendicular to it in the other. The CTTs are upgraded charge trapping devices

that self-heat. Depending on the technology, the device threshold voltage (V_T) is the voltage at which a certain event occurs and is changed by the charge trapped in (HKMG).

- Short gate bias V_G pulses of around 1.8–2.0V are commonly used for programming, together with drain bias V_D of approximately 1.3–1.6V, and source and substrate biases of 0V Fig 2.3 The existence of a strong vertical field enables the introduction of electrons into the gate, resulting in their confinement inside the high-k dielectric material. This technique successfully increases the threshold voltage (V_T). Concurrently, the utilization of a high channel current pulse results in the occurrence of self-heating within the device.

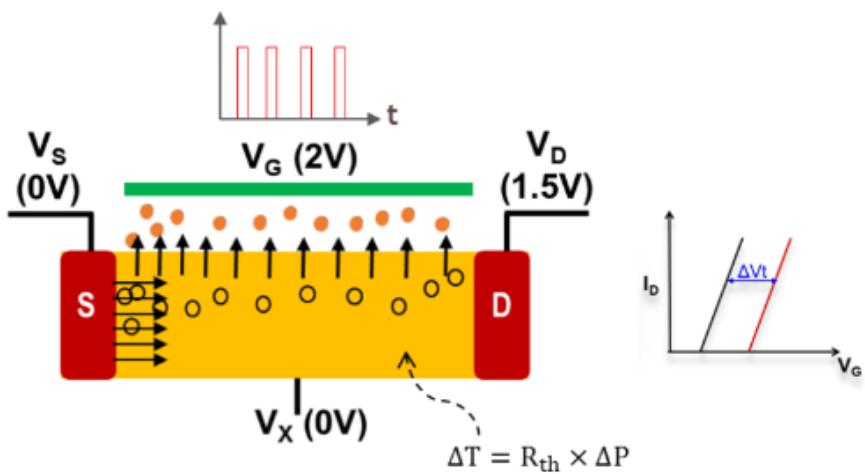


Figure 2.3: A schematic of the CTT programming operation.

The temperature that causes self-heating is positively correlated with the charge trapped. the drain (V_D) of sin 1.3 – 1.6V and the source bias (V_S) of sin 1.8 – 2.0V, while the measure ΔV_T from CTT programmed in which the self-heating device on the magnitude in stable (retention) where $\Delta T = R_{th} * \Delta P$ is displayed in figs.2.4 and fig2.5. The source bias (V_G) and substrate bias (V_X) are 0V in fig.2.3.

- The data retention lifespan of > 10 years at 125°C is demonstrated by the measuring of charge activation energies (E_a) for compiled devices at self-heating temperatures on different devices (fig. 2.5). Other chapters have a thorough explanation of all that has been said.
- Consequently, each distinct V_T value may be interpreted as a very unique bit, such as "0" and "1" for two distinct V_T curls, as the modulation device V_T is charge trapped in the high-K dielectric. Transistor V_T displays the fundamental equation, in which Q_{ox}

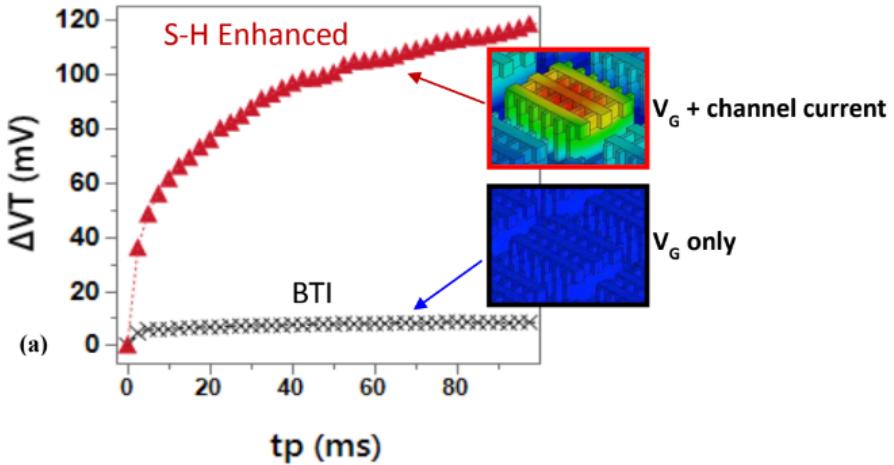


Figure 2.4: Measured ΔV_T from a CTT programmed with and without device self-heating.

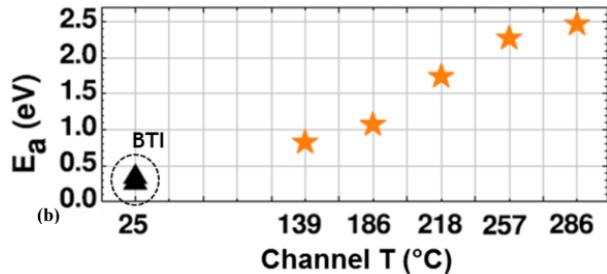


Figure 2.5: Measured activation energies (E_a) for CTTs

represents the quantity modified by the charged particles trapped in the gate dielectric.

$$V_{TO} = 2\Phi_F + \Phi_{ms} - \frac{Q_{ox}}{C_{ox}} - \frac{Q_{dm}}{C_{ox}} \quad (2.1)$$

where

- Φ_{ms} work function difference b/w gate metal and substrate
- Q_{ox} trapped charge in the dielectric
- C_{ox} capacitance of gate dielectric
- Φ_F substrate Fermi potential
- Q_{dm} maximum charge held by depletion layer

So, its shows that the amount of charge present in high-K dielectric is

$$V_T = V_{TO} + \Delta V_T$$

where

$$\Delta V_T = \frac{\Delta Q_{OX}}{C_{OX}}.$$

- Currently, we discuss the (STAR) which expands as the "Self-HGeating temperature As-sisted ehase" that has been designed to achieve high erase efficiency and issue the address, ultimately leading to a considerable improvement in the erase cycle endurance and window memory in CTTs.

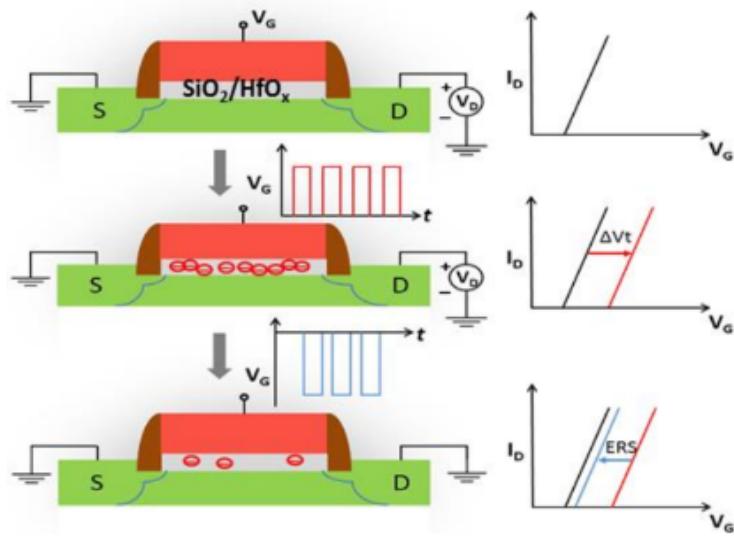


Figure 2.6: The operation of CTT memory

- This memory device is CTT. Note that FinFET technologies are also subject to the inherent problem of greater charge trapping due to self-heating. The CTT eNVM implementation has demonstrated how to accomplish various tasks in the following technology nodes: (32nm ,22nm)SOI planar,14nm bulk FinFET,14nm FinFET, and 7nm bulk FinFET.
- A prototype memory array in a fully functional product, the FinFET (Fin Field-Effect transistor) is a multi-gate device that offers voltage logic compatible operation, high density, scalability ($-0.144\mu m^2/bit$ for 22nm and $\approx 0.082\mu m^2/bit$ for 14nm technology).
- Compared to current **one-time programmable (OTP) technologies like eFUSE**, CTT technology has an advantage. A unique kind of non-volatile memory (NVM) known as an OTP allows data to be written to memory just once. Gate breakdown anti-fuse, with its low power, high density, and high scalability at no additional processing cost, is

hence more useful for improving yield, chip configuration , redundancy, and field module testing.

Chapter 3

Improved Self-Heating Charge Trapping and Optimized CTT Design

- The High-k/metal-gate (HKMG) CMOS logic devices' charge trapping behavior is examined, investigated, and described in relation to device self-heating. It goes without saying that the given gate bias, or charge taken field, affects how much charge is trapped. It is shown, however, that the charge trapping behavior is also significantly influenced—possibly even more . The temperature (T) of the channel during charge injection (programming) is affected by the thermal resistance (R_{th}) of the device. The occurrence of self-heating-induced charge trapping has been confirmed and examined in several commercial technologies, as evidenced and analyzed in the coming chapters. The technologies described above include the 32 nm SOI planar, 22 nm SOI planar, 14 nm SOI FinFET, 14 nm bulk FinFET, and 7 nm bulk FinFET nodes. This chapter provides a comprehensive illustration of the implementation of CTTs in semiconductor technology, specifically focusing on the utilization of 14 nm bulk FinFET nodes and 22 nm SOI planar nodes.
- So, In this chapter we learn about the Charge Trap Transistors in 22nm SOI planer And 14nm bulk in FinFET. i,e

$$\Delta T = R_{th} * P = R_{th} * (I_{ch} * V_D) \quad (3.1)$$

where

- I_{ch} channel current

- V_D applied drain-source bias
- In this chapter, the implications of the results concerning the use of high-k/metal-gate logic devices, or "Charge Trap Transistors" or "CTTs," as embedded memory components are discussed. Investigating the potential of high-k/metal-gate CMOS technologies in the context of non-volatile data storage is the main objective of this study in order to reduce the need for additional process complexity. Also included are elements for optimizing bitcell design and CTT memory operation conditions.

3.1 The manipulation of self heating enhanced charge trapping by bias modulation.

Understanding the dynamic aspects of charge trapping and associated bias dependency requires first using the (PVRS) technique to investigate drain bias effects. This method's inability to differentiate between the impacts of the self-heating process and the horizontal electric field is a drawback. When layout-dependent effects are carefully controlled to regulate the thermal resistance (R_{th}) of devices while maintaining all other electrical parameters constant, one can quantify and fully understand the effect of device self-heating on charge trapping in CTTs, and distinguish between the effects of thermal and electric field factors.

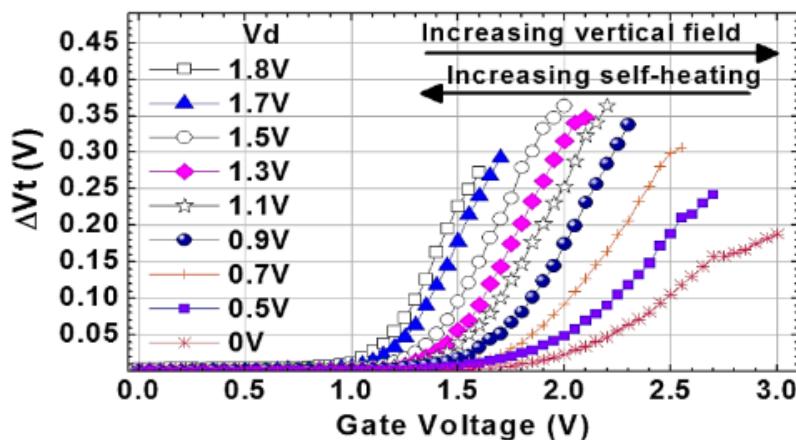
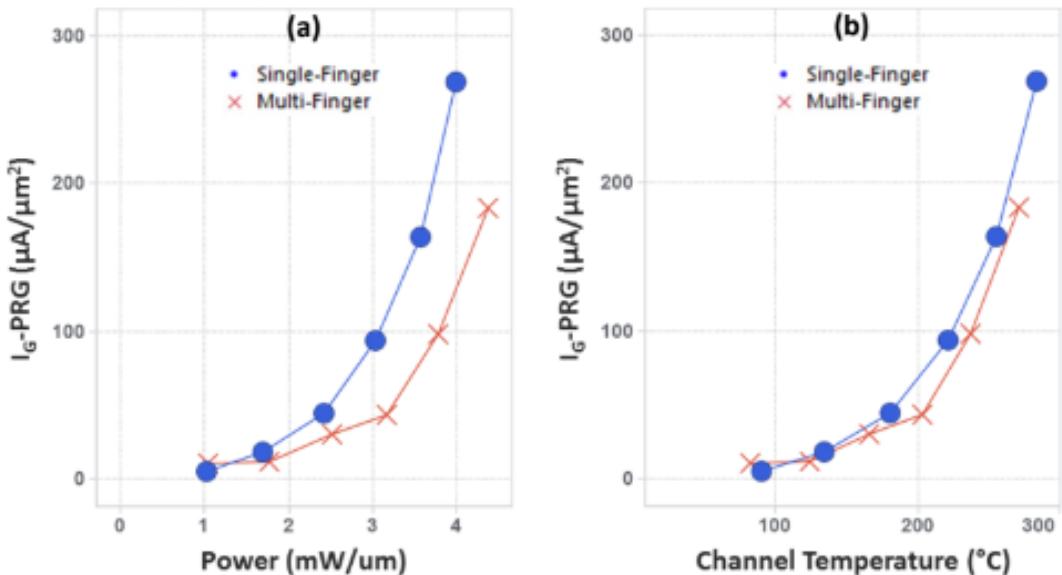


Figure 3.1: ΔV_T function of PVRS and fixed V_D values

3.2 Charge Injection Mechanism and the Charge Trapping Profile

In order to improve understanding of the charge injection behavior and its underlying mechanisms, the charge injection currents are observed at the gate terminal of 14nm Fin-FET CTTs during the programming process. The graphical representation in Figure 3.2(a) depicts the progressive augmentation of normalized charge injection currents during the programming process. The phenomena of augmentation is exhibited in devices that possess different thermal resistances (R_{th}) and are subjected to various power densities. In line with the observed trends in the relationship between threshold voltage and drain-source voltage, it is evident that, when subjected to equivalent power density conditions, the charge injection current in the single-channel device with a higher thermal resistance (R_{th}) is notably higher compared to the split-channel device with a lower thermal resistance (R_{th}). After conducting an examination of the relationship between the function and the anticipated channel temperature, as depicted in Figure 3.2(b), it becomes evident that the charge injection current characteristics of both devices demonstrate a significant level of resemblance. To improve understanding, it is important to acknowledge that there exists a notable correlation between the magnitude of charge injection currents and the temperature at which the device experiences self-heating, relative to the applied power density.



The results indicate that there is no statistically significant impact of polarity on the average values of device V_T and saturation currents, as shown in Figure 3.3 and Figure 3.4. This suggests that although there may be some degree of asymmetry, the distribution of trapped charge throughout the channel is generally uniform. The stochastic fluctuation that corresponds to the normalized deltas between forward- and reverse-mode scans (Fig. 3.4) leads to a relatively low standard deviation of 2.8%.

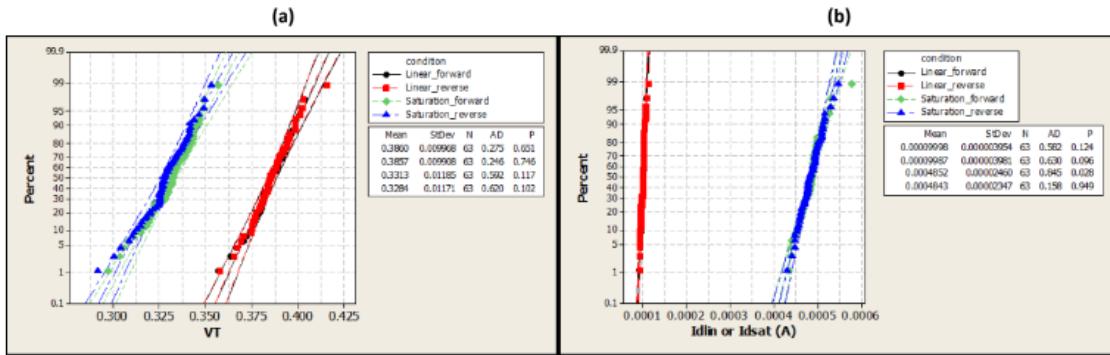


Figure 3.2: Reverse vs. forward-mode division for saturation and linear (a) V_T and (b) channel current in FinFET CTTs.

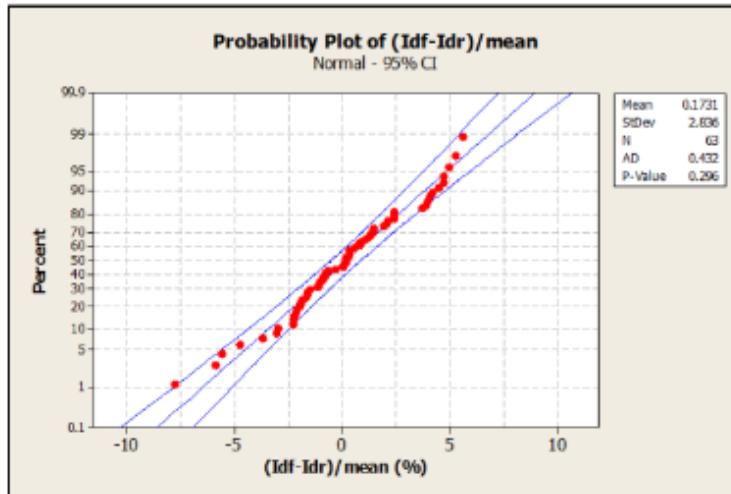


Figure 3.3: Normalizing delta between forward- and reverse-mode reads in FinFET CTTs.

3.3 DATA RETENTION

To access the charge retention characteristics, a set of identical 22nm planar silicon-on-insulator (SOI) charge trapping transistors (CTTs) is programmed at different specified

values of drain voltage (V_D). Each device is intended to have a cumulative threshold voltage shift (ΔV_T) of about 250 mV. The aforementioned gadgets are then heated to 85°C to preserve them. The retention of trapped charges in each device is evaluated by tracking its threshold voltage (V_T) for a predetermined period of time. The graph in Figure 3.5 visually depicts the percentage decline in the variable V_T , which indicates the decrease in trapped charge, with respect to its starting values. There is a correlation between the trapped charge retention and the programming drain bias, V_D . [?]

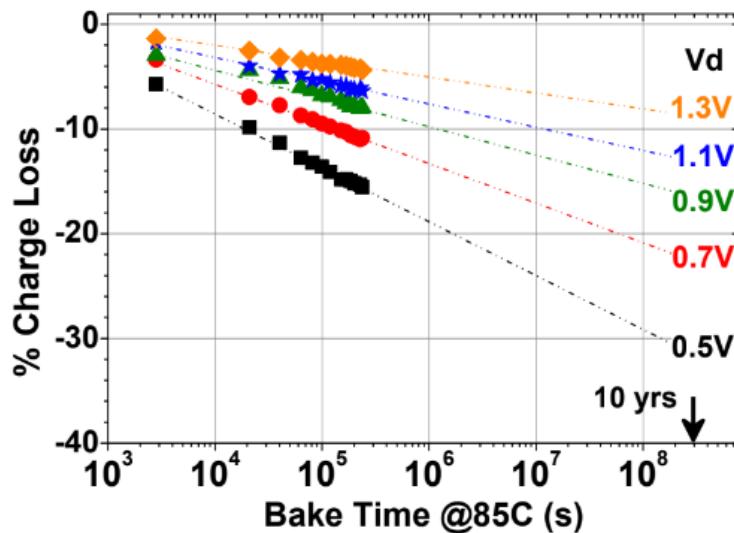


Figure 3.4: % charge loss vs. bake time 85 °C

A group of devices with varying channel widths (while maintaining the same length) and varying channel lengths (while maintaining the same width) is configured using Programmable Voltage Reference Sources (PVRS) at a voltage of 1.5V in order to attain a combined threshold voltage shift (ΔV_T) of about 265mV in each device. These devices are subsequently subjected to storage at a temperature of 85 °C. The measurement of the trapped charge retention is conducted as previously stated and is depicted in Figure 3.6. It is evident that broader and shorter devices exhibit increased retention of trapped charge.

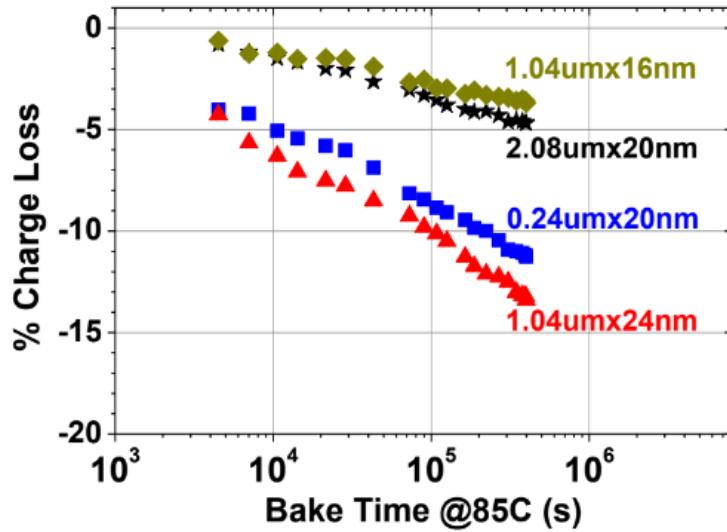


Figure 3.5

The results of high-temperature charge retention bake tests conducted on 22 nm CTTs indicate that there is an estimated charge loss of less than 25% over a period of 10 years when exposed to a temperature of 125 °C . The findings obtained from the 14 nm FinFET CTTs indicate a predicted reduction in charge of less than 25% over a period of 10 years when exposed to a temperature of 125 °C, as seen in Figure 3.7.

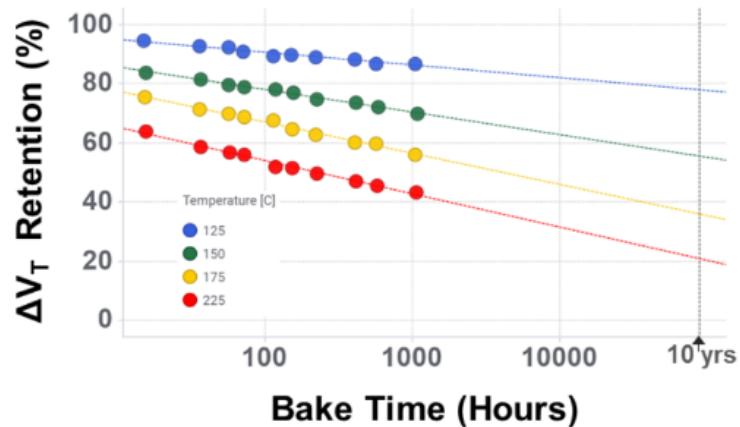


Figure 3.6: High-temperature data retention for FinFET CTTs

The utilization of the Arrhenius equation enables the determination of the acceleration factor (AF) by the following procedure:

$$AF = \frac{\tau_1}{\tau_2} = e^{[\frac{E_a}{K}(\frac{1}{T_2} - \frac{1}{T_1})]} \quad (3.2)$$

- T_2 = Accelerated stress temperature in Kelvin

- k = Boltzmann's constant
- E_a = Activation energy (eV)
- T_1 = Operation temperature in Kelvin

3.4 Fundamental Understanding

- This chapter presents a thorough examination of the occurrence of enhanced charge trapping in HKMG CMOS transistors as a result of self-heating, along with its potential ramifications for memory-related purposes. The research presents findings that demonstrate the improved trapped charge magnitude and stability (retention) of HKMG CMOS transistors, especially known as "Charge Trap Transistors" or "CTT," when exposed to device self-heating during the charge injection process, also referred to as programming operation.
- The attainment of comparable levels of charge trapping can be expedited by employing shorter durations and/or reduced gate bias, while also exhibiting enhanced stability (retention) when the devices are programmed under elevated self-heating circumstances. Additionally, this paper presents various strategies aimed at optimizing the design of the CTT bitcell in order to improve programming efficiency. The discussion has focused on the manipulation of device architecture to optimize the phenomenon of self-heating aided charge trapping, which is the fundamental concept behind the operation of Charge Trap Transistors (CTTs).
- Even when exposed to temperatures as high as 125 °C, the CTT technology exhibits excellent data retention properties and has a lifespan of over ten years. This method also exhibits compatibility with logic voltage operation and has the advantage of scalability. It is a good candidate for incorporation as an embedded non-volatile memory (eNVM) due to the aforementioned characteristics. Furthermore, it may take the place of current one-time programmable (OTP) memory technologies like gate breakdown anti-fuse and eFUSE. Many applications, including chip identification, on-chip encryption, field configurability, redundancy, repair, hardware security enhancement, yield improvement, and performance modification within HKMG CMOS technologies, may benefit from the use

of this replacement. Moreover, the CTT technique has a significant cost advantage over other memory technologies since it doesn't need any additional steps or masks.

Chapter 4

Modeling and Reliability Consideration

The comprehension and resolution of device dependability have significant relevance in all technological domains. The enhancement of reliability in the CTT eNVM technology may be achieved by the optimization of operational circumstances and bitcell architecture, as extensively described in preceding chapters. However, it is important to note that some challenges, such gate leakage current, electromigration and dielectric breakdown, may still arise. The aforementioned problems, as well as strategies for effectively mitigating them, are expounded upon in this discourse. An empirical observation has been made indicating a positive correlation between the threshold voltage shift (ΔV_T) and the rise in gate leakage current in Complementary Tunnel Transistors (CTTs). The observed outcome can be attributed to the anticipated rise in trap density inside the HfO₂ layer due to the stress induced during the programming and erasing processes. As a result, this occurrence leads to a rise in stress-induced leakage current (SILC), which can be attributed to the mechanism of trap-assisted tunneling (TAT) over preexisting flaws. A comparable pattern may be observed inside the realm of the on-state gate leakage current (IG-ON) in this particular circumstance. There exists a discernible trade-off between the memory window (ΔV_T) and the gate leakage current. The term IG-OFF is employed to denote the gate leakage current that occurs when a specific cell inside a memory array exhibits a low VG (gate voltage) and a high VD (drain voltage). The phrase "IG-ON" refers to the occurrence of gate leakage current, which is particularly evident during the read operation of a cell [?]

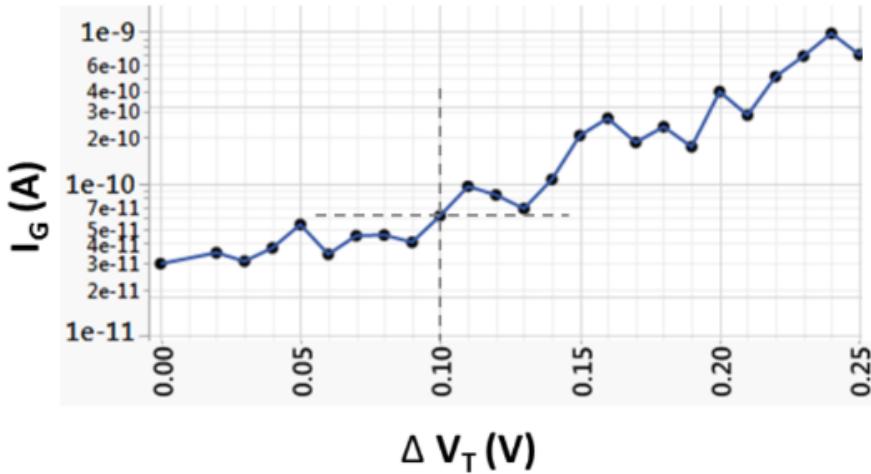


Figure 4.1: Off-state gate leakage current for CTT threshold voltage shift (ΔV_T).

This chapter provides an overview of a concise model that is capable of accurately characterizing and predicting the charge trapping phenomena in CTTs. The model presents a distinct and explicit elucidation of the correlation between electric-field and self-heating temperature, particularly within the framework of charge trapping in CTTs. Furthermore, it is possible to expand this model to incorporate charge trapping in high-k metal gate (HKMG) devices on a broader scale. The adoption of a compact model offers several benefits in the optimization of operational conditions and bitcell design for the CTT eNVM. The model has demonstrated a significant level of concurrence with experimental data collected from different bitcell designs and over a wide range of programming configurations. This comprises all feasible operating situations for CTT eNVM (Configurable Test Technology electrically Non-Volatile Memory). This chapter also includes an analysis of reliability issues related to the CTT eNVM technology, such as gate leakage current, dielectric breakdown, and electromigration. Several novel techniques, such as STAR, AC-STAR, and AC-PRG, have demonstrated their effectiveness in addressing the aforementioned issues.

Chapter 5

Conclusion

The Charge Trap Transistor (CTT) is an innovative kind of embedded non-volatile memory (eNVM) technology. It allows for the transformation of ordinary logic transistors, which are produced using high-k/metal (HKMG) CMOS technology nodes, into programmable memory components that possess the ability to be reprogrammed many times. This change is achieved without the employment of any supplementary approaches or protective coverings. The current study has successfully showcased the underlying principles that regulate the operation and physical characteristics of CTTs, as well as the potential feasibility of employing CTT eNVM technology in real-world commercial applications. The integration of the theoretical knowledge, operational principles, and innovations explored in this study has led to the accomplishment of effectively implementing a commercially accessible CTT eNVM device capable of functioning under extreme temperatures typically seen in military applications. The integration of CTT eNVM has been effectively shown in several manufacturing technologies, including 32 nm SOI planar, 22 nm SOI planar, 14 nm SOI FinFET, and 14 nm bulk FinFET. The confirmation of the application of CTT technology on 7 nm nodes has also been validated.

The CTT technology utilizes conventional logic transistors in their original state as elements for electrically non-volatile memory (eNVM). This is achieved by taking use of the self-heating increased charge trapping phenomenon. A comprehensive analysis has been conducted on the underlying principles of self-heating improved charge trapping in HKMG CMOS transistors, along with its potential implications for memory applications. The experimental findings indicate that the trapped charge's magnitude and stability, particularly its retention, are substantially improved through device self-heating. This enhancement leads to notable and consistent shifts in the threshold voltage (ΔV_T) of the device, rendering it suitable for non-volatile

memory applications that necessitate operation at high temperatures. Specifically, the device has demonstrated a data retention lifetime exceeding 10 years at a temperature of 125 °C. This paper also includes methods for optimizing the design of the CTT bitcell in order to improve programming efficiency.

References

- [1] F. Khan, E. Cartier, J. C. S. Woo and S. S. Iyer, "Charge Trap Transistor (CTT): An Embedded Fully Logic-Compatible Multiple-Time Programmable Non-Volatile Memory Element for High- k -Metal-Gate CMOS Technologies," in IEEE Electron Device Letters, vol. 38, no. 1, pp. 44-47, Jan. 2017, doi: 10.1109/LED.2016.2633490.
- [2] J. Robertson and Ka Xiong, "Defect energy levels in HfO₂," 2006 International Workshop on Nano CMOS, Mishima, Japan, 2006, pp. 175-179, doi: 10.1109/IWNC.2006.4570991.
- [3] F. Khan, E. Cartier, C. Kothandaraman, J. C. Scott, J. C. S. Woo and S. S. Iyer, "The Impact of Self-Heating on Charge Trapping in High- k -Metal-Gate nFETs," in IEEE Electron Device Letters, vol. 37, no. 1, pp. 88-91, Jan. 2016, doi: 10.1109/LED.2015.2504952.