The Effects of Total-Ionizing-Dose on Charge-Trap Transistors and Implications for Charge-Trap Memories

By

Rachel Mae Brewer

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Electrical Engineering

August 13, 2021

Nashville, Tennessee

Committee:

Brian D. Sierawski, Ph.D. (chair)

Ronald D. Schrimpf, Ph.D.

Daniel M. Fleetwood, Ph.D.

En Xia Zhang, Ph.D.

Joshua D. Caldwell, Ph.D.

This dissertation is dedicated in memory of Grandaddy, for his encouragement to pursue graduate studies, his interest in my research, and his love for me that was clearly evident.

ACKNOWLEDGEMENTS

I would first like to thank Dr. Brian Sierawski for being an excellent advisor and mentor throughout my entire time in graduate school. In all ways, you have been the perfect advisor for me: involved and always available for questions while still giving me independence to explore and discover within my topic. I am immensely thankful for your advice and suggestions on research, employment opportunities, and with every other thing I have asked for guidance. Since day one of graduate school, you have been an advocate for me and my success in both my research and professional life for which I am incredibly grateful.

I would like to next thank the Defense Threat Reduction Agency (DTRA) for sponsoring portions of the research presented in this dissertation. Also, thank you to Dr. Subramanian Iyer of UCLA for working with me and providing the test chips for the experiments conducted in this work. And thank you to UCLA graduate students Steven Moran and Jon Cox for your help with the CTTs and for answering my questions about them.

In addition, I want to thank all my committee members for their thoughtful suggestions and comments on my dissertation work and research as a whole. To Dr. Ron Schrimpf, I am thankful for your instruction and education from my first year of graduate school in both classes and research. To Dr. Dan Fleetwood, I appreciate your help in crafting the TNS paper and assisting with reviewer responses and am also thankful for your expertise on all things TID. To Dr. En Xia Zhang, thank you for your assistance in collecting experimental data in addition to your instruction on the use of the lab equipment. And to Dr. Josh Caldwell, thank you for your comments and suggestions from the perspective of a non-radiation effects researcher. All of you have been instrumental in helping shape my research and contributions to the field of radiation effects.

Thank you also to the myriad of other professors that have assisted me in some way as well. For the professors at the 3D meetings, thank you for your weekly feedback and suggestions on the current state of my research. Dr. Robert Reed, thank you for introducing me to Vanderbilt as a prospective graduate student; additionally, I appreciate your feedback on the early portions of my research. To Dr. Kevin Warren, thank you in particular for your help and advice during my job search. Dr. Mike Alles, thank you for your big-picture questions and for keeping me focused on the "why" of my research.

I also want to thank the many friends I have made in the RER group these past four years. To my dissertation buddies and closest Nashville friends Kaitlyn Ryder, Landen Ryder, and Brandon Smith, thank you for your support along this journey as we encouraged, shared with, and laughed with each other. I am thankful to have had you three to bounce ideas off of and for all our discussions that helped me gain perspective from those in similar places in grad school. To Drew Tonigan, I am thankful for your perspective and advice as a senior grad student and for being my testing trip buddy during my first heavy-ion testing experience. To Ellis Richards, I am grateful for the close friendship we formed starting at the office through similar interests that has continued even after separation by many miles and states. And to Rebekah Austin, thank you for being a senior graduate student and friend from year one that could provide me with advice and expertise on Vanderbilt and life in Nashville.

Last but certainly not least, I want to express my thanks to my family who have supported me and my educational pursuits throughout my entire life. To my parents, particularly my mom, thank you for providing the foundation of all my academic success. You gave me an excellent education by homeschooling me every year from kindergarten through 12^{th} grade. I would not be where I am today without your love or without the sacrifices you made to give me the best

education and upbringing possible. To my brothers, thank you for your love, support, and friendship. And to my grandmother and late grandfather, thank you for your love and encouragement to attend graduate school, and thank you for being supporters and cheerleaders throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Charge-trap transistors (CTTs) enhance and exploit the properties of the dielectric layers of high-*k*-metal-gate (HKMG) devices for use as memory elements [1]. The amount of charge trapped in the HKMG dielectric layer is determined by the degree of voltage-ramp-stress (VRS). The threshold voltage V_{th} shifts due to the resulting charge trapping in the gate oxide are sufficient and stable enough for use in non-volatile memories (NVM) [1]–[3]. CTTs are fabricated in commercial CMOS processes and are a potential low-power, non-volatile memory. As a commercial CMOS process, CTTs are manufactured without additional process complexity and therefore can be designed as an embedded memory along with other circuitry. Because CTTs have high integration and low power consumption, they are attractive for use as digital or analog memories in neuromorphic computing, which is brain-inspired, non-von Neumann computing. However, CTTs are not just limited for use in neuromorphic computing architectures. Due to CTTs being low-cost, low-power, multi-time programmable memories (MTPM), they can be used in conventional von Neumann architectures as well.

However, total-ionizing-dose (TID) environments can similarly introduce trapped charges within dielectric materials potentially introducing a competing memory effect. Therefore, in this dissertation, the direct current (DC) characteristics of individual charge-trap transistors manufactured in both 22 nm fully-depleted silicon-on-insulator (FDSOI) and 14 nm bulk fin field-effect transistor (FinFET) technologies are investigated for their total-ionizing-dose and programming response. Past work on FDSOI and bulk devices show that FDSOI is sensitive to

TID while bulk devices are not [4]–[16]. This work investigates and quantifies the TID response of FDSOI and bulk FinFET CTTs. FDSOI CTTs were found to be sensitive to TID as expected from the literature. It was also found that 2-fin bulk FinFET CTTs were impervious to TID up to 500 krad(SiO_2), while 40-fin bulk FinFET CTTs saw increased sub-threshold leakage current with increasing dose. It is demonstrated that charge-trap transistors can maintain their programmed state in TID environments when not degraded by the process isolation structures.

The key contributions of this work are as follows. First, the TID response is characterized for individual 22 nm planar FDSOI and 14 nm bulk FinFET CTTs, which is the first ever examination of the response of CTTs to TID. Unprogrammed transistors are exposed to TID and characterized to determine the effect of TID on the two process technologies. In addition, CTTs are programmed before irradiation, and CTTs are programmed after irradiation to determine what effect, if any, TID has on the programmability of CTTs. The collected experimental data is then applied to neural network simulations in a forward-looking view to assess the potential ramifications of a TID-environment on CTTs used as neuron weights in neural networks.

Several important results are found. The most important finding is that the tolerance of CTTs to TID is found to depend primarily on the properties of the nearby insulators in the transistors, such as the buried oxide (BOX) or shallow trench isolation (STI), *not* on the memory element. 22 nm FDSOI CTTs are found to lose their programmed state above ~150 krad(SiO_2) as a result of hole trapping in the BOX. On the other hand, 14 nm bulk FinFET CTTs with two fins are minimally affected by TID. 40-fin, 14 nm bulk transistors have increasing subthreshold leakage currents with increasing TID.

In addition, programming is shown to result in sufficient electron trapping in the gate oxide to support non-volatile memory applications for both technologies. Furthermore, the order of

programming and irradiation had no effect on programmability of 14 nm bulk FinFET CTTs, while 22 nm FDSOI CTTs saw higher shifts when programmed after irradiation due to neutralization of trapped holes in the BOX.

Finally, application of the experimental data to exploratory neural network simulations shows potential for CTTs to be used in neural networks as neuron weights. Gradual decreases in classification accuracy are projected with increasing TID. In addition, strong dependence on architecture is found, indicating there is potential for TID effects in CTT-based neural networks to be mitigated by conscientious development of the neural network architecture.

The remainder of the dissertation is organized as follows. Chapter 2 gives important background for understanding the contributions resulting from this dissertation. In addition, work done by other individuals on implementation of CTTs in neural networks is also presented. The next chapter, Chapter 3, focuses on the experimental setup for the irradiation, programming, and subsequent data analysis of the CTTs tested in this work. Chapters 4 and 5 present the experimental results collected on the CTTs. The focus of Chapter 4 is the TID response of the 22 nm FDSOI CTTs and the impact of TID on the programmability of the devices. Chapter 5 similarly focuses on the characterization and programmability of the 14 nm bulk FinFET devices, both those with 2 fins and those with 40 fins. Following the experimental results, Chapter 6 is a discussion and analysis of the data collected in the previous two chapters. Simulations exploring the potential application of CTTs to neural networks within a TID environment is studied in Chapter 7 using the experimental data collected in Chapters 4 and 5. Finally, Chapter 8 summarizes the findings and conclusions resulting from this dissertation.

CHAPTER II

BACKGROUND

Introduction to Charge-Trap Transistors

Charge-trap transistors (CTTs) enhance and exploit the properties of the dielectric layers of high-*k*-metal-gate (HKMG) devices for use as memory elements [1]. The amount of charge trapped in the HKMG dielectric layer is determined by the degree of voltage-ramp-stress (VRS). The threshold voltage V_{th} shifts due to the resulting charge trapping are sufficient and stable enough for use in non-volatile memories (NVM) [1]–[3]. Total ionizing dose (TID) irradiation can introduce competing effects in the dielectric layers of NVM transistors [17]–[19]. Thus, it is important to determine whether TID irradiation reduces the effective memory windows, based on threshold voltage shifts of CTTs, in advance of potential use in space or other high-radiation environments.

In this work, the programming and TID irradiation responses of hafnium-based high-*k* CTTs fabricated in GlobalFoundries 22FDX and 14LPP processes [20], [21] using their conventional CMOS process flow have been evaluated. Each type of CTT has three initial conditions: unprogrammed, programmed (Fig. 1), and erased [1]. The unprogrammed state is the initial state as the as-processed device. After initial programming of the as-processed device, the multi-time programmable CTT can be cycled between programmed and erased states [1], [3], [22]–[24]. In each case, the programmed state of the CTT is established by electron trapping in the HKMG dielectric. For the 22 nm fully depleted silicon-on-insulator (FDSOI) devices, the radiation response is determined primarily by hole trapping in the buried oxide (BOX). For 14 nm bulk

devices, minimal V_{th} shifts are observed when devices are irradiated, but multi-fin devices show increasing leakage with increasing TID.

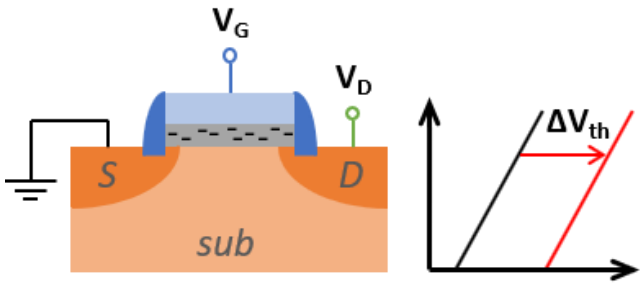


Fig. 1. Programming the CTT causes accumulation of electrons in the gate oxide. The threshold voltage shifts positively. *(Adapted from [1].)*

The Physics of Charge-Trapping in HKMG Oxides

Most advanced CMOS technologies already use HfO_2 in the gate oxide, so CTTs can be manufactured as NVM devices without additional process complexity [22], [23]. As a result, CTTs are inexpensive to manufacture [1]. In Hf-based oxides, electron trapping is possible due to oxygen vacancies [22]. The formation energy of these O vacancies is low, so many states exist for electron trapping. Trap densities in HfO_2 are estimated in [25] to be as high as $10^{18}/cm^3$. Fig. 2 shows several energy levels of oxygen vacancies in HfO_2 [26].

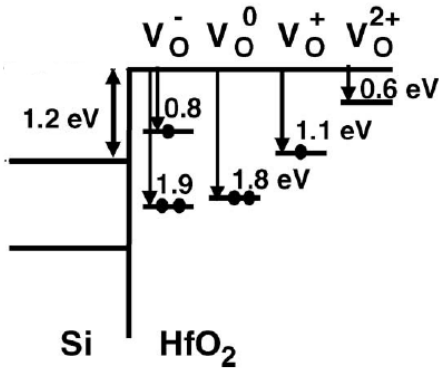


Fig. 2. Oxygen vacancy trap levels in HfO_2 in relation to the conduction and valence bands of silicon. *(From [26].)*

Applying a modest positive bias to the gate can lead to electron trapping through hot-carrier injection. While hot-carrier injection has typically been undesirable in traditional use of transistors due to reliability concerns, the application of a modest positive bias to the drain at the same time as to the gate produces predictable, stable threshold voltage shifts that can be exploited for non-volatile memory applications [27]. The trapped electrons in the gate oxide create a positive V_{th} shift that is large enough to make small fluctuations in threshold voltage insignificant in comparison. Therefore the V_{th} shifts in HfO_2 gates are stable enough on which to establish a sufficient memory window to enable application as a NVM [23], [28]. In addition to enhanced hot carrier injection allowing for electrons to be trapped in the gate oxide, these electrons can also be removed via tunneling by simultaneously applying a modest negative voltage to the gate and positive voltage to the drain. Fig. 3 shows the band diagram as well as the process of electron trapping and detrapping for a SiO_2/HfO_2/poly-Si gate stack [26].

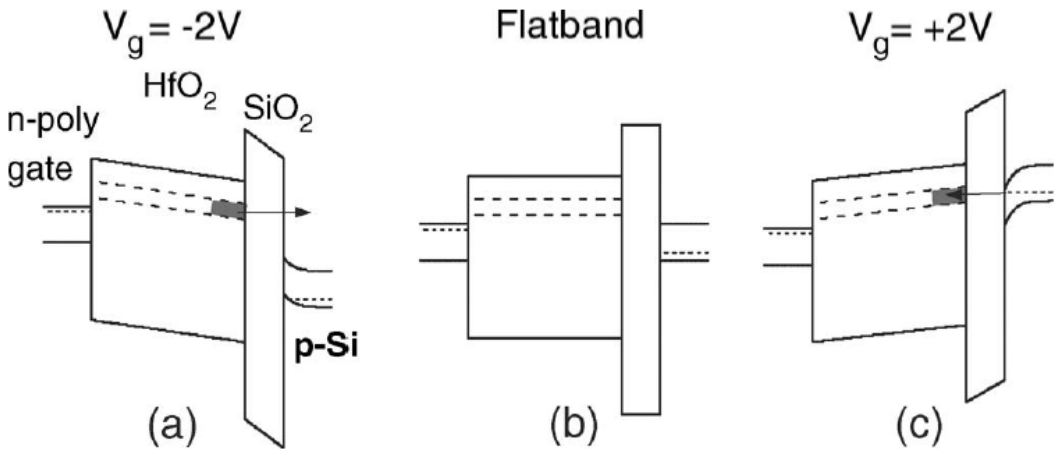


Fig. 3. Band diagrams for SiO_2/HfO_2/poly-Si gate stack under negative gate bias (a), flatband (b), and positive gate bias (c). In (a) and (c), the trapping and detrapping of electrons in the HfO_2 gate oxide as they tunnel though the silicon oxide is shown. *(From [26].)*

CTT-Based Programmable Memories

Several works have explored programmable memories created from both 22 nm FDSOI and 14 nm bulk FinFET CTTs [1], [3], [22]–[24], [27]–[37]. Transistors are programmed via hot-carrier injection under gate and drain bias conditions that favor electron trapping in HfO_2 over interface-trap generation [3], [23], [24], [29], [31], [34]. An 80 kb, multi-time programmable memory was successfully implemented using 22 nm FDSOI CTTs in [22], [30], and it was demonstrated that the same principles would apply to 14 nm bulk FinFET CTTs showing scalability of CTTs. In [23], 64 kb DRAM and SRAM arrays are constructed from HfO_2-based 22 nm SOI devices. The use of 22 nm SOI CTTs for use in unsupervised learning in neuromorphic computing was analyzed in [27]. The creation of both digital and analog memories constructed from CTTs was also studied in-depth in [36].

From the various studies on non-volatile memories constructed from CTTs, it has been established that programming one cell does not affect V_{th} for adjacent cells, cycling between programmed and erased states does not cause an increase in leakage current, and retention of trapped charge is greater than 80% after 10 years [23], [24], [27]. The 14 nm bulk FinFET based devices have been shown to share similar memory transistor properties with the 22 nm FDSOI-based devices [22], [24], [29]–[31]. Scalability to later-generation technology nodes, even down to 7 nm FinFETs, has also been demonstrated [24].

The use of these GlobalFoundries CTTs fabricated in 22 nm FDSOI as analog memories is explored in [28]. The state is stored as trapped charge in the CTT, and the state is read as a current. The different states within the CTT are linearly spaced out with regards to drain current. The targeted drain currents ranged from -800 nA to +800 nA, and the read voltages are V_G = 200 mV and V_D = 50 mV. A twin-CTT cell is used where the state is represented by the difference

between the two devices: $w = w^+ - w^-$. As a result of this twin-cell arrangement, both positive and negative states can be implemented, and the effect of device variation is reduced. When using this twin-cell arrangement, only one of the CTTs is programmed. Since programming shifts the threshold voltage positively (Fig. 1), the drain current decreases with programming. Therefore, the "negative" CTT is programmed to increase the drain current, and the "positive" CTT is programmed to decrease the drain current. Analog CTTs have a lower margin for variability than their purely digital counterparts since the difference between two states is smaller. However, it was found in [28] that even with accounting for variation between devices and programming differences, 5 bits are possible for individual CTTs and 3 bits each for a CTT array.

Total-Ionizing Dose Effects in MOSFETs

When metal-oxide-semiconductor field-effect transistors (MOSFETs) are exposed to ionizing radiation, electron-hole pairs are generated in the oxide as the ionizing particle passes through it [38], [39]. The total absorbed dose due to this ionizing radiation is aptly named total-ionizing dose (TID). TID is an expression of energy absorbed per unit mass, and the common base unit for TID is rads, and the SI unit for dose is the gray (Gr) where 1 Gr = 1 J/kg = 100 rads [38]. Since the target material affects the amount of energy absorbed due to TID being a function of mass, the unit "rad" is usually proceeded by a material designator, for example, rad(SiO_2).

The Physics of TID

While some of the TID-induced electron-hole pairs recombine, a number of the electrons and holes move through the oxide due to the electric field resulting from an applied voltage. Fig. 4 uses a band diagram to demonstrate the physics of TID in the case of an SiO_2 *n*MOS transistor

under positive bias [39]. The generated electrons from the electron-hole pairs are swept by the electric field to the gate and quickly collected. On the other hand, the holes move more slowly and in the opposite direction of the electrons toward the SiO_2/Si interface. Some of these holes become trapped in the oxide near the interface, and these are called oxide traps. However, some holes make it to the SiO_2/Si interface and subsequently interact there resulting in interface traps. Annealing, or recovery of some oxide and interface traps, can occur over time and is accelerated by elevated temperatures [40], [41].

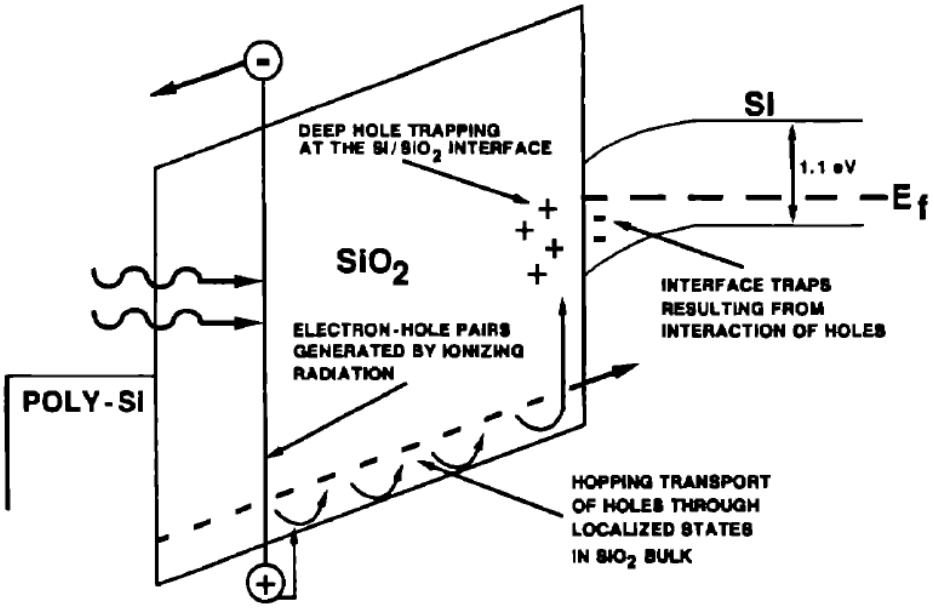


Fig. 4. The physics of TID in an *n*MOS transistor under positive bias. *(From [39].)*

The effect of oxide and interface traps on a transistor can be seen in a semi-log I_D-V_G curve. Fig. 5 shows such curves for a *n*MOS device that is measured before and after irradiation to 300 krad(SiO_2) [42], [43]. The oxide traps cause a shift in the curve to the left. However, the interface traps cause the slope of the linear region of the curve to decrease. If annealing results in the removal of interface or oxide traps, the curve recovers by changing from the post-irradiation curve toward

the pre-irradiation one. The number of traps that anneal is directly related to how much the curve recovers toward the pre-irradiation curve. With respect to the post-irradiation curve, the annealing of interface traps results in a steeper slope while the annealing of oxide traps shifts the curve to the right.

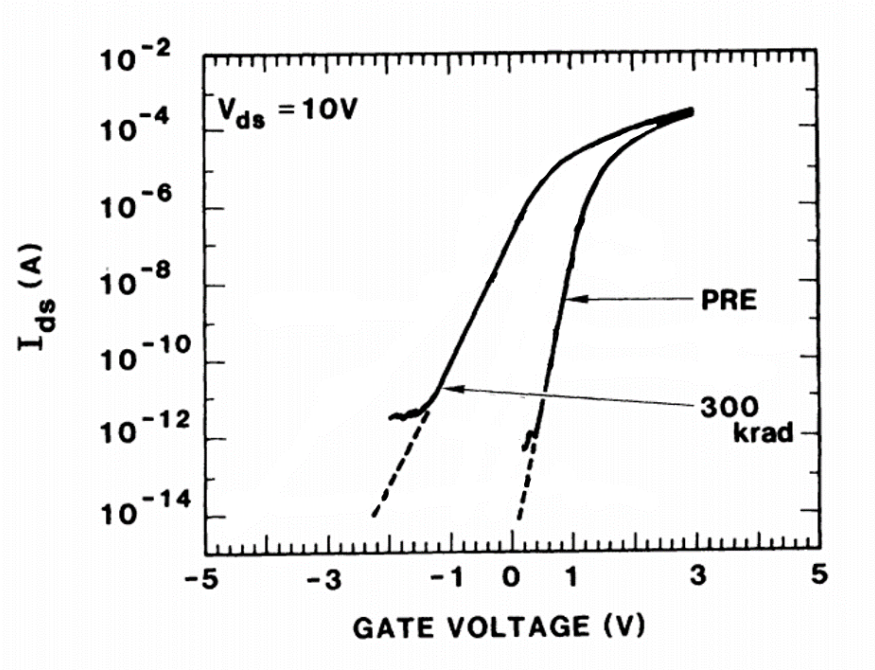


Fig. 5. I_D-V_G response of a transistor before and after irradiation to 300 krad(SiO_2). *(Adapted from [42], [43].)*

TID Response in Various Technologies

A number of studies have been conducted on the total-ionizing-dose response of various transistor technologies. A non-exhaustive list of some of these studies conducted on recent technology nodes include planar bulk transistors [4], [5], [9], planar FDSOI transistors [10]–[12], bulk FinFETs [6]–[8], [13]–[16], and FDSOI FinFETs [13], [14].

Experiments on the TID response of 28 nm planar bulk transistors have been conducted using X-rays [4], [5], [9]. Ultra-high doses of up to 1 Grad(SiO_2) were achieved, and both *n*- and *p*MOSFETs were tested for several dimensions and at various bias conditions. The 28 nm planar

bulk transistors were found in general to be TID-tolerant with very small threshold voltage shifts. However, some TID effects were seen depending on device geometry. Some subthreshold leakage current was seen for the *n*MOSFETS, and long, narrow channel *p*MOSFETs saw threshold voltage shifts after irradiation to very high doses.

Planar FDSOI transistors exposed to X-rays were found to be TID-sensitive with increasing threshold voltage shift with increasing dose [10]–[12]. These threshold voltage shifts are attributed to charge-trapping in the buried oxide of these transistors. Little to no sub-threshold leakage current was seen in these types of transistors.

A number of TID studies have been conducted on bulk FinFETs [6]–[8], [13]–[16], including the 14 nm bulk FinFET technology node [15], [16]. Threshold voltage shifts are minimal in these technologies up to high doses, but some interface trap formation is seen. The TID effects resultant from radiation are attributed to charge-trapping the shallow-trench isolation oxide regions.

Finally, some TID studies have been conducted on FDSOI FinFETs and their results compared to bulk FinFETs [13], [14]. In these works, it was found that the SOI devices exposed to TID had larger threshold voltage shifts than their bulk counterparts. The TID response was attributed to charge-trapping in the buried oxide for the SOI devices and charge-trapping in the shallow-trench isolation for the bulk devices. Changing the fin width resulted in similar responses in both bulk and SOI transistors. Lower temperatures were found to cause more charge-trapping in the BOX of the SOI devices resulting in larger threshold voltage shifts than room temperature.

In general, it was found that SOI devices were more sensitive to TID than their bulk counterparts. This is attributed to the trapping of charge in the buried oxide of these silicon-on-insulator devices. Comparing planar and FinFET devices, the shallow-trench isolation regions in

the FinFETs also contributed to TID effects not seen in planar devices such as increased subthreshold leakage current and interface trap generation.

Non-Volatile Memories

The amount of information that is available is multiplying at a rapid rate. As the amount of known information grows, there is an increasing need both for storage to be more compact and for the data to be processed more efficiently. Currently, memories can be classified as either non-volatile or volatile memories. Volatile memories are very fast, but they require constant power in order to keep the data. The speed of these memories makes them ideal for calculations. On the other hand, non-volatile memories are slower to read and write; however, the data does not require constant power and can be retained for years. Not surprisingly, these are primarily used for long-term data storage.

Floating-Gate Based Memories

A floating gate MOSFET (FG MOSFET) is a non-volatile CMOS-based memory device based on a modification to the structure of a regular MOSFET [44], [45]. In a FG MOSFET, the gate stack is modified (compared to a regular MOSFET) to include a metallic floating gate. This metallic floating gate is isolated from the control gate and channel via insulators as shown in Fig. 6 [45]. The floating gate contains the data in the form of charge. An array of FG MOSFETs can be used to construct memory arrays, such as electronically-erasable programmable read-only memories (EEPROMs) and Flash memories.

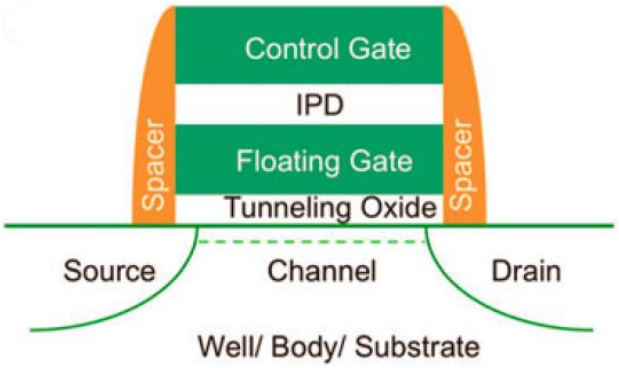


Fig. 6. Cross-section cut of a floating gate MOSFET. *(From [45].)*

Floating-Gate Based Memories: EEPROMs

EEPROMs are a type of memory constructed from FG MOSFETs where individual bytes of data can be modified [44], [45]. Writing to the EEPROM memory is accomplished through use of a high electric field that either deposits or removes electrons in the FG MOSFET gates depending on the bias conditions. Under the high electric field, the electrons transport through the tunnel oxide into or out of the floating gate via Fowler-Nordheim tunneling. In order to read the bit value stored in an FG MOSFET memory cell, a voltage is applied to the gate, and the current between the drain and source is measured. Unlike some memories, reading the data from EEPROMs does not destroy the data. EEPROMs are capable of storing data for many years, and they have high endurance allowing for many cycles of writing. However, since individual bytes can be modified, EEPROMs suffer from very long read/write times and have lower areal density than that of their successor, Flash memories.

Floating-Gate Based Memories: Flash Memories

Flash memories are another type of floating-gate based memory composed of poly-silicon that is very common today [44], [45]. In contrast to their predecessor, the EEPROM, Flash

memories are fast, boasting quick write and erase times due to their ability to erase in groups. Within Flash memories, there are two types: NOR Flash and NAND Flash. NOR Flash is faster than NAND at reading, while NAND Flash is faster at erasing. As a result, NAND Flash is more suited to large data storage and is currently being used in solid state drives (SSDs).

Data is written and erased from the Flash cells by removing or adding electrons via tunneling though the oxide thereby changing the threshold voltage of the device [44], [45]. Flash is inexpensive to manufacture and has higher areal density than EEPROMs. However, Flash is sensitive to total-ionizing-dose. When a Flash cell is irradiated, holes are injected into the floating gate [19]. These holes then recombine with the electrons that were previously injected to establish the state of the Flash memory cell. As a result, data in Flash memories is corruptible by TID.

Although Flash is a valuable memory device, it has some drawbacks even outside of the radiation realm, particularly when it comes to scaling [45]. As the thickness of the oxide has decreased with scaling, it has become harder to retain the stored charge. In addition, as the device dimensions decrease, the coupling between neighboring devices increases. As a result, one cell can affect an adjacent one as shown in Fig. 7. Additionally, there are leakage paths that are able to form in the tunneling oxide (TOX). Combined with oxide defects, low-resistance paths form through which leakage current can flow. In addition, since the total amount of charge capable of being stored in the floating-gate is decreasing with transistor size, charge leakage is also becoming more of a factor due to scaling.

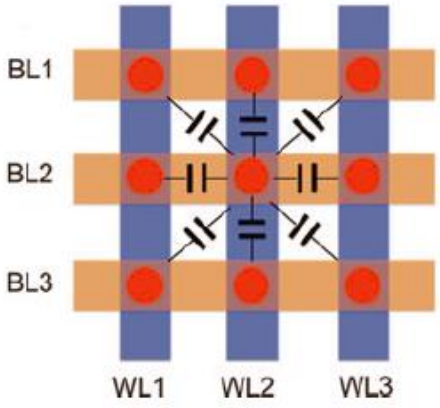


Fig. 7. Coupling between adjacent Flash cells in highly scaled technologies. *(From [45].)*

Charge-Trap Memories

In an attempt to overcome some of the scaling issues with Flash memories, a new memory concept was introduced: charge-trap memories [44], [45]. Over time as it deteriorates, a MOSFET begins to have an unwanted leakage current, and this extra current further degrades the performance of the device. By purposely replacing the floating gate in an FG MOSFET with a material that has atomic defects, a charge-trapping layer (CTL) is created as shown in Fig. 8. Memories created by exploiting the trapping charge properties of the atomic defects in the CTL are a form of aptly-named charge-trap memories (CTMs). While CTTs do not require the addition of a CTL, CTTs are a type of charge-based memories since their operation is determined by the amount of charge trapped in the gate oxide.

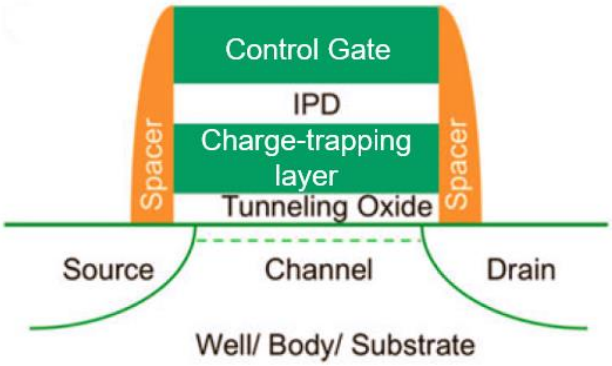


Fig. 8. Transistor structure similar to a FG MOSFET that replaces the floating gate with a CTL. *(From [45].)*

One type of CTL is based on silicon nitride. Electrons are able to tunnel from the channel through the thin silicon dioxide layers into the many electron traps located in the silicon nitride CTL under an electric field produced by an applied bias [44]. The electrons as a result of the traps are able to be stored in a localized place, therefore allowing for the possibility of multiple bits to be stored on a single transistor. In the case of two separate bits, threshold voltage readings are taken on both sides of the transistor. Fig. 9 shows an example of the threshold voltage readings and how they correlate to the values of the bits stored on that particular transistor [44].

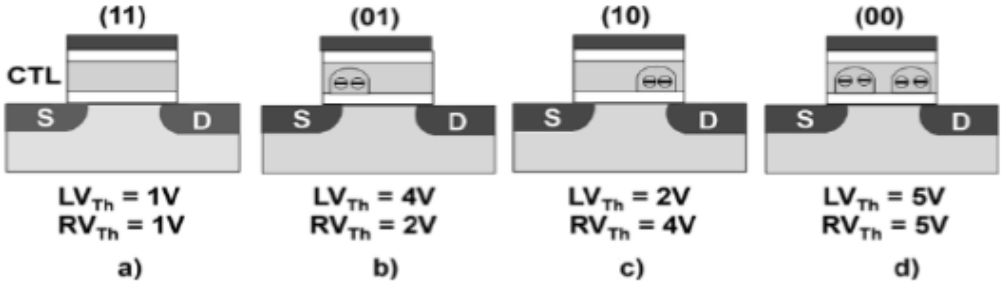


Fig. 9. The establishment of two bits in a single transistor due to the possibility of localized electron trapping in a CTL. *(From [44].)*

SONOS Memories

A common type of charge-trap memory that pre-dates CTTs is SONOS, and the CTL in SONOS is comprised of silicon nitride (commonly Si_3N_4) [44], [45]. Whereas CTTs do not require additional manufacturing steps or process complexity for CMOS implementation, SONOS necessitates the addition of the charge trapping layer during fabrication. Therefore, since SONOS is not a CMOS process and requires additional manufacturing steps, the fabrication of SONOS transistors is more expensive than CTTs.

Compared to Flash memories, a major advantage of SONOS is their charge trapping density which can be as high as 10^{19} or 10^{20} cm^{-3} [46]. Flash data is stored in a sea of charge within the floating gate while SONOS structures contain the individual charges in traps located throughout the entire CTL. As a result, one defect in the SONOS structure is less likely to cause complete charge removal. In SONOS, only a small percentage of the total charge would be lost from the cell from a single defect while Flash could have instant loss of charge due to a defect in the tunneling oxide [45]. Because of the reduced likelihood of complete charge loss in SONOS, SONOS can withstand a higher accumulation of defects than floating-gates before failure. Therefore, SONOS is more resistant to total ionizing dose effects than Flash [46].

SONOS is not completely immune to TID however. TID still causes deposition of positive charge in the CTL which reduces the charge in programmed cells and increases the charge in erased cells [47]. While SONOS structures provide some clear advantages to Flash memory, they are much more expensive to produce making them better suited for niche markets than general consumption. Additionally, retention in these SONOS is not ideal due to silicon-rich nitrides that compose the CTL [45].

Crossbar Architectures

As increasingly larger amounts of data are being collected and analyzed in all areas of life, it has become necessary to find ways to process these massive amounts of data. As previously mentioned, it has become increasingly important to use data analysis methods that are efficient while also being able to handle large and complicated data sets. A promising technique is through use of neural networks, and some promising neural networks have been created using crossbar architectures [48]. Research is currently being conducted on crossbar arrays and how they might be integrated with current CMOS technology to advance the next generation of computing technology [49], [50].

In a neuromorphic computing architecture, crossbar arrays are used for the synapse portion of the neural network, discussed previously in the neuromorphic computing section of this chapter. A crossbar array consists of m word line inputs and n bit line inputs where each possible bit line and word line combination is connected through a weight W_{ij} [51]. The simplest implementation of this crossbar weight is through a transistor/resistor pair (1T1R). This transistor/resistor pair is implemented as the weight corresponding to the particular word line and bit line, and this weight is stored as a conductance using the 1T1R. A diagram of a single layer neural network in the 1T1R architecture along with its circuitry implementation is shown in Fig. 10.

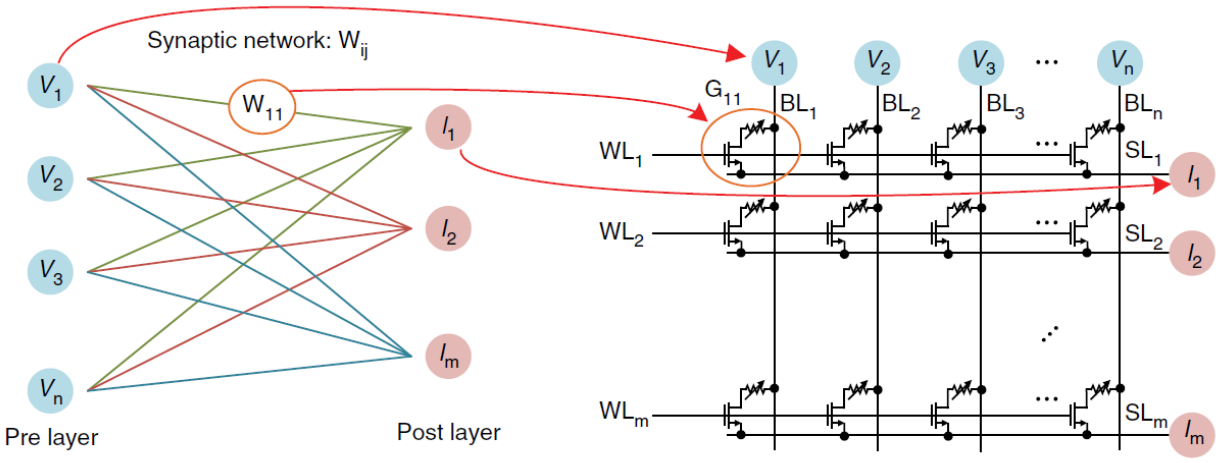


Fig. 10. Mapping from a single layer neural network (left) to the circuit-level crossbar array (right) in the 1T1R architecture. *(From [51].)*

It is possible to either have set weight values that are unchanged or to have weights that can be updated during training. The particular application for which the neural network is meant for determines which is used. Regardless of which method is selected, the weights must be written at least once. This is accomplished through the word lines and bit lines. The inputs to the crossbar are voltage inputs to the *n* bit lines, while the *m* outputs from the crossbar array are currents. The outputs come from the additive currents resulting from sources of the transistors in the 1T1R pair. This structure parallels a single-layer neural network which has *n* voltage inputs ($V_1, \ldots, V_n$) and *m* current outputs ($I_1, \ldots, I_m$). Each input is connected to each output through a weight (W_{ij} for V_i and I_j) specific to the individual connections.

To translate from crossbar architectures to a general neural network, the inputs become the axons, the weights become synapses, and the outputs become neurons [49]. Fig. 10 shows the translation between a single layer neural network and the crossbar array. Each output *I* is the

weighted sum of the product of the input voltages V and conductances W. The mathematical equation governing this process is a simple summation

$$I_j(n) = \sum_{i=1}^{n} W_{ij} V_i(n).$$

The weight W_{ij} in this equation is the conductance G_{ij} for a 1T1R synapses, or in the case of a differential pair at the ij node, the weight is the difference in the conductances ($W_{ij} = G_{ij}^{+} - G_{ij}^{-}$) [52]. This equation requires a vector-matrix multiply which can be expensive in terms of power consumption when implemented in circuitry [53].

One way to reduce the energy consumed by the vector-matrix multiply is to implement the crossbar architecture using analog components, such as a CTT. As a by-product, this also allows for a wider range of bit values to be stored. An example of implementing analog components in the architecture would be to encode the word line and bit line voltage inputs in time and voltage height respectively, and an example of such an implementation is shown in Fig. 11. As a result, the "strength" of the weight is governed by the voltage height of the bit line. Additionally, the weight can only be written to when the word line value is non-zero [53].

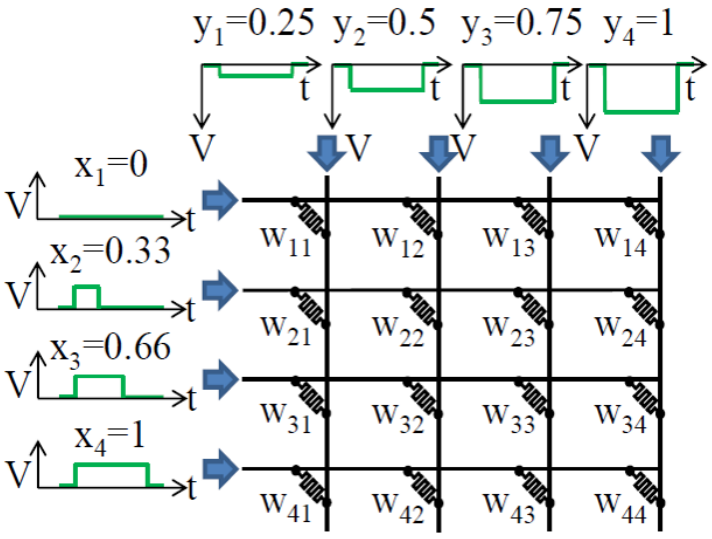


Fig. 11. Parallel write with x and y encoded in time and height respectively. *(From [53].)*

A major attraction for using crossbar architectures is the simplicity of it. However, there are currently still disadvantages to this structure that must be addressed before wide-spread usage; one such disadvantage is power consumption as discussed previously. In addition, parasitic leakage currents can also be a concern. Popular devices to use for crossbar implementation are a resistive memory element, and a simple implementation of a crossbar architecture uses a 1T1R pair. However, an issue faced by these resistive memories is very large resistance requirements which in turn results in large parasitic currents. Then as the dimensions of a crossbar array increase, these parasitic currents accumulate and reach unacceptable levels [54].

Therefore, for efficient use of the crossbar array, it is likely that at least some components of the crossbar will need to be constructed in novel technologies. One possible implementation would be to use novel, two-terminal devices as the weights, but continue to use traditional CMOS for the surrounding circuitry [49]. Several works focus on the individual novel technologies that could possibly be used as the weights in a crossbar architecture. A few novel technologies analyzed include memristors [48], [50], carbon nanotube FETs [49], [54], and spin transfer torque (STT) magnetic RAM (MRAM) [55]. Some of the aspects examined in these novel technologies include optimization [50], [56], power consumption, timing [50], defect and fault tolerance [54], and transistor density [55]. CTTs have been successfully implemented in prototype crossbar arrays, demonstrating the feasibility of using CTTs in crossbar arrays for neuromorphic computing [32]. In addition, since CTTs are a commercial CMOS process, it is possible to manufacture the entire crossbar structure including the peripheral circuity at the same time thereby making it inexpensive.

Use of crossbar architectures in neuromorphic computing and neural networks is still an emerging research area. While some companies have specialized hardware that implements crossbar architectures for use in neural networks, crossbar architectures have not yet seen

widespread adoption. However, foundational research into crossbar architectures has begun. Exploratory research has been conducted into important aspects including how to operate them, possible technologies to use with them, and potential issues that may arise with use of these architectures. The hope is that this foundational research paves the way for an eventual new, robust, efficient architecture for use in neural networks and neuromorphic computing.

Neural Networks Implemented with CTTs

A few works have investigated the potential for CTTs to be implemented in neural networks [27], [28], [32], [33]. The first work on the subject explored the use of CTTs as analog synapses (weights) for unsupervised learning in a neural network [27]. The devices used were 22 nm planar SOI CTTs with a hafnium-oxide based gate dielectric. A simple, single-layer, winner-take-all neural network was simulated based on CTT experimental data collected on individual CTTs. This proof-of-concept neural network was simulated using 27 CTTs as the weights in the network. For this neural network, there are three possible output classifications based on nine input neurons. Each of the neurons could take on one of two states representing a pixel that is either black or white. Fig. 12 shows the possible inputs to the neural network. The left-most column is the ideal classification while the other nine inputs in a given row are single-pixel modifications of the left-most inputs. Fig. 13 shows the fully connected, single-layer neural network. The CTTs in the simulation correspond to the weights (shown as lines) connecting each input to each output.

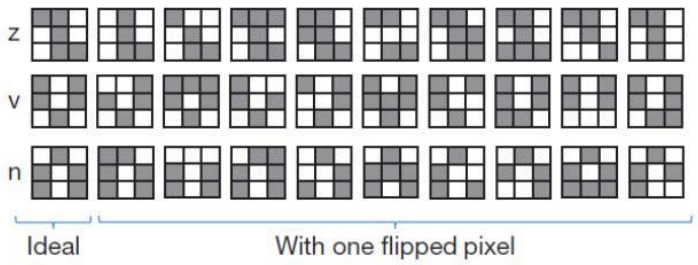


Fig. 12. The thirty possible 3x3 pixel inputs to the neural network where the left-most column corresponds to the desired classification. *(From [27].)*

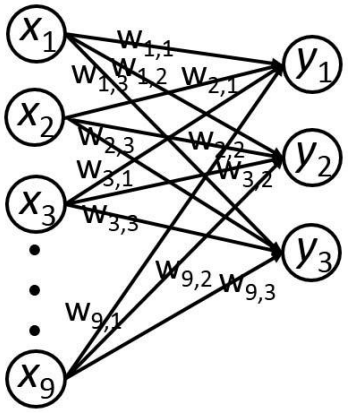


Fig. 13. Fully-connected single layer neural network with 9 inputs and 3 outputs. *(From [27].)*

The CTTs were initially programmed to a random value, and subsequent training was conducted by alternating between presenting one of the three ideal inputs (left-most column of Fig. 12) and then presenting a single-bit flip version of this same input. This process was repeated until 100% classification accuracy was achieved, and this occurred on average after 24 training cycles over 10,000 simulations. This demonstrated that it is indeed possible to use CTTs for unsupervised learning in a neural network.

Following this initial study [27] demonstrating proof-of-concept, two other studies followed expanding on the results of the first one utilizing more complex networks [28], [32]. Using the experimental data of [27], [32] simulated the use of a 784 x 784 array of CTTs as weights

in a neural network to classify handwritten digits from the MNIST database [57]. Over 95% classification accuracy was achieved using a four layer, fully-connected analog neural network with two hidden layers. This demonstrates that CTTs can be used in more complex neural networks involving practical applications such as that of handwritten digit classification.

The use of CTTs as analog memories for neural networks has been explored in [28]. The focus of this work is the programming accuracy and stability of CTTs with application as analog synapses in mind. In neural networks sometimes a twin-cell architecture, where the state is determined by the difference between the two weights (Fig. 14), is implemented to allow for bipolar states and reduce the effects of device-to-device variation. In this work, five arrays, each constructed with forty sets of twin-CTT cells, were programmed and tested. It was found that fairly accurate analog values could be programmed to these CTTs in an array. Accounting for programming variation, drift, and device variation, a total of ten separate digital values, or more than three distinct bits, were able to be programmed in the array of CTTs. These results demonstrate that CTTs can be used in neural networks as analog weights.

Finally, [33] looks at the fault tolerance of CTTs implemented as multi-level cell (MLC) weights in neural networks. An MLC application increases the likelihood of faults compared to a single-level cell (SLC) due to a decreased distance between states. However, in this work, it was found that careful design and implementation of MLC resulted in no discernable loss in accuracy in the neural network during fault injection. In addition, the resulting MLC-CTT neural network implementation reduced the area requirement by an order of magnitude compared to an SRAM neural network implementation.

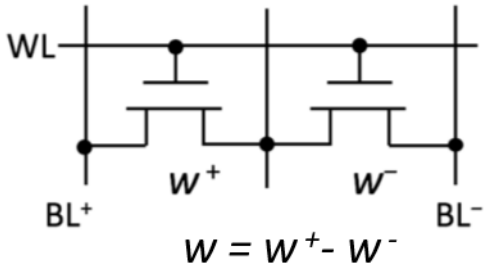


Fig. 14. Twin-cell structure for a neural network weight where the resulting output weight is the difference between the weights of the two transistors, $w = w^+ - w^-$. *(From [28].)*

CHAPTER III

EXPERIMENTAL SETUP

Devices

The CTTs used in this work were *n*MOS devices fabricated by GlobalFoundries in their 22FDX® and 14LPP process-design kits. The 22FDX® CTTs are 22 nm fully-depleted silicon-on-insulator (FDSOI) MOSFETs, and the 14LPP devices are low power, 14 nm bulk FinFET MOSFETs [20], [21]. Fig. 15 shows cross-sections of both planar FDSOI and bulk FinFET transistors, which corresponds to the structure of 22 nm FDSOI and 14 nm bulk FinFET CTTs respectively. The CTTs have three modes: initial, programmed, and erased. The initial state is a one-time, pre-programmed state (virgin). After the first programming, the multi-time programmable CTT can be cycled between the programmed and erased states by trapping and de-trapping electrons in the gate oxide. This cycling process is shown in Fig. 16.

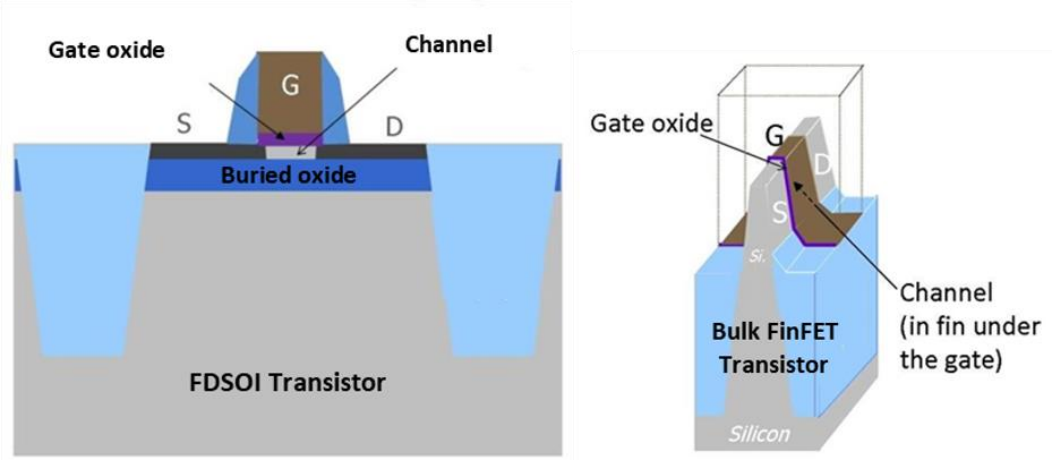


Fig. 15. Cross-section showing the structure of a planar FDSOI transistor (left) and a bulk FinFET transistor (right). *(Adapted from [58].)*

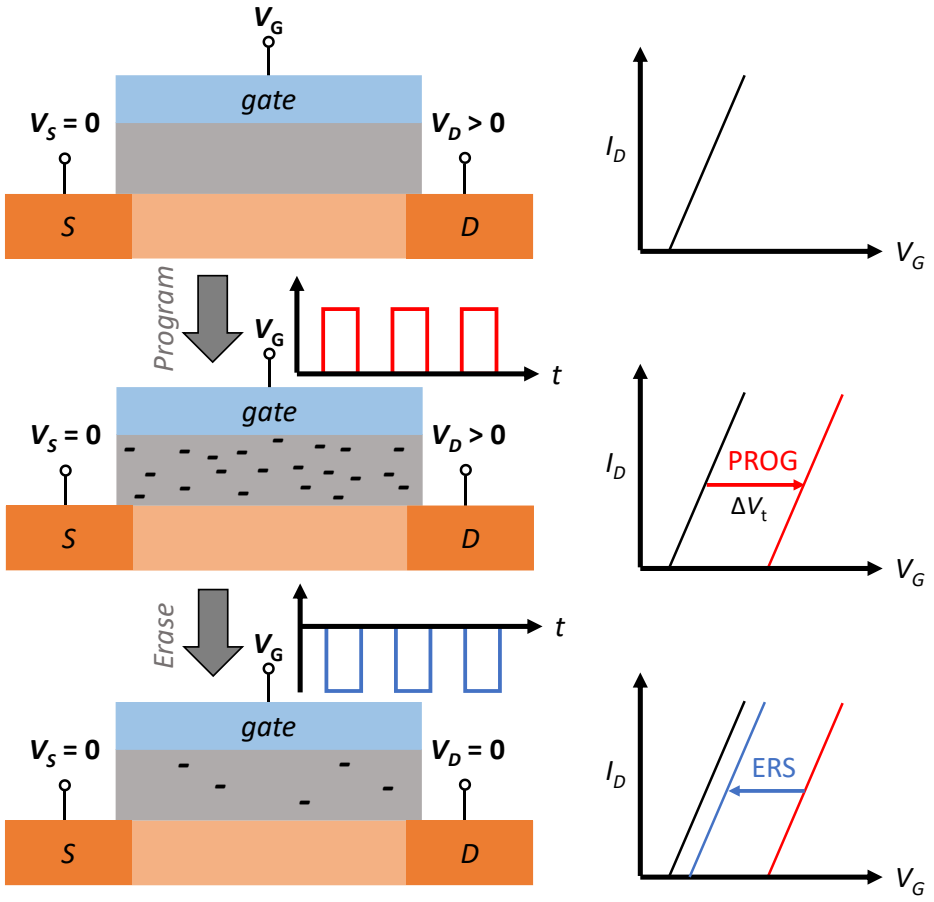


Fig. 16. The initial, virgin state (top), the programmed state after positive gate pulses (middle), and the erased state after negative gate pulses (bottom). *(Adapted from [1].)*

Mounting and Bonding

The dies containing the CTTs were mounted on custom high-speed packages [59]. Fig. 17 shows a photomicrograph of a die containing two rows of 22 nm FDSOI CTTs (boxed in red.) Each row contains 25 pinouts and seven total devices per row, where each CTTs varies from the others in the row in length (or width) while the width (or length) stays constant for the row. The seven CTTs in a row all share the *n*-well, *p*-well, and substrate contacts while each device has separate drains, sources, and gates. Table I shows the pinouts for these 22 nm FDSOI devices, where pad 1 corresponds to the far left pad in a row in Fig. 17.

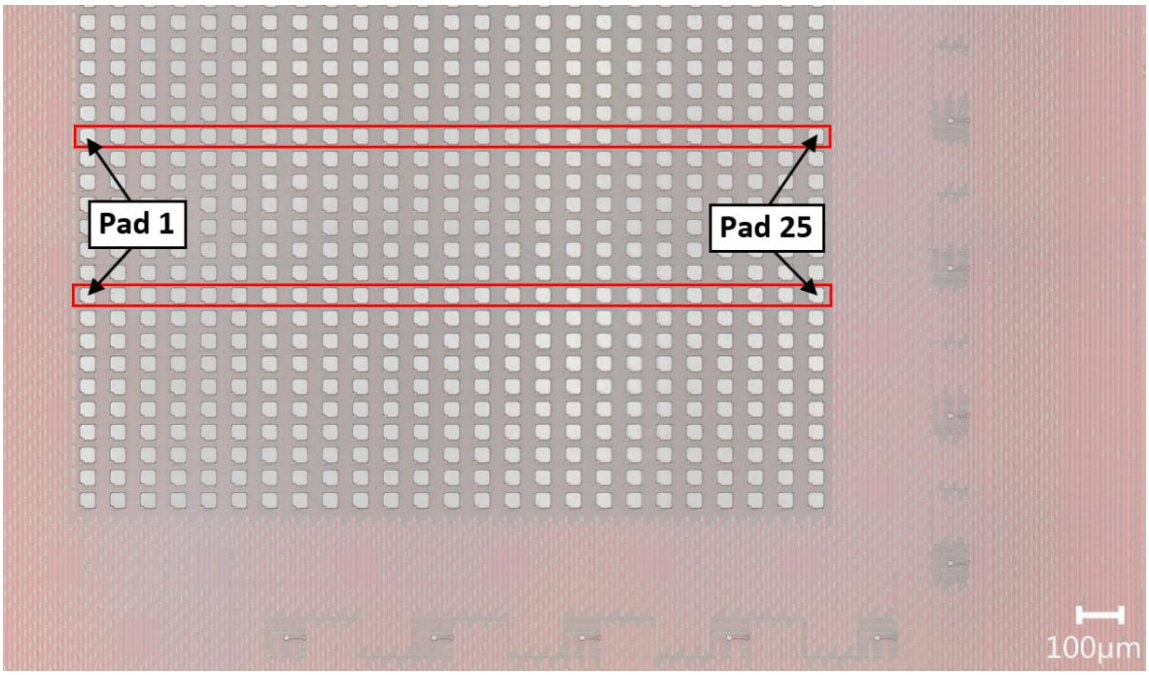


Fig. 17. Photomicrograph of a die containing two rows of 22 nm FDSOI CTTs with seven CTTs per row.

Table I. Pinout diagram for the 22 nm FDSOI CTTs. The substrate, *n*-well, and *p*-well are shared between the seven devices, while the drains, gates, and sources are separate. Pads 1 through 25 correspond from left to right in Fig. 17.

22 nm Device	**Pad Number**					
	Drain	**Gate**	**Source**	**Substrate**	***N*-Well**	***P*-Well**
1	2	3	1	25	24	23
2	5	6	4	25	24	23
3	8	9	7	25	24	23
4	11	12	10	25	24	23
5	14	15	13	25	24	23
6	17	18	16	25	24	23
7	20	21	19	25	24	23

The 14 nm bulk FinFETs had similar-looking dies with slightly smaller pitch between pads, but different pinouts from the 22 nm CTTs. The 14 nm CTTs had 14 devices per row, no pads for the *n*-well and *p*-well, shared gates, and drains shared between pairs of CTTs. Table II shows the pinouts for the 14 nm bulk FinFET CTTs. Each 14 nm CTT has two gates due to being a multi-

fingered device, and these two gates are shared throughout the whole row of 14 devices. During testing, the two gates are connected to the same potential, therefore during the bonding process they could be connected to the same output on the high-speed package. Table II also shows that a single drain pad is shared between two devices. In order to test two devices with a single drain separately, the source from one device was grounded while the other was left floating. To avoid this issue entirely, only one device from a shared drain was bonded out at a time.

Table II. Pinout diagram for the 14 nm bulk FinFET CTTs.

14 nm Device	**Pad Number**				
	Drain	**Gate**	**Gate2**	**Source**	**Substrate**
1	2	23	24	1	25
2	2	24	23	3	25
3	5	23	24	4	25
4	5	24	23	6	25
5	8	23	24	7	25
6	8	24	23	9	25
7	11	23	24	10	25
8	11	24	23	12	25
9	14	23	24	13	25
10	14	24	23	15	25
11	17	23	24	16	25
12	17	24	23	18	25
13	20	23	24	19	25
14	20	24	23	21	25

After mounting to the high-speed package, the pads of the transistors are attached to the outputs of the package via thin gold wires using a ball bonder. Fig. 18 shows a picture taken through a microscope of Device 3 from Table I that is bonded out to the high-speed package, and Fig. 19 shows the bonded out die mounted to the middle of the high-speed package. At minimum for each device, each drain and gate must be bonded to separate outputs. In addition, the sources

of the devices, shared substrate, *n*-well, and *p*-well are either connected to ground on the high-speed package or bonded to an output that is grounded during testing. Since each high-speed package contains between 6-8 outputs, up to four devices can be bonded out and tested at a time. However, typically two or three devices were bonded out and tested at a time. I_D-V_G sweeps were conducted on the CTTs using a HP 4156 Parametric Analyzer, and the parameter analyzer had four inputs therefore limiting the number of devices that could be tested simultaneously. Three devices could be tested simultaneously if their three gates were connected together to one input of the parameter analyzer. Then the remaining three inputs to the parameter analyzer could be used for the three drains of the devices. Typical settings on the ball bonder during the bonding process were as follows: ball 1.0, initial power 1.09, final power 1.52, loop 8.9, time 9.9, force 1.0, step 0.0, and tail 6.0.

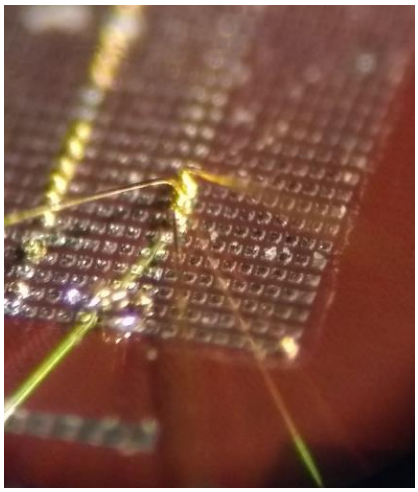


Fig. 18. Photo taken through a microscope of the gate, source, and drain of a CTT that is bonded out to the high-speed package (not visible in this picture.)

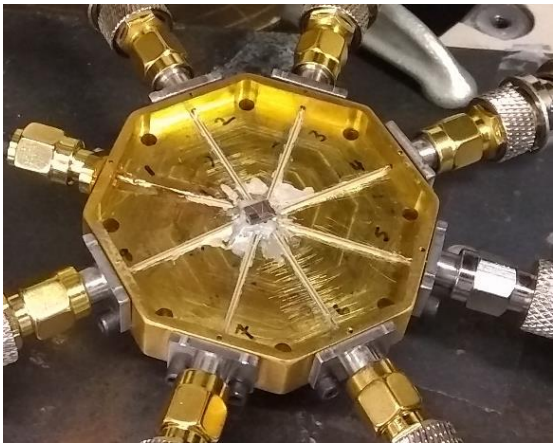


Fig. 19. High-speed package (gold-colored octagon) with the die mounted at the center.

Programming and Erasing

Programming and erasing are accomplished using pulsed-voltage ramped stress [2], [3]. Stressing is done by applying high gate-voltage V_G and drain-voltage V_D pulses; sensing is performed at lower V_G and V_D values. The degree of programming is determined primarily by the strength of the gate electric field. Retention and stability of the V_{th} shift depends primarily on drain voltage [3]. For programming in this work, V_D is set at 1.2 V, pulse times are 10 ms, and the peak V_G is set initially at 1.4 V and incremented in magnitude in a series of 39 pulses until reaching a maximum V_G of 2.7 V for the 22 nm FD SOI devices, and 27 pulses until reaching a maximum of 2.2 V for 14 nm bulk FinFETs. Fig. 20 shows the voltage versus time relationship for the pulsed-voltage ramped stress applied to these devices. The erase pulses for both the 22 nm FDSOI and 14 nm bulk FinFET CTTs consist of 51 V_G pulses starting at -2.4 V, increasing in magnitude to -3.1 V. The drain is grounded during the erase pulses. For sensing pulses following programming or erasing, V_G is 0.6 V and V_D is 0.1 V. The sensing time is 50 ms per cycle.

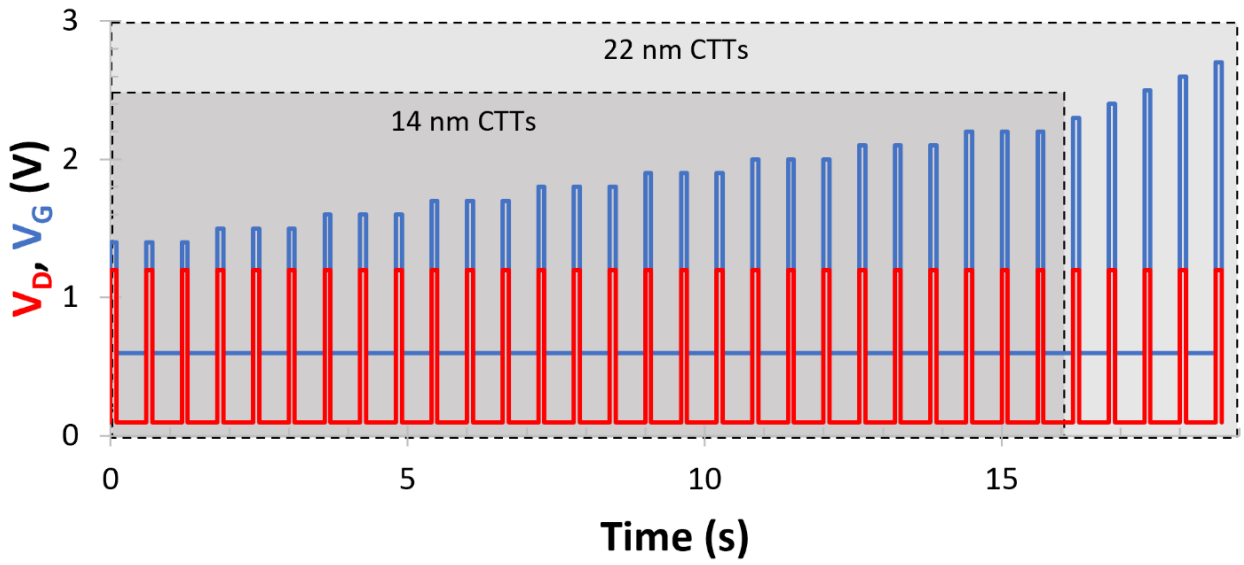


Fig. 20. Pulsed-voltage ramped stress applied to both the 22 nm FDSOI CTTs and 14 nm bulk FinFET CTTs. The inner box shows the pulses applied to the 14 nm CTTs while the outer box shows the additional pulses applied to the 22 nm CTTs. The peak gate voltage (blue) increases in magnitude until 2.7 V for the 22 nm CTTs and 2.2 V for the 14 nm CTTs. The peak drain voltage (red) is 1.2 V during pulsing for both types of CTTs.

Irradiation

Devices were irradiated up to 500 krad(SiO_2) with ~10 keV X-rays at a dose rate of ~30 krad(SiO_2)/min using an ARACOR Model 4100 X-ray Irradiator [60], [61]. A picture of the irradiation test setup is shown in Fig. 21 where the CTTs are mounted in the center of the high-speed package which is located directly in-line with the X-ray beam line. The peak energy is ~10 keV, with 95% uniformity for a 3 cm diameter beam [60]. All device pins were grounded during irradiation, typical for non-volatile memory applications [17]–[19]. All irradiation and annealing steps were performed at room temperature. Before irradiation and after each dose increment, I_D-V_G sweeps at V_D = 50 mV were performed using the medium integration time setting of a HP 4156 Parametric Analyzer. For consistent results, exactly one minute of annealing time

was taken after irradiation before the I_D-V_G sweep was conducted. At least two (typically three or more) devices were tested at each condition. Results for devices programmed and/or irradiated under similar conditions vary by less than ±10 %.

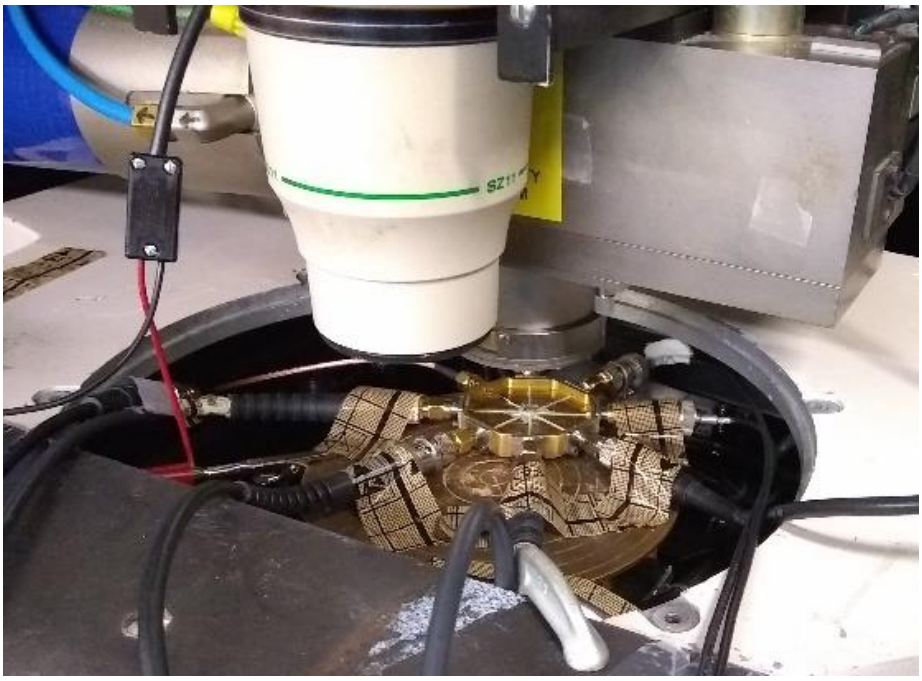


Fig. 21. Experimental test setup for X-ray irradiation. The CTTs to be irradiated are located in the center of the high-speed package directly underneath the beam line (gray). The package was secured with ESD-safe tape.

Data Analysis

A script written in Python was used to extract the relevant parameters from the I_D-V_G curves resulting from the Parameter Analyzer. Both the threshold voltage (V_{th}) and the transconductance (G_m) were of interest and collected for the various I_D-V_G curves. For a given I_D-V_G curve, a five-point stencil was used to find the maximum slope over five consecutive data points. The window of five data points was used instead of the maximum between any two consecutive points to reduce the effect of noisy data. The slope found then is the transconductance (G_m). This transconductance

was then used at with the center data point from the five-point stencil to find V_{th} by calculating the x-intercept and subtracting $V_D/2$. V_D in this work was always 50 mV, and the resulting equation for finding V_{th} is

$$V_{th} = V_G(G_m) - \frac{I_D(G_m)}{G_m} - 0.025\ V.$$

Fig. 22 shows an example I_D-V_G curve with the five point window of maximum slope, and Fig. 23 shows a closer view showing the location of G_m, the x-intercept, and V_{th}. After calculating the threshold voltages and transconductances for all the I_D-V_G curves in a data set, the changes in threshold voltage or transconductance compared to the pre-irradiation, virgin devices were calculated. Since everything was with respect to the pre-irradiation curves, ΔV_{th} values greater than zero indicate a positive threshold voltage shift, or a shift of the I_D-V_G curve to the right compared to the pre-irradiation device. Similarly, ΔV_{th} values less than zero indicate negative threshold voltage shifts and a shift of the curve to the left compared to the initial state of the device.

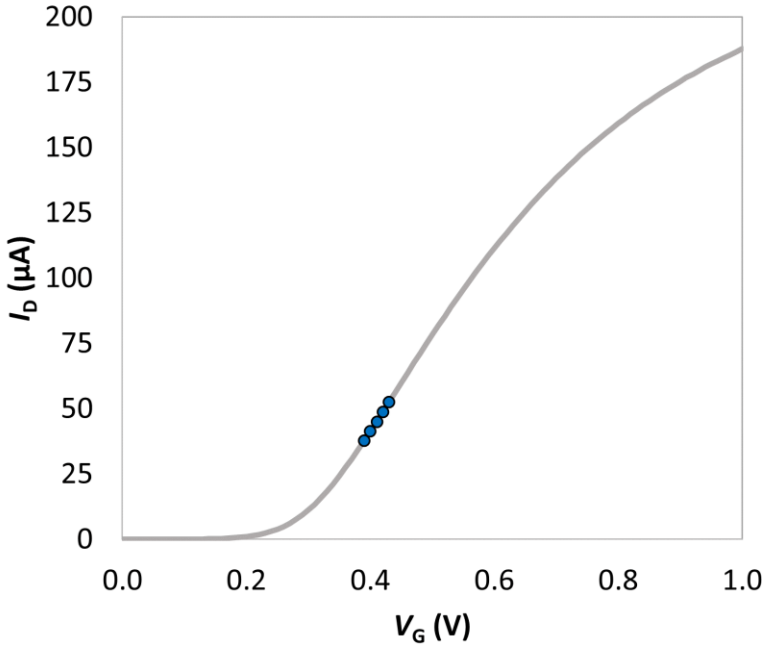


Fig. 22. Location of maximum slope using a five-point stencil where the five points are denoted by markers.

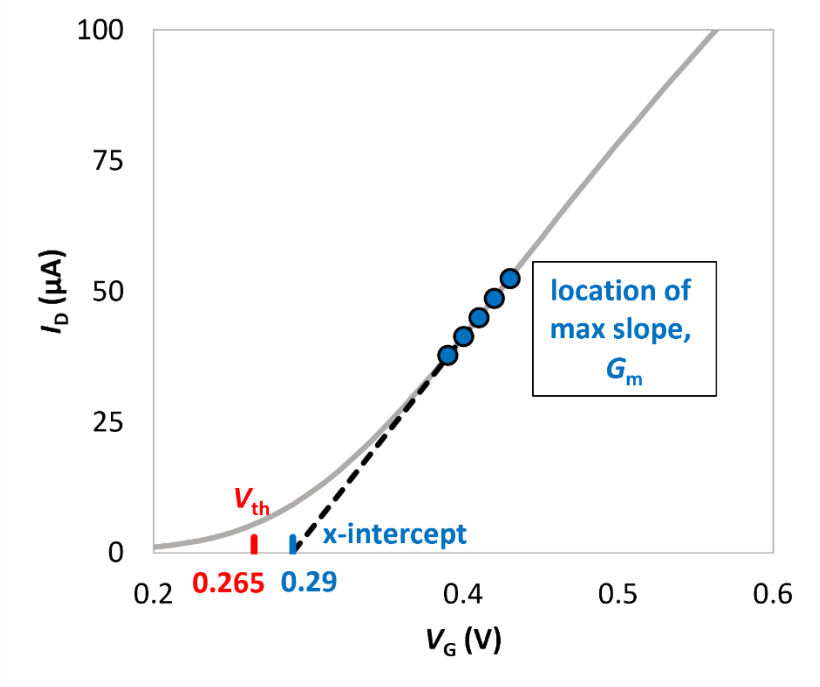


Fig. 23. Zoomed in version of Fig. 22 where the locations of G_m, the x-intercept, and calculated V_{th} are denoted.

Experimental Sequence

Both the 22 nm FDSOI and 14 nm bulk FinFET charge-trap transistors are tested in two sequences as shown in Fig. 24. Some CTTs are programmed before irradiation while others are programmed after irradiation. The two sequences are then compared to determine whether irradiation effects the programmability of the devices. It is found that the memory and TID effects are largely independent for the 14 nm bulk FinFET CTTs. However, the 22 nm FDSOI CTTs saw higher threshold voltage shifts when programmed after irradiation, and this is attributed to neutralization of radiation-induced trapped holes in the buried oxide.

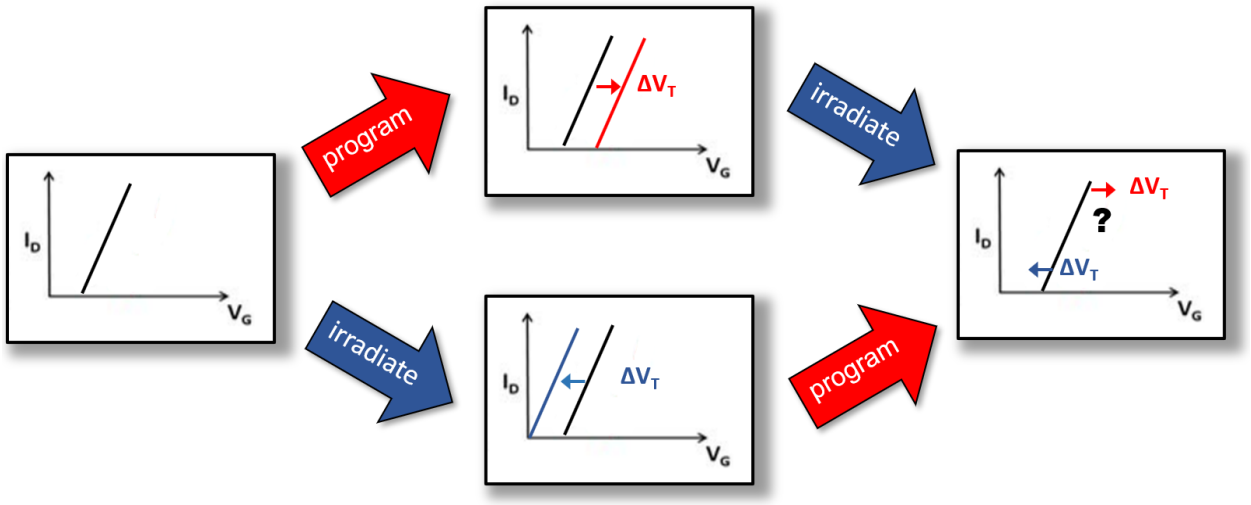


Fig. 24. Programming sequence for both the 22 nm FDSOI and 14 nm bulk FinFET CTTs. Some devices were programmed and then irradiated while others were irradiated and then programmed. These two sequences were then compared to determine the effect of irradiation on the programmability of the CTTs.

CHAPTER IV

TID RESULTS ON 22 NM FDSOI CTTS

The experimental data and results from TID irradiation on the 22 nm FDSOI CTTs are presented here. The majority of this chapter has been published in [62].

Programming Before Irradiation

Fig. 25(a) and Fig. 25(b) show typical I_D-V_G curves for single-finger, 22 nm FDSOI devices of width W = 120 nm and length L = 20 nm programmed before irradiation. Fig. 26 shows extracted ΔV_{th} data and error bars for several devices tested in the same manner. Programming the device results in a threshold voltage shift (ΔV_{th}) of ~100 mV. The nearly parallel shifts in I_D-V_G curves show that this shift is caused primarily by electron trapping, and not interface traps [1]–[3], [61]. When the device is irradiated to 50 krad(SiO_2), V_{th} shifts by about −40 mV from its programmed value. By ~150 krad(SiO_2), V_{th} has returned to its initial value, before programming, so any potential memory window would be fully closed. Irradiating devices to 500 krad(SiO_2) leads to a further V_{th} shift of approximately −50 mV. To assist in the evaluation of trapping mechanisms, the device was annealed at room temperature (RT) and reprogrammed after irradiation without performing an erase cycle (Fig. 25(b)). The second programming results in a positive V_{th} shift of only ~45 mV, leaving V_{th} still below its initial value. Hence, the programming sequence is insufficient to re-establish a usable memory window in a potential NVM operation.

Fig. 27(a) and Fig. 27(b) show typical I_D-V_G curves for 22 nm FDSOI devices of width W = 300 nm and length L = 20 nm, programmed before irradiation. Fig. 28 shows extracted ΔV_{th} data

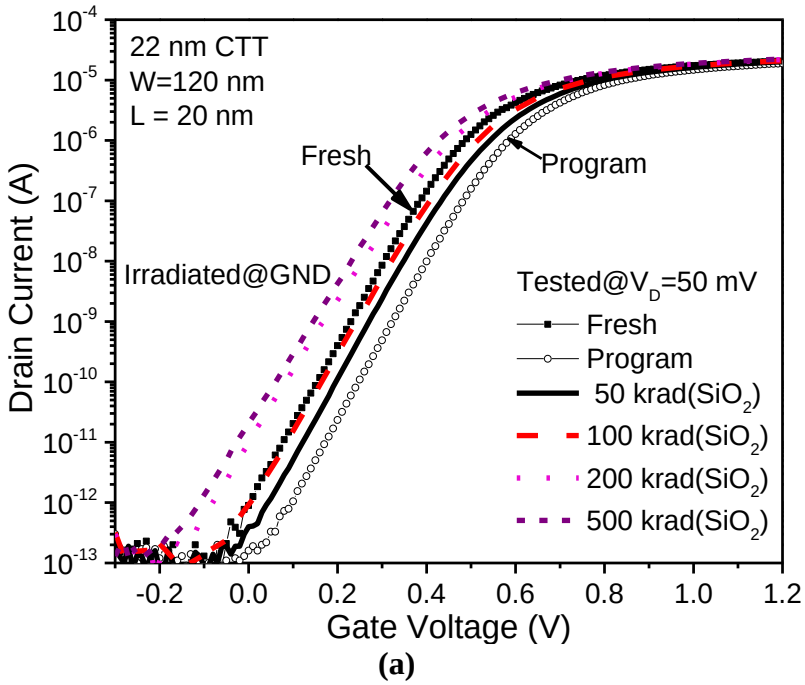


Fig. 25. Typical I_D-V_G curves for 22 nm FDSOI CTT devices with channel width of 120 nm after programming, (a) irradiation to 500 krad(SiO_2), and (b) 30 minutes of room temperature (RT) annealing, reprogramming, and additional annealing. “Fresh” indicates the response of as-processed devices that were not programmed or erased.

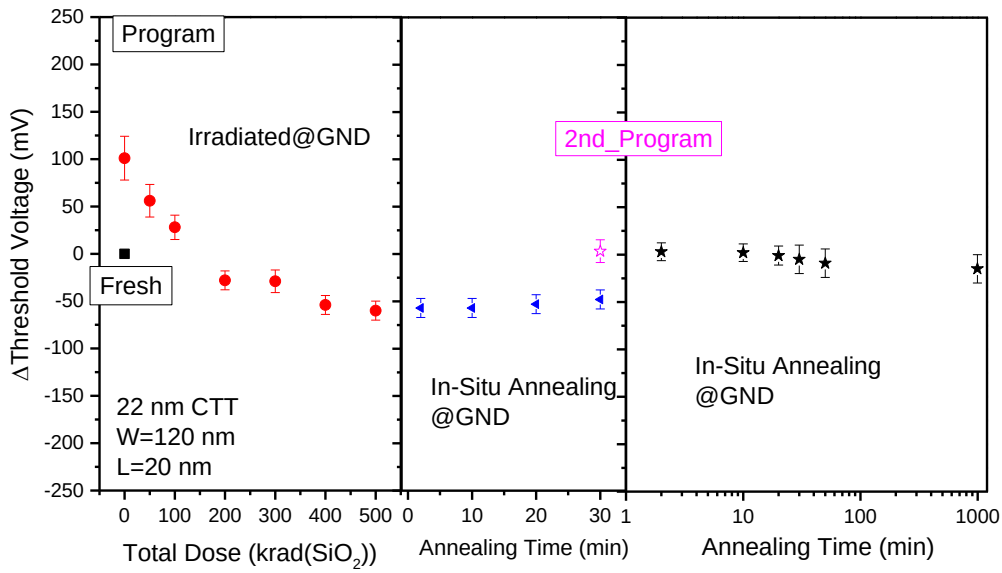
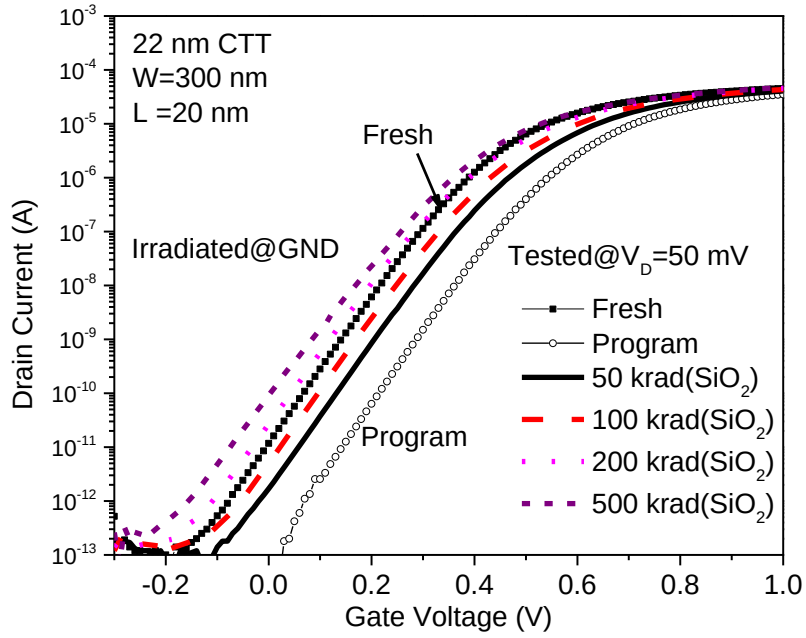


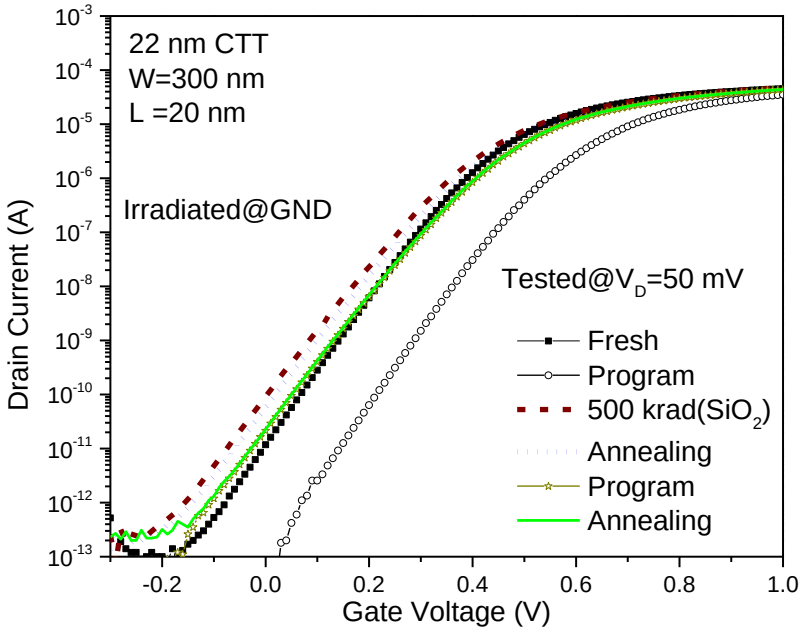
Fig. 26. Extracted threshold voltage shifts with error bars denoting standard deviation for three 22 nm FDSOI CTT devices programmed and irradiated using sequences similar to those shown in Fig. 25.

and error bars for several devices tested in the same manner. Programming these devices results in V_{th} shifts of ~170 mV, due to the additional electron trapping that occurs during programming in wider devices [3]. For the devices in Fig. 27 and Fig. 28, V_{th} does not return to its initial value, before programming, until a dose of ~400 krad(SiO_2) at which point the amount of charge trapped in the gate oxide and buried oxide cause equal and opposite direction V_{th} shifts. Moreover, in contrast to the narrower devices of Fig. 25 and Fig. 26, a second programming pulse is able to restore the value of ΔV_{th} to ~60% of its initial, programmed value. Thus, if memory density requirements allow the use of wider devices, doing so may significantly facilitate their radiation tolerance. However, because CTTs are of most interest in more highly scaled device applications [1]–[3], the focus is primarily on the responses of the 120 nm devices in this work.

The smaller V_{th} shifts produced by post-irradiation programming than before irradiation in Fig. 26 and Fig. 28 strongly suggest that (1) the first programming sequence efficiently fills most available electron traps in the HfO_2 layer, and (2) the majority of these traps remain filled after



(a)



(b)

Fig. 27. Typical I_D-V_G curves for 22 nm FDSOI CTT devices with channel width of 300 nm after programming, (a) irradiation to 500 krad(SiO_2), and (b) 30 min. of RT annealing, reprogramming, and additional annealing.

devices are irradiated to 500 krad(SiO_2) [1]. Hence, the significant, negative V_{th} shifts in these devices are caused most likely by the electrostatic effects of radiation-induced trapped holes in the BOX on these FDSOI devices [63]–[65], not by neutralization of programming-induced charge.

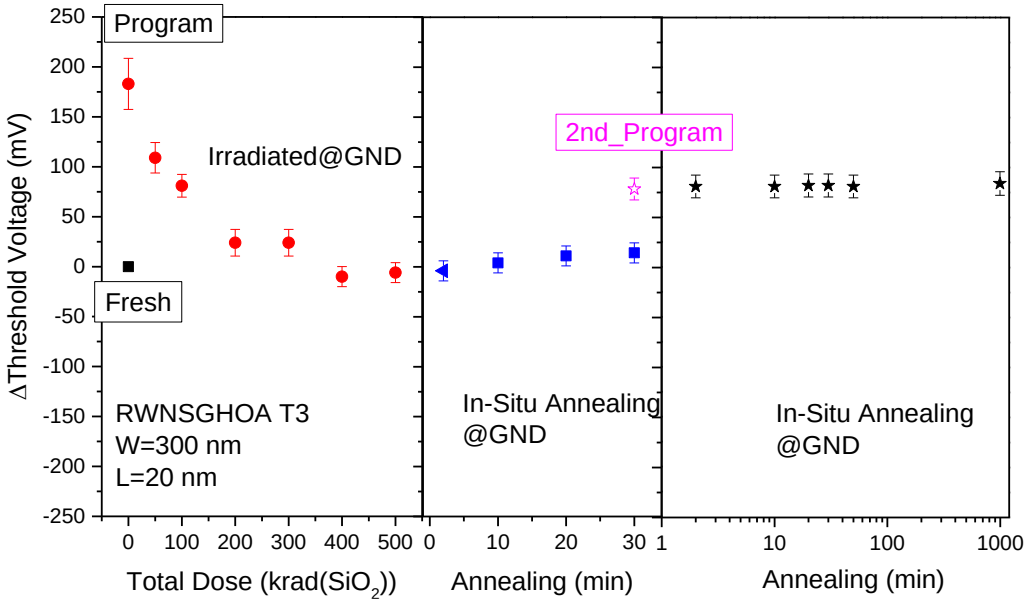


Fig. 28. Extracted V_{th} shifts, with error bars denoting standard deviation for three 22 nm FDSOI CTT devices programmed and irradiated using sequences similar to those shown in Fig. 27.

The results of Fig. 25-Fig. 28 contrast with the efficient neutralization of trapped charge often observed in Flash memories [18], [19] and/or SiO_2 gate dielectrics [66] when programmed devices are irradiated to similar doses. The gate dielectric responses are similar instead to those of earlier-generation charge-trapping memories based on silicon nitride that are quite robust to ionizing radiation [17]. Similar to HfO_2 [67]–[71], Si_3N_4 dielectric layers also contain high densities of hole and electron traps that are not readily neutralized by radiation-induced charge [17], [72], [73].

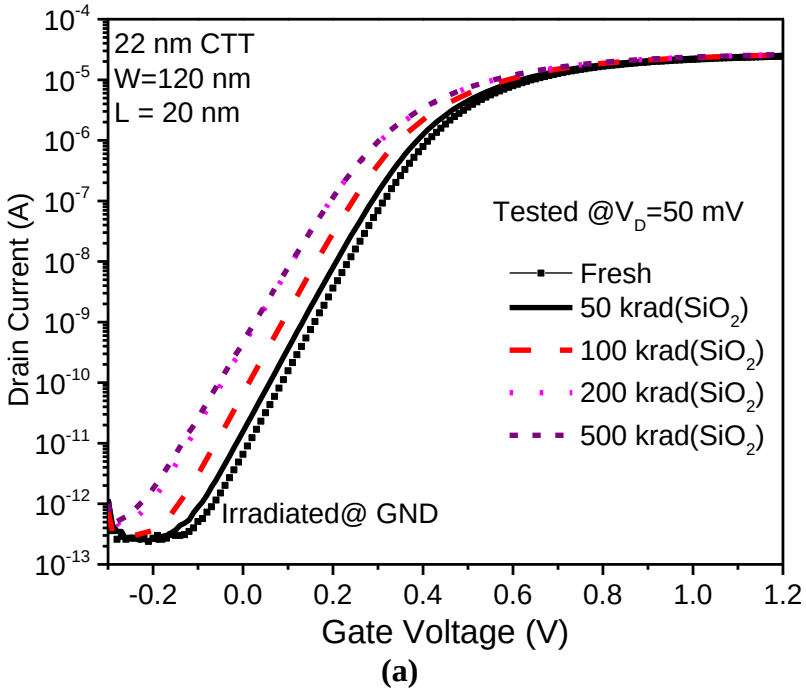
Programming After Irradiation

Fig. 29 shows typical I_D-V_G curves for 22 nm FDSOI devices (W = 120 nm, L = 20 nm) programmed after irradiation. The black dotted curve (labeled "Fresh") shows results before programming or irradiation. Irradiation to 500 krad(SiO_2) in Fig. 29(a) leads to $\Delta V_{th} \approx -120$ mV. Fig. 29(b) shows annealing after TID, programming, and annealing after programming. Fig. 30 shows extracted V_{th} shifts and error bars for several devices irradiated and programmed via similar sequences. Programming after irradiation in Fig. 29(b) results in a positive V_{th} shift of 145 mV, which is ~45% larger than shifts in unirradiated devices when programmed in Fig. 25 and Fig. 26.

A significant decrease in slope of the subthreshold curve is observed in the post-programming I_D-V_G curve and quantified in Fig. 31. If this were due to the buildup of interface traps, one would expect to see significant degradation in the peak transconductance G_m of the devices after programming. Fig. 31(a) shows selected G_m-V_G curves for the devices of Fig. 29(b), and Fig. 31(b) shows peak G_m, normalized to initial values, through the full irradiation, annealing, and programming sequence of Fig. 29(b). Programming the devices after irradiation leads to a slight *increase* in peak G_m. Hence, the decrease in subthreshold slope observed in Fig. 29 and quantified in Fig. 31 is most likely not due to the buildup of a significant density of interface traps [61]. Instead, it is more consistent with lateral nonuniformities in the oxide-trap charge distribution [74] and/or a significant border-trap density, as discussed in Chapter VI.

The negative radiation-induced V_{th} shifts in Fig. 30 are similar to those observed after programming in Fig. 26 and Fig. 28. When devices are programmed after irradiation and annealing, a positive V_{th} shift of ~160 mV is observed in Fig. 30. The ability to fully program devices after irradiation in Fig. 29 and Fig. 30 strongly suggests that little radiation-induced charge is trapped in the HfO_2 layer when devices are irradiated at 0 V [67], [69], [70]. This result is

consistent with the above inference that the radiation response of these FDSOI devices is dominated by radiation-induced hole trapping in the BOX [63]–[65].



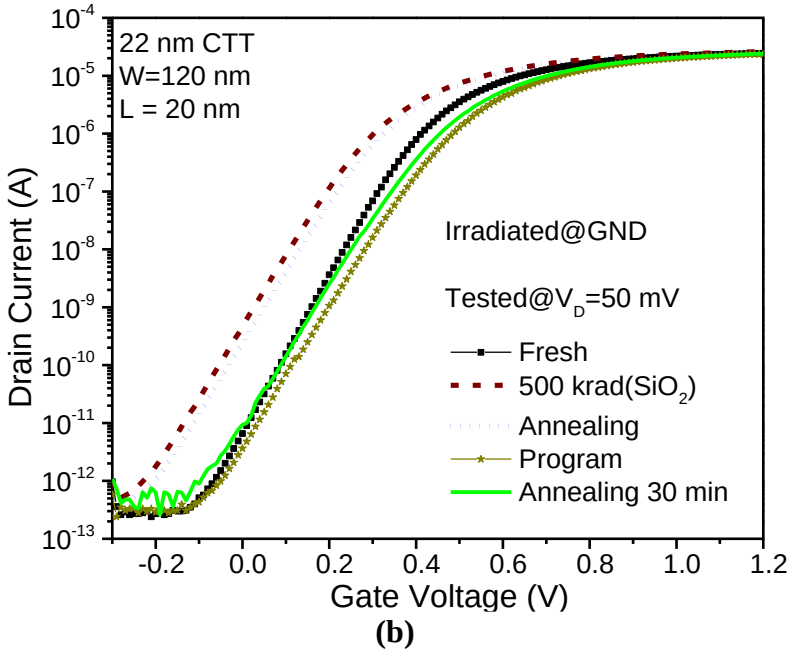


Fig. 29. Typical I_D-V_G curves for 22 nm FDSOI CTT devices after (a) irradiation to 500 krad(SiO_2) and (b) 30 minutes of RT annealing, reprogramming, and additional annealing.

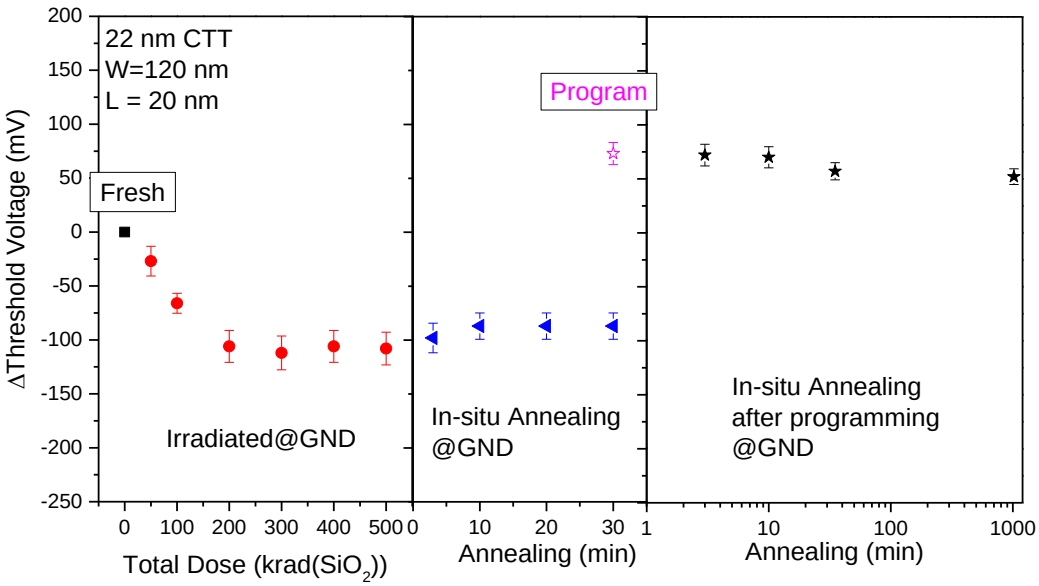


Fig. 30. Extracted threshold voltage shifts for three 22 nm FDSOI CTT devices programmed and irradiated using sequences similar to those shown in Fig. 29.

Erasing Before and After Irradiation

Fig. 32 and Fig. 33 investigate the effects of sequential erasure and programming steps on a 22 nm FDSOI CTT (W = 120 nm, L = 20 nm) before and after irradiation. The fresh device is initially stressed using the erase sequence. A negligible change is seen in V_{th} since no electrons had yet been trapped in the gate to remove via erasing. Hence, the erased state is the same as the initial state in these devices. Programming the device leads to a 70 mV shift. Erasing and reprogramming the devices shows these processes are repeatable and reversible to within ± 5 mV in these 22 nm FDSOI CTTs. After devices are irradiated to 100 krad(SiO_2), $\Delta V_{th} \approx$ -100 mV, and little change is observed in V_{th} during annealing. After irradiation and annealing, the device can still be erased and reprogrammed, consistent with the inference above that the radiation-induced changes in V_{th} are caused primarily by charge trapping in the BOX.

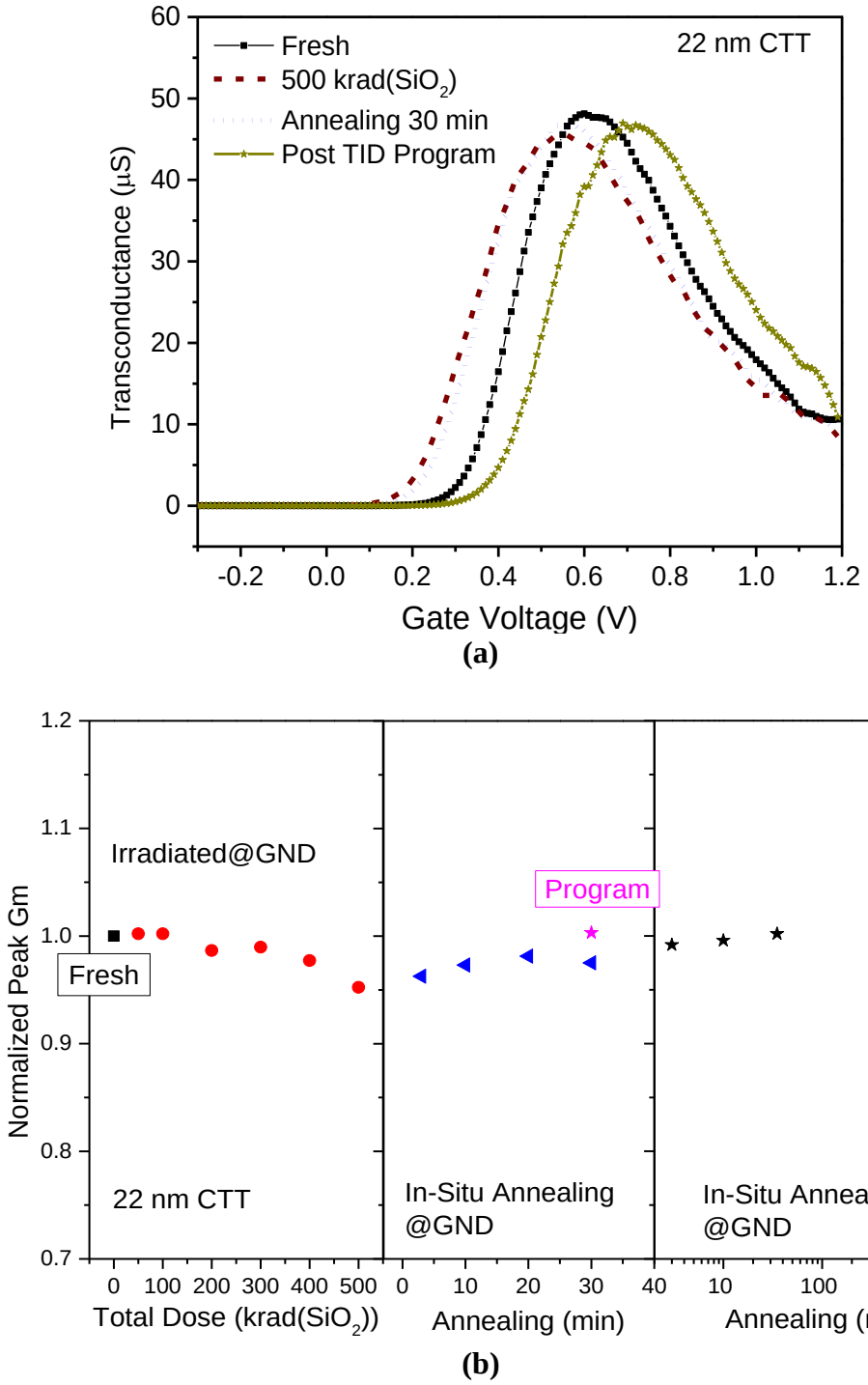
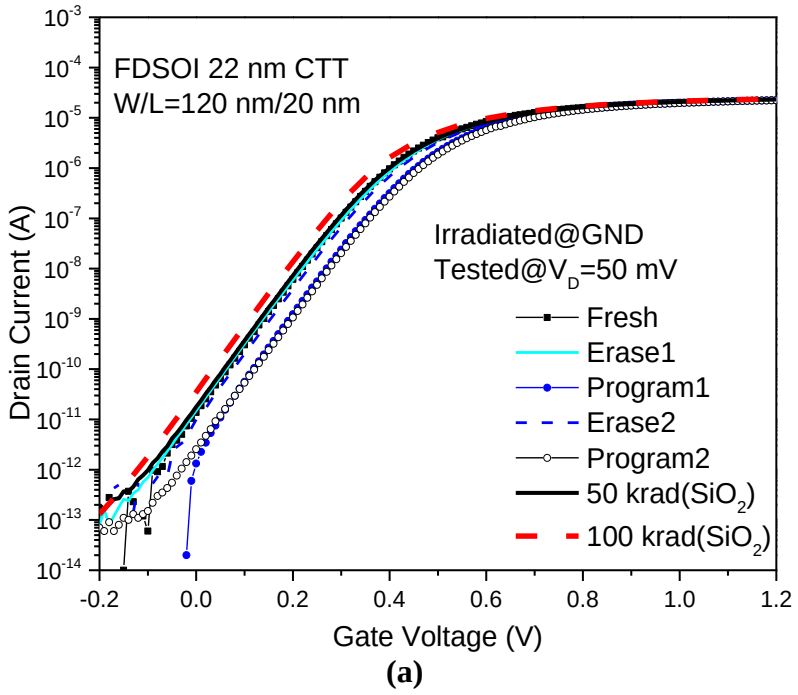
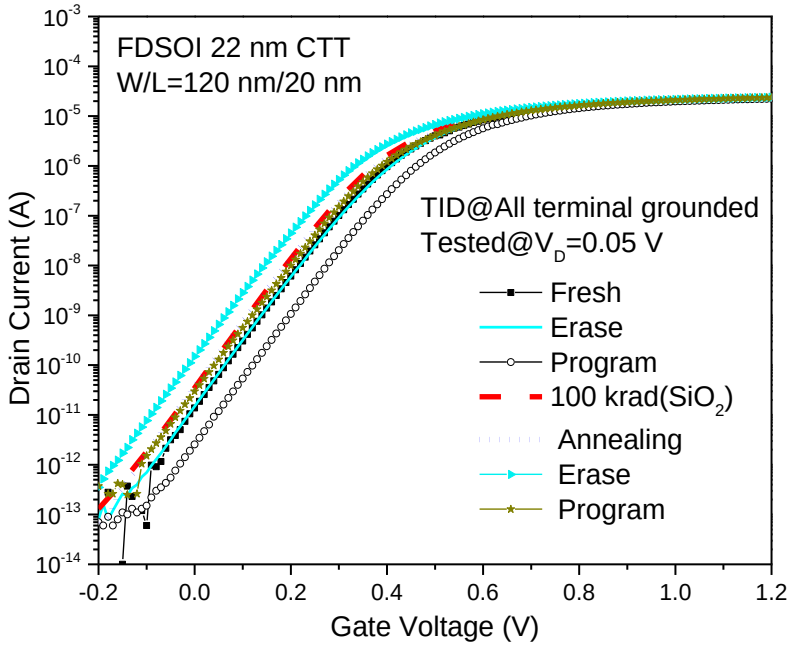


Fig. 31. (a) Selected G_m-V_G curves and (b) peak transconductance normalized to the initial value for the 22 nm FDSOI CTT device (W = 120 nm, L = 20 nm) of Fig. 29(b).



(a)



(b)

Fig. 32. Selected I_D-V_G curves for 22 nm FDSOI CTT devices after (a) two erase and program cycles, irradiation to 100 krad(SiO_2), (b) 90 minutes of RT annealing, and an erase and program cycle.

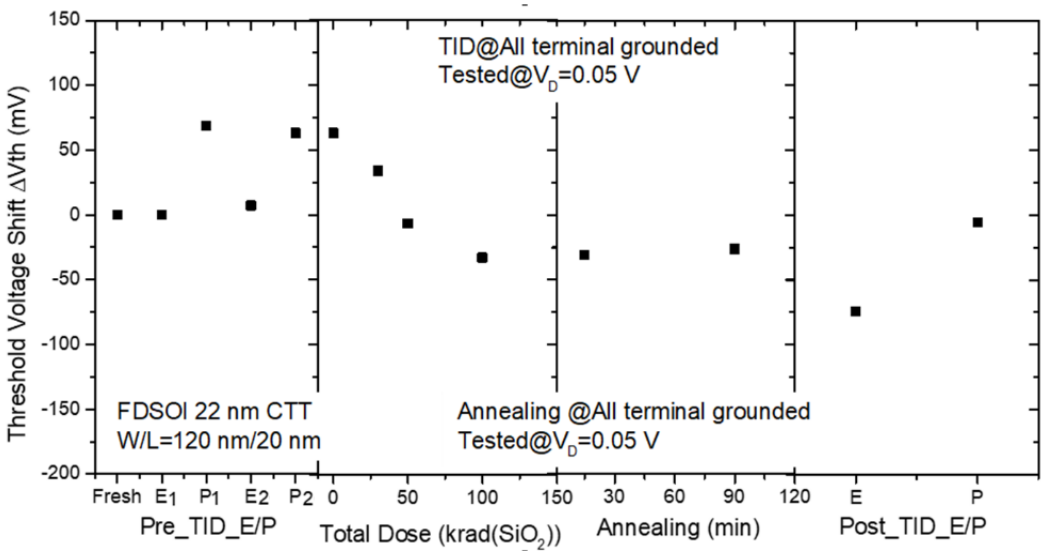


Fig. 33. Extracted V_{th} shifts for the 22 nm FDSOI CTT device in Fig. 32 that was cycled between erasing and programming, irradiated, annealed at RT, and then erased and programmed again.

CHAPTER V

TID RESULTS ON 14 NM BULK FINFET CTTS

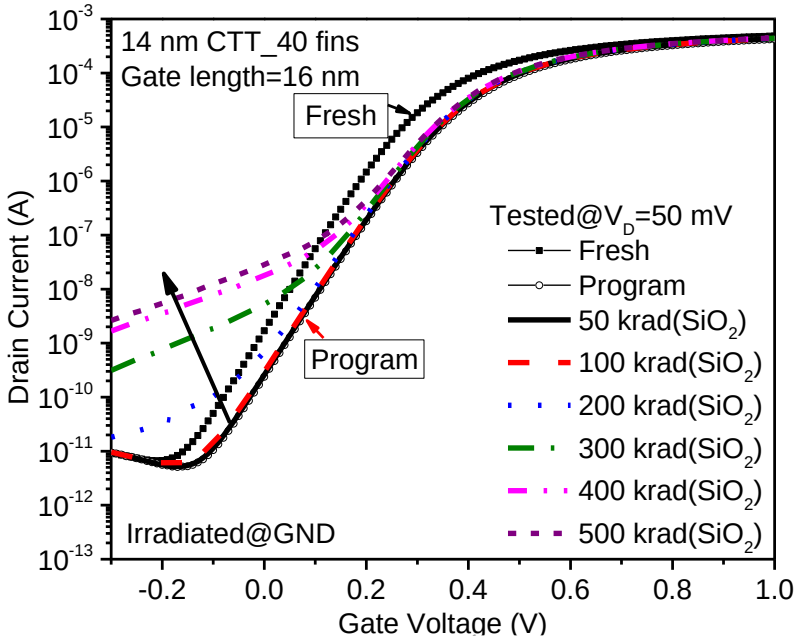
The experimental data and results from TID irradiation on the 14 nm bulk FinFET CTTs are presented here. The majority of this chapter has been published in [62].

Programming Before and After Irradiation

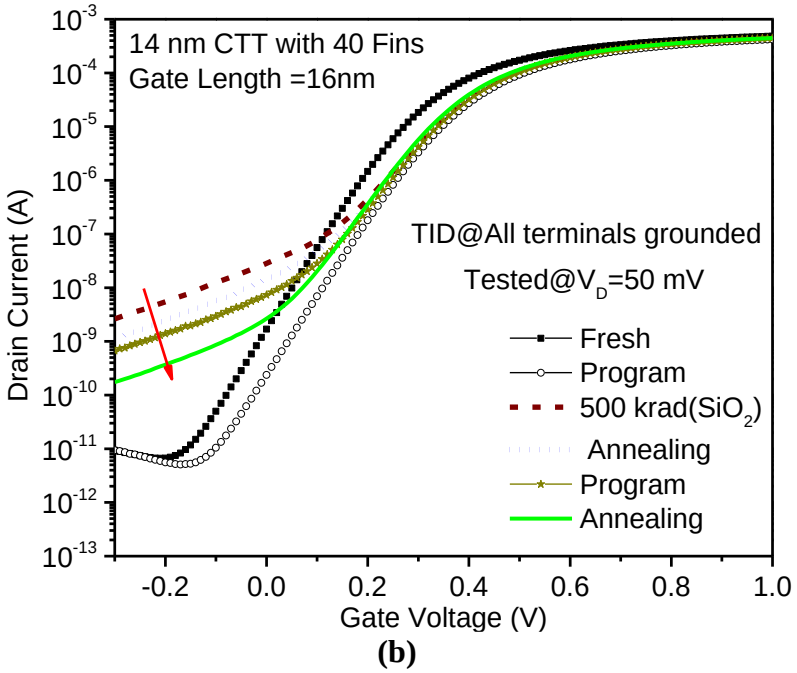
CTTs with 40 Fins

Now the responses of CTTs built in a 14 nm bulk FinFET technology are evaluated. Fig. 34(a) shows I_D-V_G curves for a device with L = 16 nm and 40 fins of width 75 nm (total effective fin width of 3 μm), before and after it was programmed and irradiated. Post-irradiation annealing and a second programming sequence are shown in Fig. 34(b). Fig. 35 shows extracted ΔV_{th} data and error bars for devices tested in the same manner. Programming unirradiated bulk FinFETs results in an ~80 mV V_{th} shift, again due primarily to electron trapping [1]–[3], [61]. This shift is about 20% less than shifts observed for unirradiated 22 nm FDSOI devices in Fig. 25 and Fig. 26 as a result of the reduced programming voltage, and concomitant decrease in hot carrier injection. Irradiating devices to 500 krad(SiO_2) leads to V_{th} shift of −15 mV, with little change in V_{th} during annealing. Only ~10 mV of positive V_{th} shift is observed when the device is programmed for a second time, with a V_{th} shift of −15 mV during the subsequent annealing period. In contrast to the FDSOI devices of Fig. 25 and Fig. 26, the 14 nm bulk FinFETs retain a significant memory window through the full irradiation and annealing sequence. This reinforces the above conclusion

that the closure of the memory window for the SOI devices is due to charge trapping in the BOX, and not the neutralization of programming-induced charge in the HfO_2.



(a)



(b)

Fig. 34. Typical I_D-V_G curves for 14 nm CTTs with 40 fins and total effective fin width of 3 μm after programming, (a) irradiation to 500 krad(SiO_2), and (b) 35 minutes of RT annealing, reprogramming, and additional annealing. Arrows denote directions of (a) increasing and (b) decreasing leakage current over time.

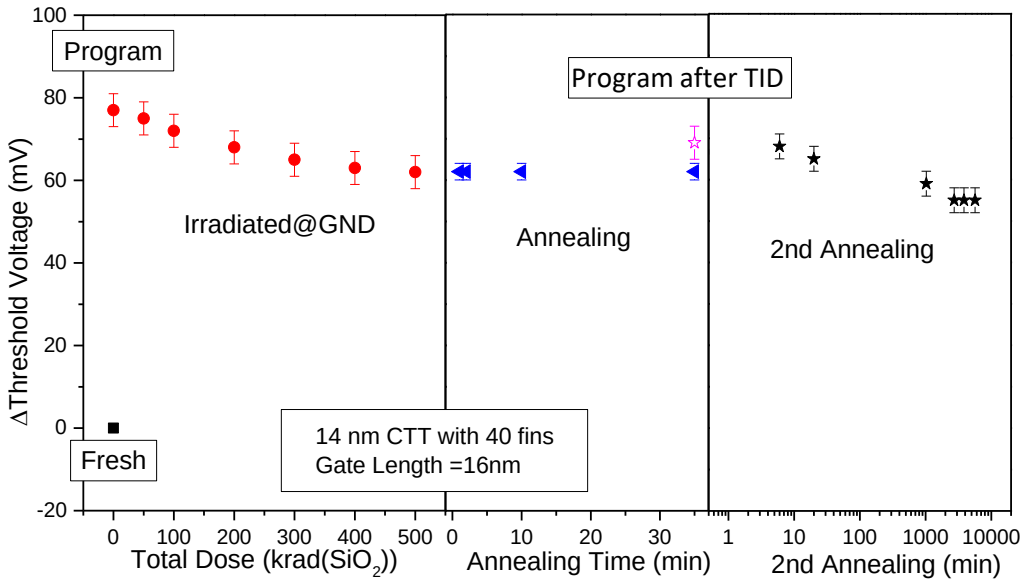
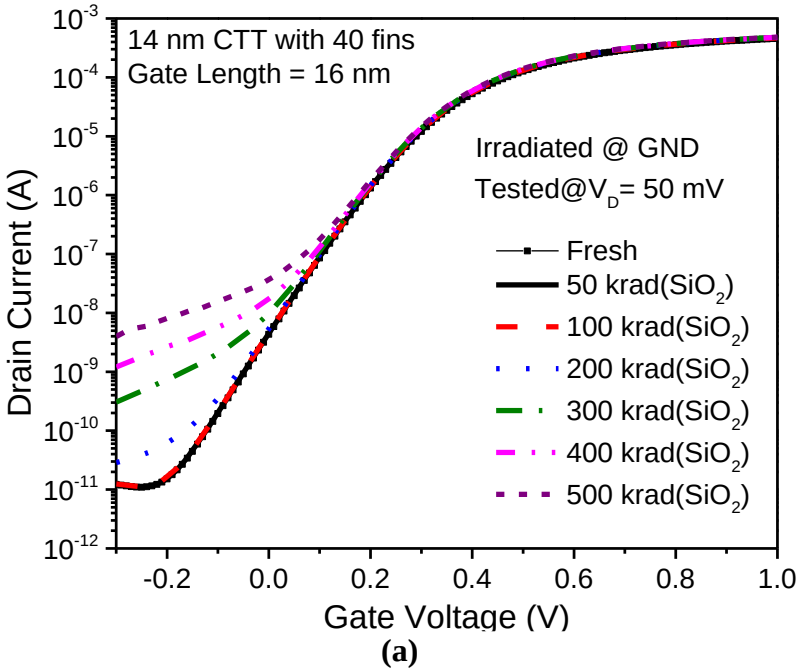


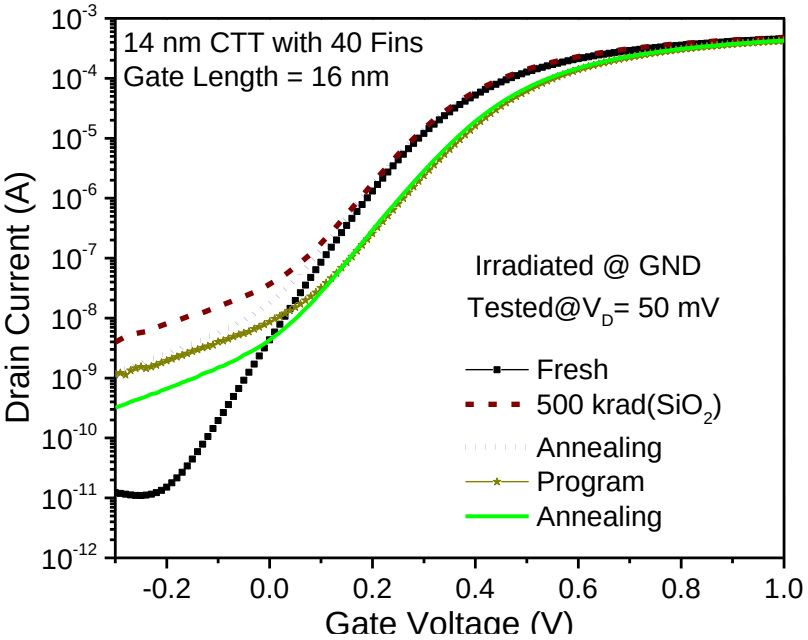
Fig. 35. Extracted V_{th} shifts of three typical 14 nm CTT devices with 40 fins and total effective fin width of 3 µm, programmed and irradiated using sequences similar to those in Fig. 34.

Significant subthreshold leakage is observed in Fig. 34 at doses above ~200 krad(SiO_2), due most likely to charge trapping in the shallow trench isolation (STI) in proximity to the sub-fin regions of the devices [8], [75], [76]. This contrasts with the FDSOI devices of Fig. 25-Fig. 33, for which no STI- or BOX-related leakage current is observed. The leakage current decreases less than an order of magnitude during annealing.

Fig. 36(a) shows I_D-V_G curves before and after a device with 40 fins, W = 3 µm, and L = 16 nm was irradiated up to 500 krad(SiO_2) and annealed in-situ. The device was then programmed and annealed, as shown in Fig. 36(b). Fig. 37 shows extracted ΔV_{th} data and error bars for several devices tested in the same manner. Irradiating these bulk FinFETs resulted in a V_{th} shift of −10 mV, with little change during annealing. A positive V_{th} shift of 80 mV is observed when the device is programmed after irradiation, with less than 10 mV of V_{th} shift during annealing. That it is possible to fully program irradiated FDSOI and bulk FinFET CTTs confirms that the programming



(a)



(b)

Fig. 36. Typical I_D-V_G curves for 14 nm CTT devices with 40 fins and total effective fin width of 3 µm after (a) irradiation to 500 krad(SiO_2), and (b) 30 minutes of RT annealing, programming, and additional annealing.

sequence is much more effective in filling traps in the HfO_2 dielectric than exposure to irradiation. Hence, the radiation response of these devices is determined primarily by charge trapping in isolation oxides.

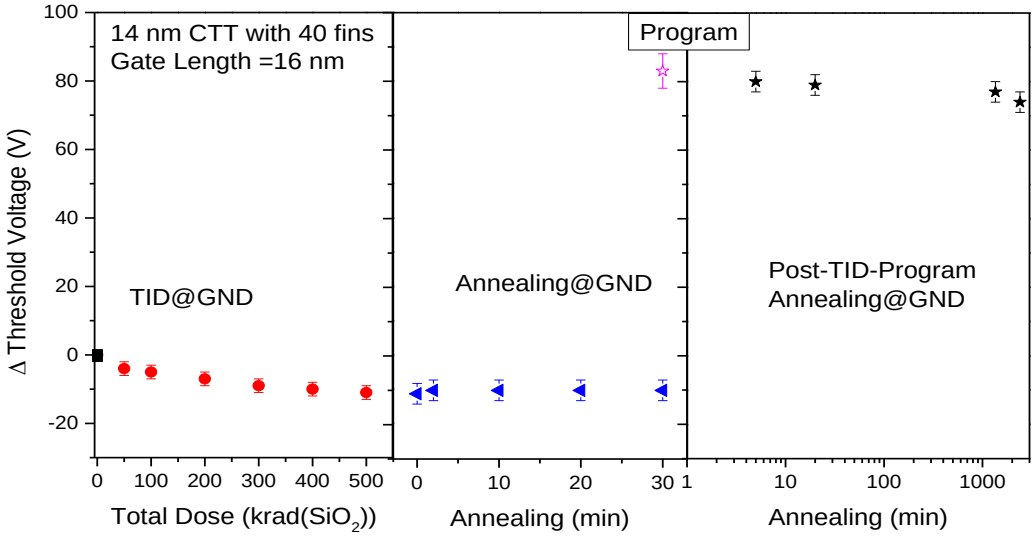
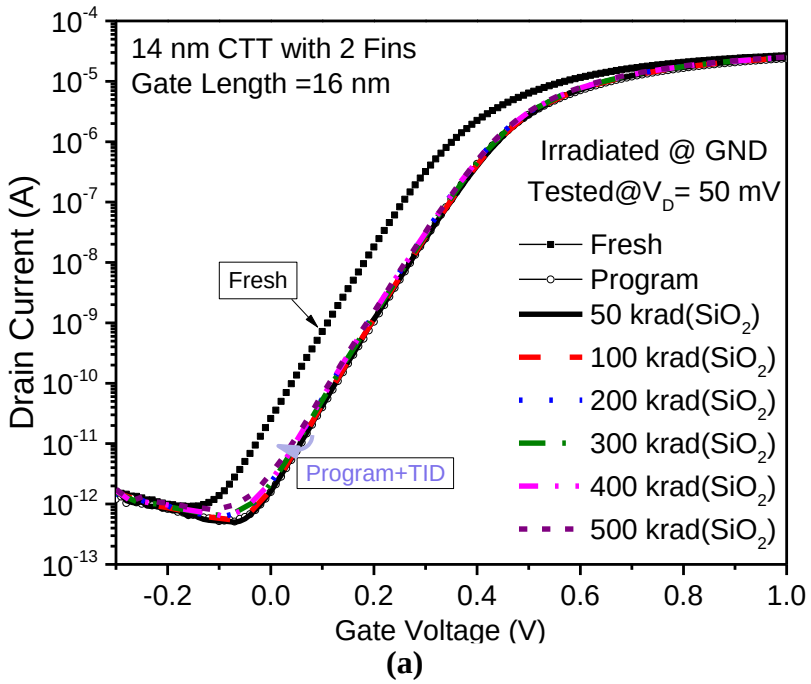


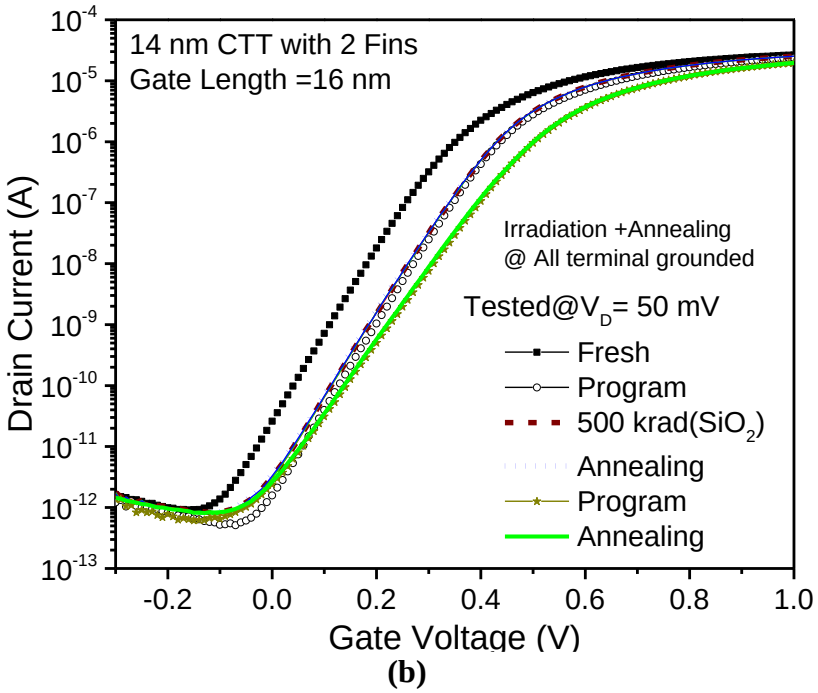
Fig. 37. Extracted V_{th} shifts for three typical 14 nm CTT devices with 40 fins and total effective fin width of 3 μm, programmed and irradiated using sequences similar to those in Fig. 36.

CTTs with 2 Fins

Fig. 38 shows I_D-V_G curves for a 14 nm bulk FinFET with two fins, effective fin width of 150 nm, and gate length of 16 nm. Fig. 39 shows extracted ΔV_{th} data and error bars for devices tested in the same manner. Programming unirradiated devices leads to V_{th} shifts of 85 mV, and irradiation to 500 krad(SiO_2) causes almost no shift in V_{th}. Programming the device without an erase cycle after irradiation leads to an additional V_{th} shift of 70 mV. In Fig. 40, it is demonstrated that this decrease in subthreshold slope is accompanied by a significant reduction in peak G_m. Hence, the decreased subthreshold slope for the 14 nm FinFETs is due most likely to the generation of interface traps, as discussed in Chapter VI.



(a)



(b)

Fig. 38. Typical I_D-V_G curves for 14 nm CTT devices with 2 fins and total effective fin width of 150 nm after programming, (a) irradiation to 500 krad(SiO_2), and (b) 35 minutes of RT annealing (obscured by the initial program and irradiation curves), reprogramming, and additional annealing.

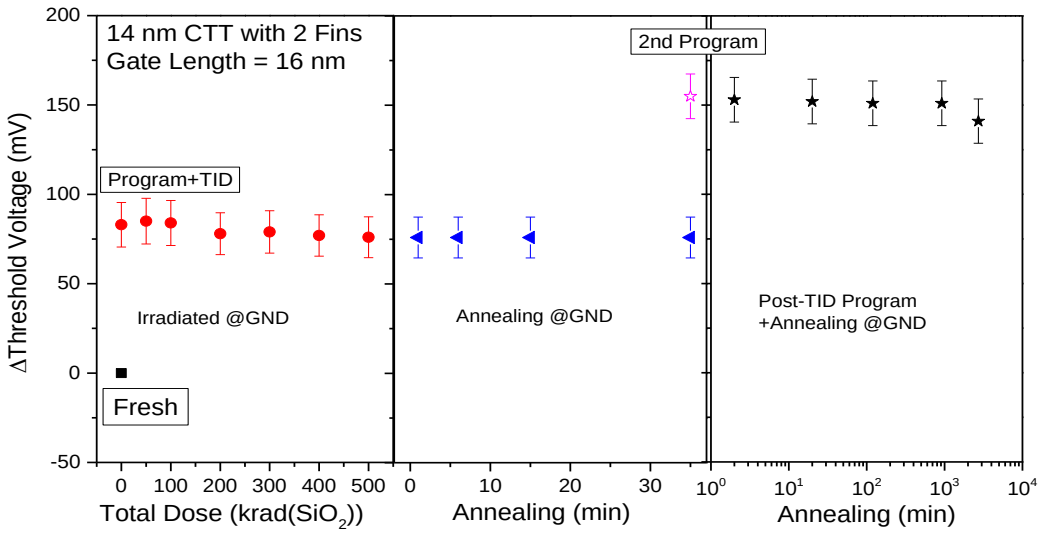
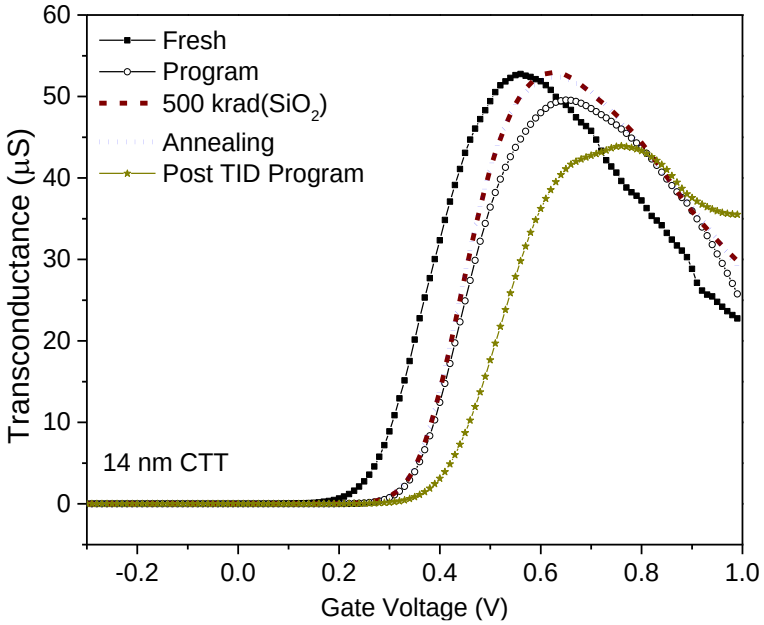
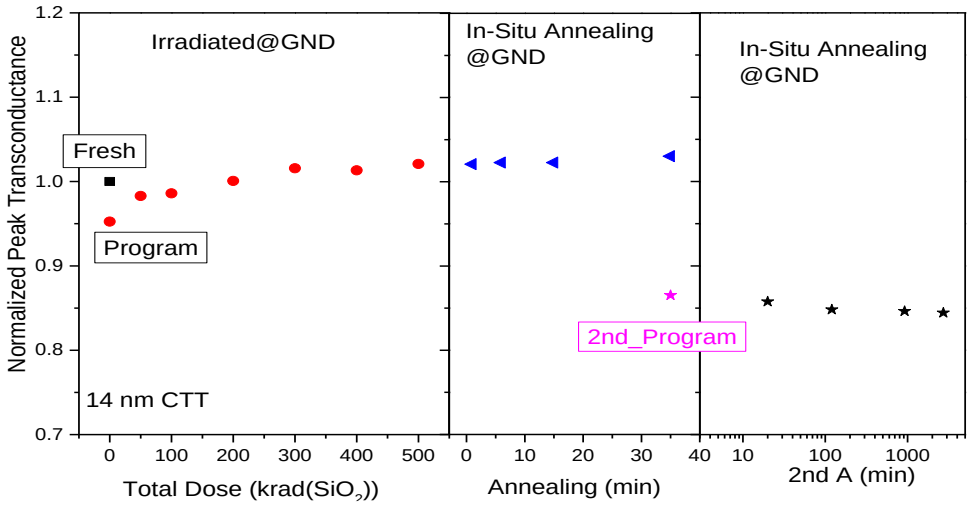


Fig. 39. Extracted V_{th} shifts for 14 nm CTT devices with 2 fins programmed and irradiated using sequences similar to those shown in Fig. 38.

Fig. 41 shows I_D-V_G curves for 14 nm bulk FinFETs with two fins and total effective fin width of 150 nm that were irradiated and then programmed after annealing. Fig. 42 shows extracted ΔV_{th} data and error bars for several devices tested in the same manner. Irradiation to 500 krad(SiO_2) causes almost no V_{th} shift. Programming the device after irradiation leads to a V_{th} shift of ~80 mV, similar to that of unirradiated devices in Fig. 34. In contrast to the 40-fin devices of Fig. 36, no STI-related leakage is observed for the 2-fin devices of Fig. 41. Comparing Fig. 34 and Fig. 38, the reduction in leakage (up to a factor of 1000 or more) at the highest doses is much greater than the 20:1 ratio of the numbers of fins. Thus, it is likely that the enhanced leakage in the devices with 40 fins and total effective fin width of 3 μm is due to the increased probability that one or more fins will have a more defective sub-fin/STI interface than average, due to the much greater widths, and/or differences in layout of the two structures [8], [76].



(a)



(b)

Fig. 40. (a) selected G_m-V_G curves and (b) peak transconductance normalized to the initial value for the 14 nm FinFET device with 2 fins and total effective fin width of 150 nm of Fig. 38(b).

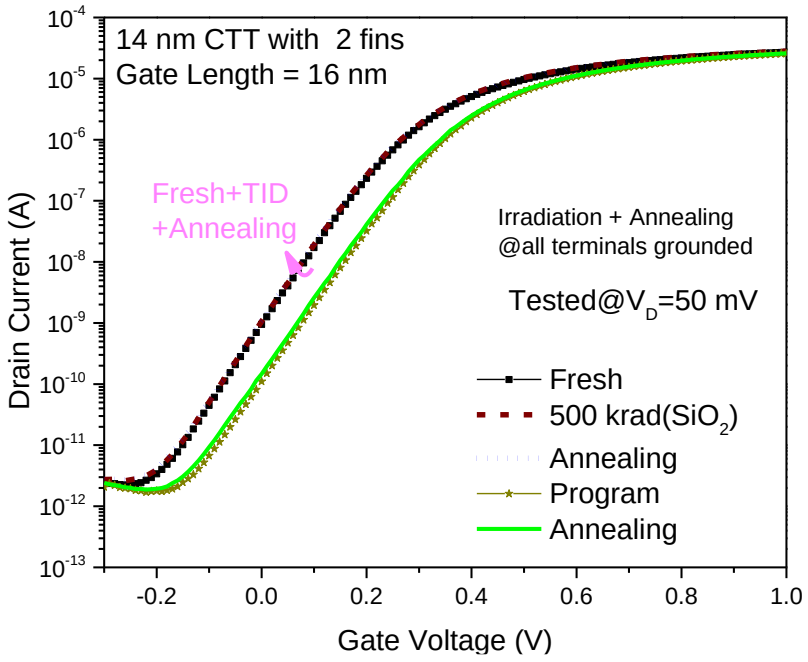


Fig. 41. Typical I_D-V_G curves for 14 nm CTT devices with 2 fins and total effective fin width of 150 nm after irradiation to 500 krad(SiO_2), 30 minutes of RT annealing, and programming.

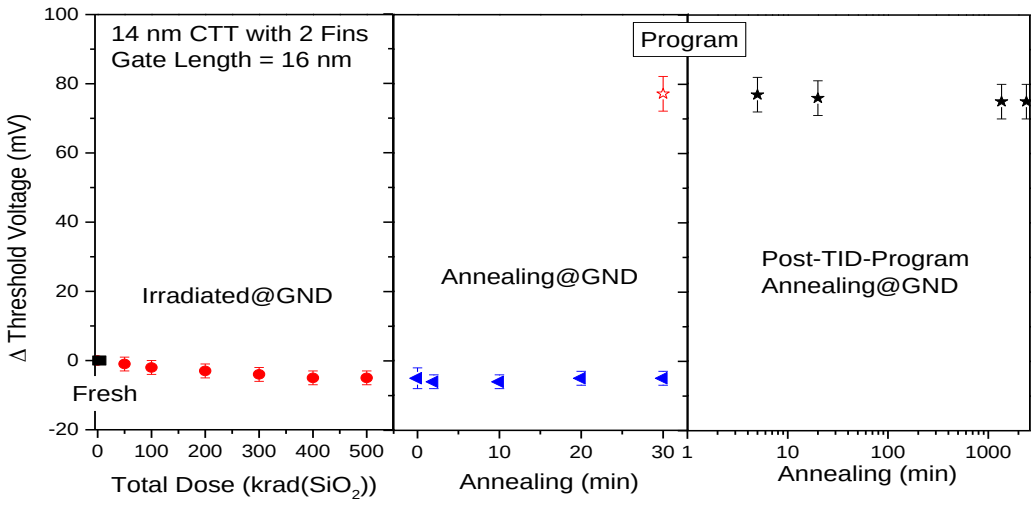
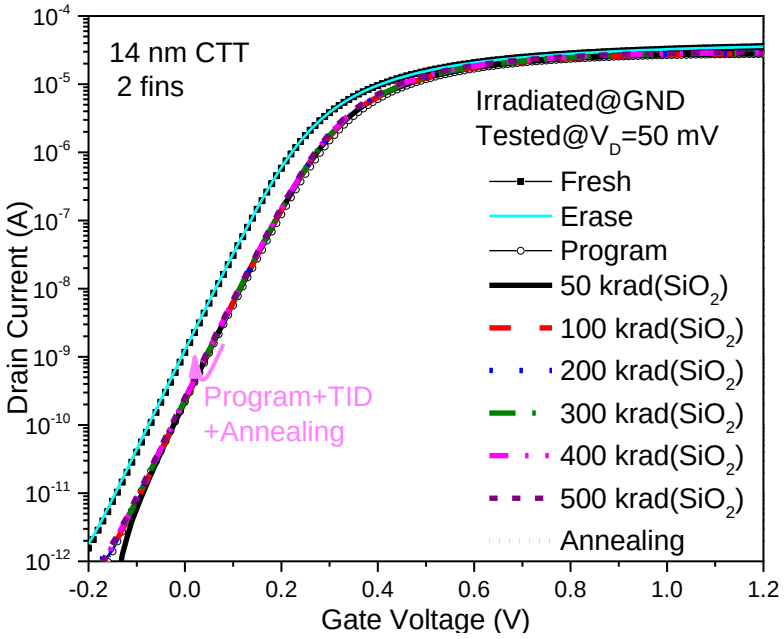


Fig. 42. Extracted V_{th} shifts for three 14 nm CTT devices with 2 fins and total effective fin width of 150 nm programmed and irradiated using sequences similar to those shown in Fig. 41.

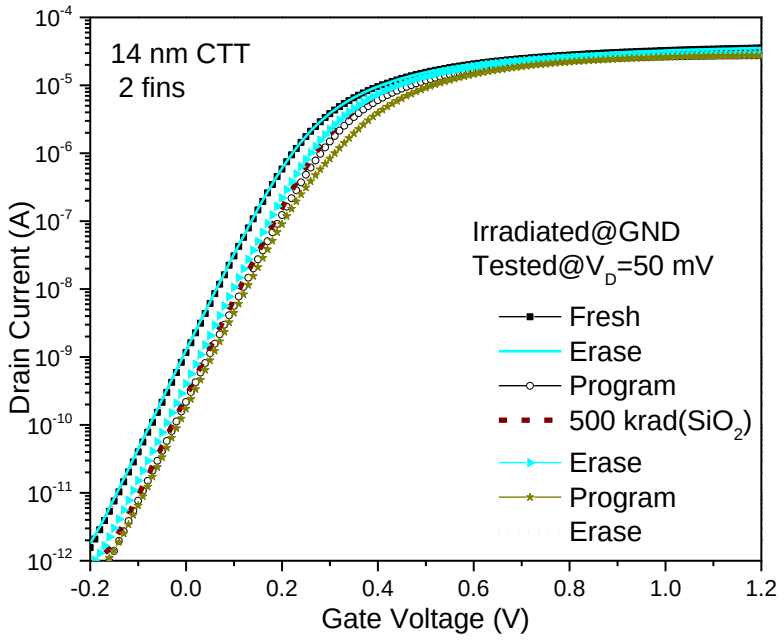
Erasing Before and After Irradiation

Fig. 43 and Fig. 44 show selected I_D-V_G curves and a summary of results, respectively, for a 14 nm bulk FinFET CTT with 2 fins and total effective fin width of 150 nm that was erased and programmed before and after irradiation in a sequence similar to that of the 22 nm FDSOI device of Fig. 32 and Fig. 33. The as-processed device is first given an erase cycle resulting in no V_{th} shift. Subsequent programming results in a threshold voltage shift of 55 mV, which is ~30% smaller than the averages of values for initially programmed devices in Fig. 36-Fig. 39 above. This reduction in shift is due most likely to sample-to-sample variation, since responses are otherwise similar. Erasing the programmed device returns the device to its initial state. The device is then reprogrammed and irradiated to 500 krad(SiO_2). Negligible changes in V_{th} are observed during irradiation and post-irradiation annealing.

In contrast to the CTT devices of Fig. 32 and Fig. 33, the 14 nm CTTs that are programmed and irradiated to 500 krad(SiO_2) cannot be erased using the same sequence of pulses employed to successfully erase these devices before irradiation and erase the 22 nm CTT devices before and after irradiation (Fig. 34 and Fig. 35). Reprogramming leads to a more positive shift, similar to the response observed in Fig. 39 in the absence of erase cycles. After reprogramming, again the charge is stable during an erase cycle, due most likely to an increase in density of stable electron traps during irradiation [71], [77], [78]. Attempts to increase the magnitude, duration, or number of negative pulses for these devices lead to damage or device destruction. Hence, additional refinement may be required before these 14 nm CTTs are suitable for use in high-dose radiation environments.



(a)



(b)

Fig. 43. Selected I_D-V_G curves for 14 nm CTTs with 2 fins and total effective fin width of 150 nm after (a) initial erase, program, irradiation to 500 krad(SiO_2), 45 minutes of RT annealing, and (b) erase, reprogram, and erase.

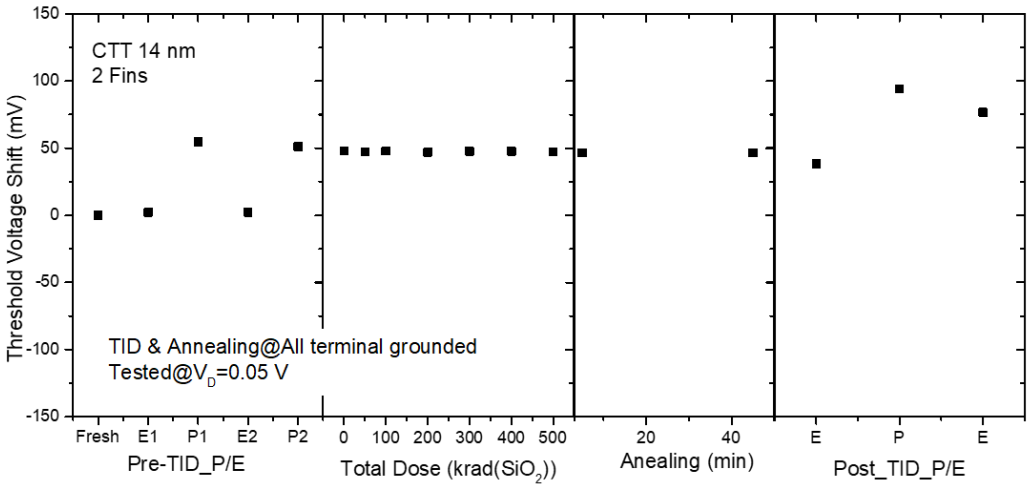


Fig. 44. Extracted V_{th} shifts for the 14 nm CTT of Fig. 43.

CHAPTER VI

DISCUSSION OF TID RESULTS

The results of Fig. 34-Fig. 43 show that the 14 nm FinFET-based CTTs show similar programming-induced V_{th} shifts to those of the 22 nm FDSOI devices of Fig. 25-Fig. 33. The HfO_2 gate dielectrics show similar resilience to radiation exposure before and after programming for the 22 nm FDSOI and 14 nm FinFET based CTTs. STI-related leakage is observed in 14 nm devices with 40 fins and total effective fin width of 3 μm, but not in 14 nm devices with 2 fins and total effective fin width of 150 nm. Taken together, the results of Fig. 25-Fig. 44 strongly suggest that neither radiation-induced charge trapping nor neutralization of trapped negative charge in the HfO_2-based gate dielectrics limit the radiation tolerance of these CTTs. Thus, the tolerance of CTT memory devices based on transistors similar to those evaluated in this work are likely to be limited primarily by the charge-trapping properties of surrounding insulators. This contrasts with most Flash memory devices, where the radiation tolerance typically is limited by the neutralization of charge on the floating gate [18], [19].

Consideration is now given to why increased subthreshold stretchout (decreases in subthreshold slope of the I_D-V_G curves) is observed in Fig. 27 and Fig. 29 for FDSOI devices after irradiation and programming, and in Fig. 38 for 14 nm FinFETs when devices are irradiated after programming. Because there is no degradation in peak G_m in Fig. 31, the increased subthreshold stretchout for the 22 nm FDSOI device is evidently not caused by interface-trap buildup. Fig. 45(a) shows a schematic diagram of a 22 nm FDSOI CTT being programmed after irradiation. A fraction of hot electrons generated by the programming sequence will be trapped in the gate

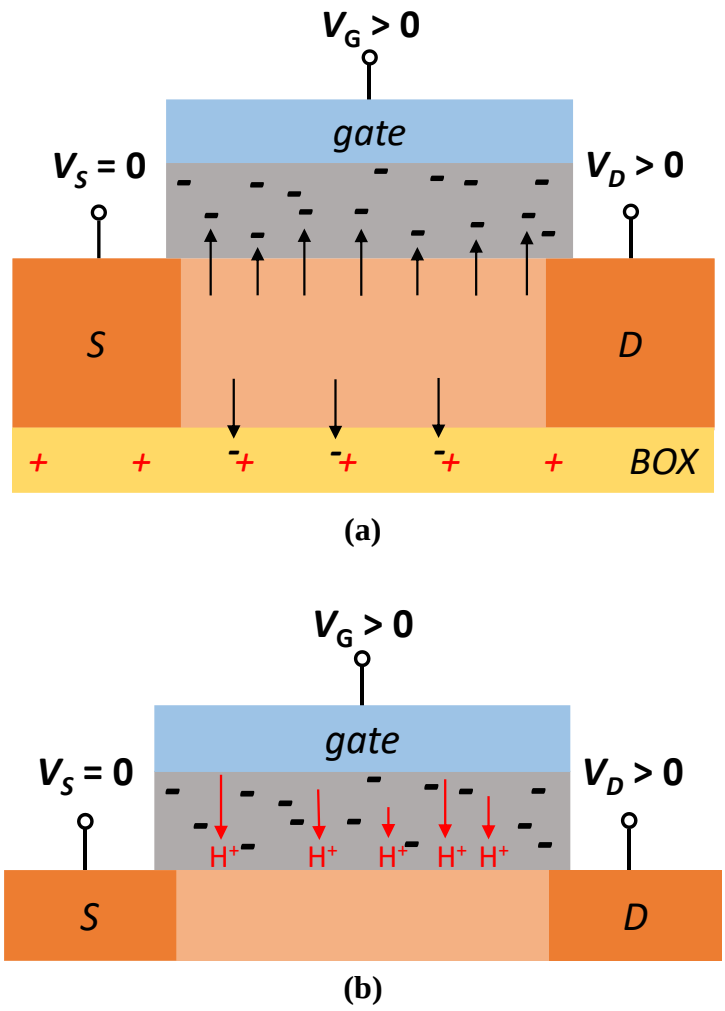


Fig. 45. Schematic illustrations of programming under positive gate bias (a) an irradiated FDSOI CTT and (b) an irradiated 14 nm FinFET (shown in cross section for an individual gated region). In (a), most hot electrons are trapped in the gate as desired, but some scatter into the BOX where they neutralize radiation-induced trapped positive charges. In (b), protons are released within the gate dielectric during irradiation and/or programming and transport to the interface under the influences of the large positive programming bias and/or built-in electric field, creating border or interface traps.

dielectric, as desired [1]–[3]. While the majority of hot electrons generated by the programming pulses are attracted to the gate, some hot electrons inevitably scatter into the BOX, enabling the neutralization of a portion of the trapped positive charge [61], [79]. The resulting variations in

densities of near-interfacial trapped charge along the channel can lead to spatial variations in V_{th}. These local variations in charge density and V_{th} evidently result in a significant portion of the observed increase in subthreshold slope of the I_D-V_G curves for the 22 nm FDSOI devices in Fig. 27 [65], [74], [80]. Non-uniform trapping of electrons [1]–[3] and generation of border traps in the HfO_2 gate dielectric [61], [81] can potentially add to the resulting increase in subthreshold stretchout.

As shown by the results of Fig. 40, the increase in subthreshold stretchout for the 14 nm FinFETs evidently is due at least in part to the buildup of interface traps. As shown schematically in Fig. 45(b), one possible mechanism of interface-trap formation during programming is the release of a proton in the gate dielectric during the stressing sequence. These protons can transport to or near the Si/SiO_2 interface under the influences of the large positive programming bias and/or built-in electric field due to the gate-to-substrate work function difference [67]–[70] and react to form interface and/or border traps [61], [82]–[86]. DiMaria *et al*. have shown that the threshold for hydrogen release via hot carrier injection into SiO_2 is ~2 eV [84]. Marinopoulos *et al*. have shown that this barrier for hydrogen release in a HfO_2 dielectric is reduced near a Hf atom [87]. While the probability of hydrogen release by an injected electron is initially much lower than the probability of trapping, this mechanism of interface- and/or border-trap formation may become significant after the majority of electron traps are filled, and/or when devices are programmed for a second time after initial programming and irradiation (e.g., for the bulk 14 nm FinFET devices of Fig. 35).

Especially if programming conditions are not optimized, hot electrons can also create interface traps along the channel via impact ionization [84], [86] or multi-carrier-scattering induced multi-vibrational excitation release of passivating hydrogen atoms from dangling Si bonds

at the Si/SiO_2 interface [88]–[90]. Recently, Reaz *et al*. [91] also have shown that electrons with energies higher than ~1.3 eV [92] can lead to efficient release of hydrogen from passivated phosphorus dopant atoms (~1.4 eV for arsenic [93]) within the drain of nanoscale gate-all-around devices [91]–[93]. The diffusion of these hydrogen atoms can lead to interface-trap formation in the nearby channel region via processes similar to those inferred to lead to interface-trap creation as a result of negative-bias instability and/or irradiation [81]–[91]. Hence, this mechanism may also play a role in the interface-trap buildup in these CTT devices during a second programming cycle without erasure. Additional work is required to determine whether this is the case. The resolution of this issue is important not only to understand the irradiation response, but also to understand the role of interface and border traps on degradation due to cycling in CTT devices [1]–[3], [17], [94], [95].

CHAPTER VII

IMPLICATIONS OF TID IN CTT-BASED NEURAL NETWORKS

As mentioned in Chapter II, research has been conducted into the potential application of CTTs for use in neural networks [27], [28], [32], [33]. In these neural networks, CTTs were applied as non-volatile memories to store neuron weights. This chapter leverages simulations of both circuits and neural networks to explore the potential effects that total-ionizing dose could have on neural networks that employ CTTs as neuron weights.

Neuromorphic Computing

A promising application of CTTs is in the field of neuromorphic computing. Neuromorphic computing is brain-inspired, non-von Neumann computing. The goal of neuromorphic computing is to emulate the way the brain processes information as well as imitate the way it approaches and solves complex problems, such as pattern recognition. There are a variety of applications for neuromorphic computing which span across disciplines and include signal and image processing, high performance computing, text and audio processing, learning and optimization among others [96]. Neural networks can be used for neuromorphic computing, and these neural networks can be implemented in software or emulated in hardware. Interest in specialized hardware implementations to accelerate these computations has increased recently. Several companies including IBM, Intel, Qualcomm, and Nvidia have produced specialized hardware solutions implementing neuromorphic computing architectures in digital CMOS [97]. These early generation products will ideally evolve into highly-integrated, three-dimensional, non-volatile

technologies. These future technologies will require implementation of advanced devices that are both low-power and high density. CTTs are one such possible technology.

Neural Networks

Neuromorphic computing architectures are constructed using neural networks. There are several different types of neural networks, including artificial, convolutional, and deep neural networks to name a few. The idea for neural networks to be used in computing is based on biological neural networks, such as those found in the human brain. A biological neural network consists of interconnected neurons which communicate with each other using axons and synapses. Communication signals travel through the neuron via the axon, and this information is transferred to other neurons though a synapse connection between two neurons. The "strength" of the signal is modulated by the weight of the synapse connection. In a similar fashion, a neural network used in computing transfers signals from one neuron to the next, and the strength (or magnitude) of the signal is based on a weight which is specific to the two connected neurons [48]. Neural networks connected in such a fashion where all inputs are connected to all outputs via weights are called fully-connected neural networks. This type of network forms a crossbar array structure, which was covered earlier.

Neuromorphic Architectures

Neuromorphic architectures, composed of neural networks, are densely interconnected with each output sharing large portions of the same input space as inputs from several neurons. Consequently, the neuromorphic architecture possesses increased redundancy compared to traditional computing architectures and can be fault tolerant if specifically designed that way.

However, this comes at a computational cost [98]–[101]. The way neuromorphic computers "learn" is through the use of self-tuning weights (synapses). Neuromorphic architectures utilize a type of non-von Neumann computing. Traditional, von Neumann computer architectures have physically separated memory and computing components that communicate with each other using a bus. This bus can be a bottleneck for computing, particularly in neuromorphic computing applications. Instead, neuromorphic architectures utilize non-von Neumann computing architectures which feature a distributed memory structure allowing for computations to be performed locally with the data storage [102]. As a result, neuromorphic architectures have increased redundancy and a reduction in the memory bandwidth bottleneck compared to traditional computing [98]. **Error! Reference source not found.** shows these two types of architectures.

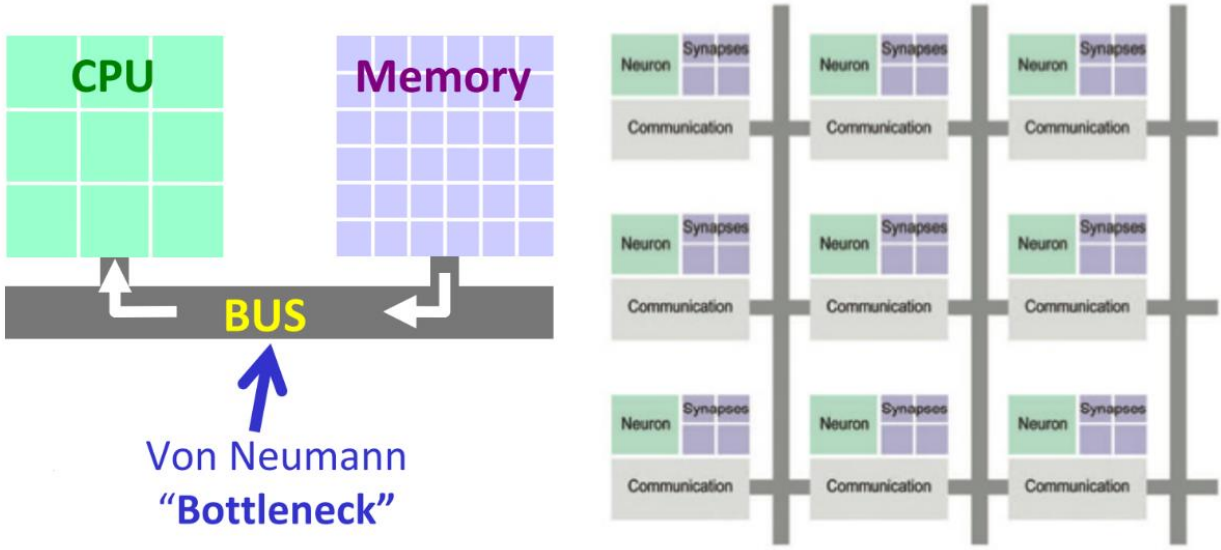


Fig. 46. Comparison of traditional, von Neumann architecture (left) with a non-von Neumann architecture (right) where the calculations and data storage are grouped together in small groups. *(From [102].)*

Artificial Neuron

Artificial neurons are the basic building blocks used in the construction of the abovementioned neural networks [103], [104]. These artificial neurons are a mathematical way to

model biological neurons for use in neuromorphic architectures. Each input x_i to the neuron (a number) is weighted by its corresponding weight w_i, and the sum of all these weighted inputs is computed. This weighted sum is then applied to an activation function φ which then results in the output y. The equation for the weighted output y from a neuron is

$$y = \phi \sum_{i=1}^{n} w_i x_i.$$

The output y from the artificial neuron can then be used as an input to one or more neurons in a subsequent artificial neuron. Artificial neurons can be implemented in software, digital circuits, or analog circuits. **Error! Reference source not found.** shows a schematic of an artificial neuron.

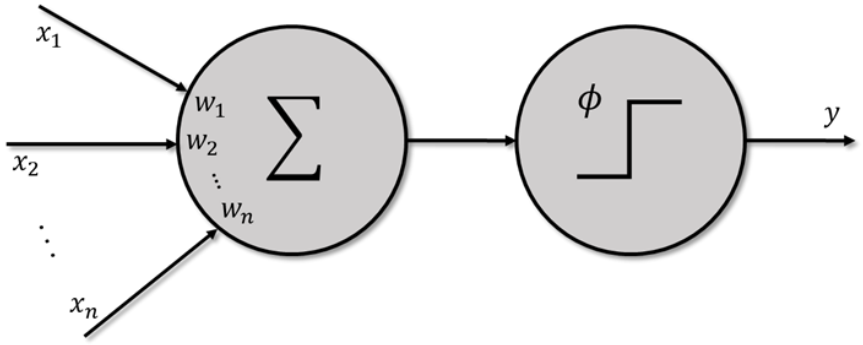


Fig. 47. Representation of an artificial neuron. The inputs to the system are denoted by x_i and the output by y.

Activation Functions

The weighted sum of the inputs can take on a range of values. The purpose of the activation function is to map these weighted sums to one of the possible outputs. In the case of the activation function being a step function, values above a certain point produce a high output, while values below this point result in the output y being low. Thus, even though the weighted sum can take on a variety of values, there are only two possible outputs for a step activation function. Other activation functions exist and have various applications and complexity. Some additional

activation functions include sigmoid and hyperbolic tangent functions, which have bounded outputs; linear functions, whose outputs are unbounded; and functions that are not one-to-one, where different inputs can result in the same outputs.

Artificial Neural Networks

Artificial neural networks can be constructed by linking a number of artificial neurons together where outputs of individual neurons are used as inputs to neurons in the next layer [104]. These networks have more processing power, flexibility, and applicability than single neurons. Artificial neural networks are claimed to be intrinsically fault tolerant; however, most artificial neural networks cannot be considered fault tolerant without a proper design [100], [101]. Passive fault tolerance can be achieved through redundancy or by modifying learning. Including faults during training promotes better generalization than other limited solutions, but it comes at both a higher computational cost and more time necessary for training. Active fault tolerance such as resetting the neural network to a faultless state after a fault occurs and propagates is the most widely used design of fault tolerant neural networks in hardware.

Feedforward Neural Network

A feedforward neural network is an artificial neural network in which artificial neurons are linked together in multiple layers where information propagates one direction through the network [104], [105]. This neural network is composed of many neurons and can solve a wide variety of classification problems. The first layer is called the “input layer” and receives the inputs to the neural network. The final layer is called the “output layer” since this layer produces the final output of the neural network. All layers (if any) between the input and output layers are called “hidden

layers" because their inputs and outputs are not directly observable outside the neural network [105]. The number of hidden layers can vary depending on the application and the complexity needed in the neural network. Each layer can have any number of neurons in it, and the number of neurons in one layer has no bearing on the number of neurons allowed in subsequent layers. If every neuron in one layer is connected to every neuron in the next layer, the layers are considered to be "fully connected." **Error! Reference source not found.** is an example of a fully-connected, feedforward neural network that contains a single hidden layer.

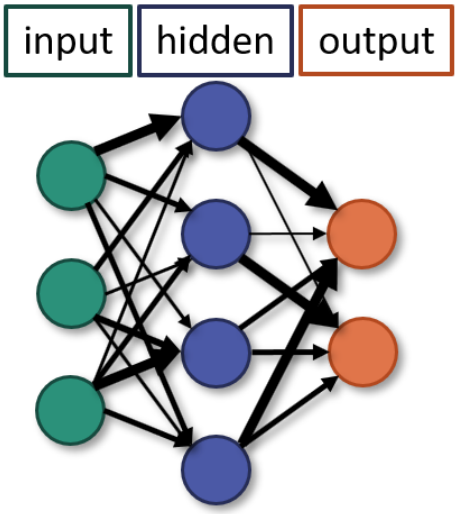


Fig. 48. Example of a fully-connected, feedforward neural network where every neuron is connected to all the neurons in the next layer. The network shown has a single hidden layer, and the arrows represent the weights between the neurons where the thickness represents the strength of the weight.

Perceptron

The perceptron is one of the most basic neural networks and is a type of feedforward neural network [103]. A perceptron is composed of at least one layer of artificial neurons with input and output neurons, and it uses the step function as its activation function. As a result of using the step

function, there are only two possible outputs from each neuron. Depending on the application, the two possible outputs could be *1* or *0*, or they could be *1* and *-1*. The bias *b* is used in the activation function to determine the output state of the perceptron. This bias is independent of the input value and is simply an offset from the origin. The perceptron output is 1 (high) when the neuron's weighted sum is greater than the bias, and the output is *0* or *-1* (low) if the weighted sum is less than the bias. More complex neural networks that incorporate more complicated aspects, such as feedback loops for example, are capable of better classification and more complex problem-solving than a simple perceptron network. However, only perceptron networks are considered in this dissertation.

Circuit Modeling

The circuit simulations were conducted using LTSpice, a SPICE-based software for the simulation of analog devices and circuits [106]. The models for the 22 nm FDSOI and 14 nm bulk FinFET transistor simulations come from [107] and are based on the work from these papers: [108], [109].

Modeling Radiation and Programming

In order to model the effects of radiation and programming of the CTTs within the circuit simulation, additional voltage sources were included to model the threshold voltage shifts resulting from radiation and programming in the CTTs. Fig. 49 shows the LTSpice circuit schematic for a single transistor where the radiation and programming effects on ΔV_{th} are modeled by the parameters *r* and *p* respectively. Since radiation and programming result in threshold voltage shifts in opposite directions, the difference between *p* and *r* is taken.

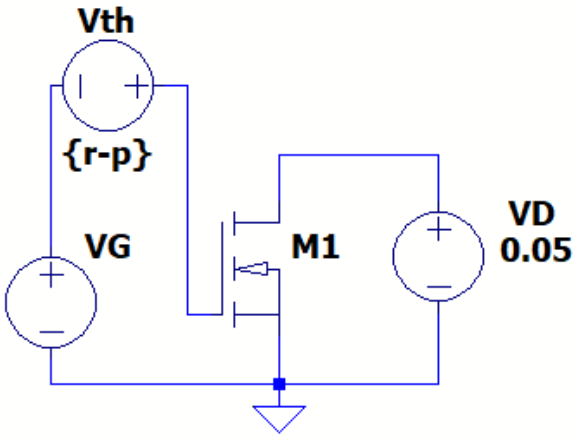


Fig. 49. Inclusion of the effects of radiation and programming on the threshold voltage on the simulation of a CTT.

Fig. 50 shows the simulated I_D-V_G curves for the schematic in Fig. 49 using the dimensions and threshold voltage shifts from Fig. 25, a 22 nm CTT that was programmed before irradiation. The simulation was run with a 22 nm CTT model with L = 20 nm and W = 120 nm, and V_G was swept from -0.4 to 1.2 V. The fresh curve (black) is obtained from setting $p = r = 0$ to simulate no radiation or programming. The programming curve (red) is when p = 100 mV, r = 0 to simulate a device programmed before irradiation. Finally, the irradiation curve is from setting p = 100 mV, r = 150 mV to simulate irradiation to 500 krad(SiO_2) after programming. Both Fig. 25 and Fig. 50 result in threshold voltage shifts of 100 mV for programming and -50 mV for programming and irradiation to 500 krad(SiO_2) with respect to the fresh, pre-programmed curve. Threshold voltage shifts due to radiation and programming are modeled for the rest of this section in this same fashion by applying an additional voltage source to the gate of each transistor.

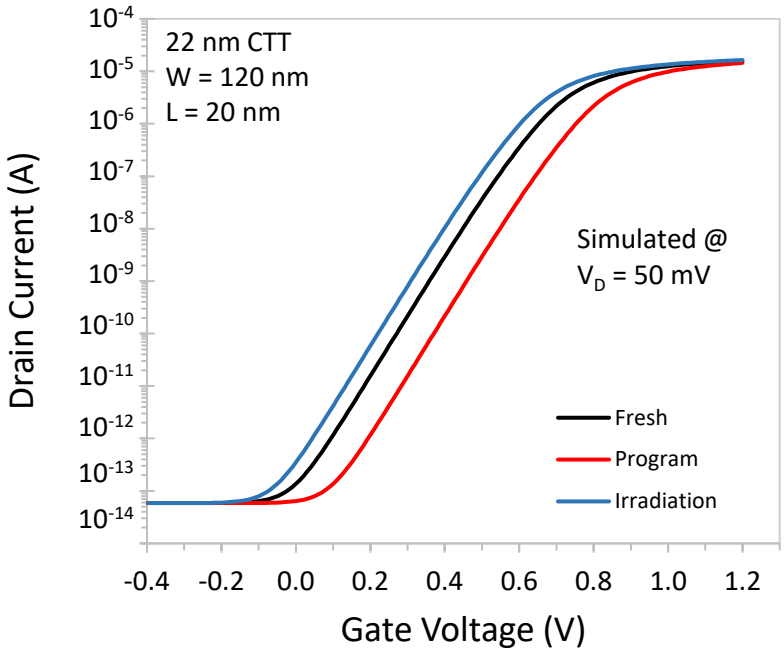


Fig. 50. Simulated results from the model in Fig. 49 for a single 22 nm FDSOI CTT programmed before irradiation to 500 krad(SiO_2) to model the experimental results in Fig. 25.

Modeling an Artificial Neuron

The artificial neuron simulated in this subsection is the three input model presented in [110]. Fig. 51 shows the relation between the artificial neuron with three inputs and weights (left) and its corresponding simplified circuit schematic using twin-cell CTTs (right). The inputs, x_i, are voltages applied to the gates of the transistors; the weights, w_i, are the transconductances of the CTTs; and the output, y, for the whole neuron is the current found by taking the sum of each current difference through the drains, $\sum(\Delta I_D)$. The weight that an individual CTT is programmed to corresponds to a drain current based on the voltages applied to the drain and gate. Since all the experimental data presented applied 50 mV to the drain, all the circuit models use V_D = 50 mV as well. In addition, since the neuron model being studied utilizes a twin-cell architecture, the output current for each weight is determined by the difference between the two drain currents of the CTTs,

shown by the purple box in Fig. 51. The resulting output from the neuron, *y*, is found by summing up the three sets of differential drain currents.

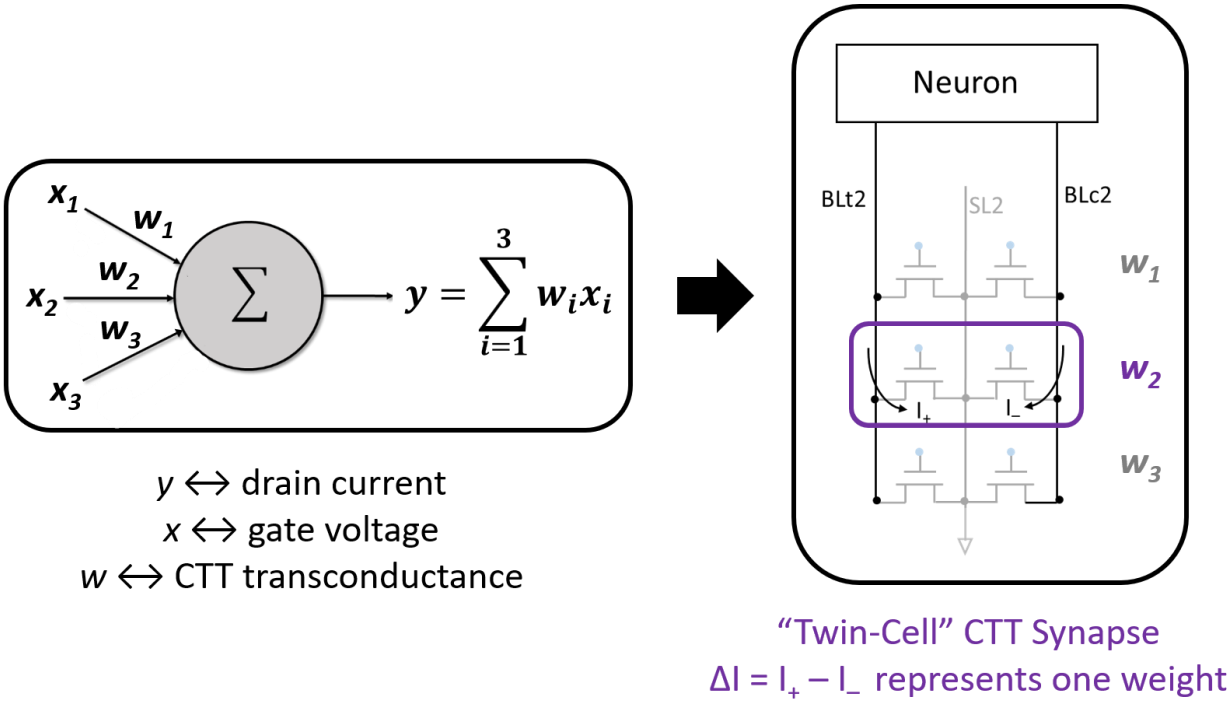


Fig. 51. Implementation of the three-weight artificial neuron (left) in a simplified circuit representation using twin-cell weights (right). The inputs are voltages, the weights are transconductances, and the output is the drain current. *(Adapted from [110].)*

Fig. 52 shows the LTSpice implementation of the simplified circuit diagram shown on the right side of Fig. 51. The six CTTs modeled are all identical and shown in Fig. 52 as *M1t*, … *M3t*, and *M1c*, … *M3c*, where the number indicates which of the three weights, and *t* and *c* indicate the respective true and complement values for the twin-cell. As before, ΔV_{th} due to radiation and programming are modeled by changeable parameters. Radiation is modeled as *r*, and programming modeled by *1t*, *2t*, *3t* for the true bit line transistors and *1c*, *2c*, *3c* for the complement transistors. For these simulations, only one of the two CTTs in a twin-cell is programmed as was done in [28]. The true side is programmed to achieve positive programming values, and the complement side is

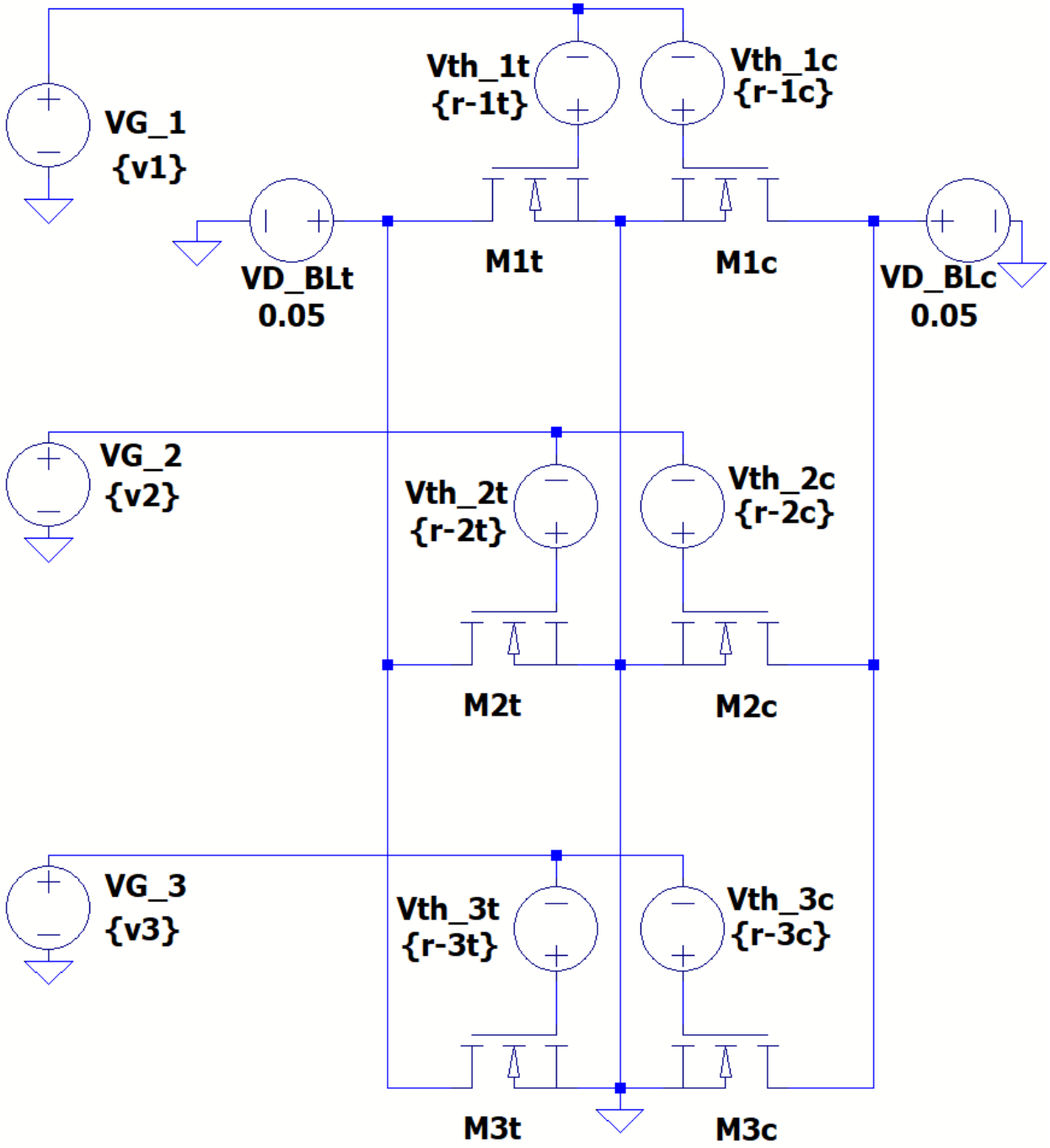


Fig. 52. Implementation of a three-weight, twin-cell CTT neuron in LTSpice.

programmed to achieve negative differential programming threshold voltage shifts. In addition, the applied gate voltage is set by assigning values to *v1*, *v2*, *v3*. V_G = 200 mV in [28], [37], so this is the gate voltage used in simulations as well unless otherwise noted. *VD_BLt* = *VD_BLc* = 50 mV are the drain voltages for the true and complement sides of the twin-cell. Finally, as indicated in Fig. 51 on the right in the purple box, the current output for each synapse is the difference between the left and right drain current. As discussed in Chapter II, the output value for the neuron is determined by the sum of the three differential currents, which is equivalent to the difference between the currents traveling through the two V_D = 50 mV sources.

Modeling Results

Various simulations were performed in LTSpice using the circuit model in Fig. 52 with many different values of programming ΔV_{th}. A representative example is presented here where Fig. 53 shows the simplified circuit representation with the programmed values indicated. The three synaptic weights were programmed to have differential threshold voltage shifts of 80 mV, -130 mV, and 50 mV by setting *1t* = 0.08, *2c* = 0.13, *3t* = 0.05, and *1c* = *2t* = *3c* = 0.

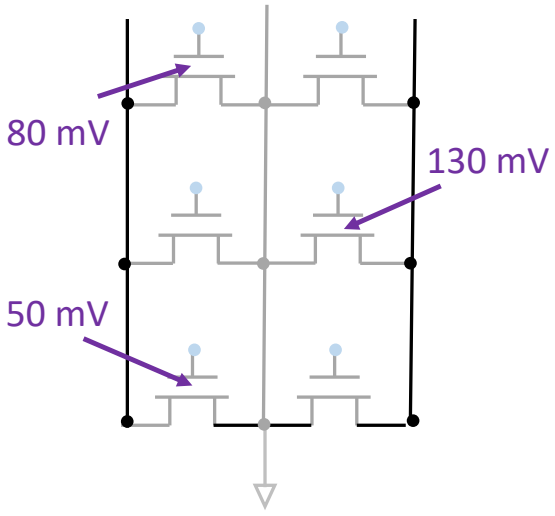


Fig. 53. Simplified circuit schematic showing the threshold voltage shifts to which the CTTs were programmed.

Radiation-induced threshold voltage shifts were also simulated on all six transistors in 50 mV increments. As ΔV_{th} increased, so did the output drain current from the neuron. Therefore, the current gain was calculated to quantify the increase in drain current due to the modeled radiation effect. The gain factor is the ratio between the drain current with radiation-induced ΔV_{th} compared to the modeled drain current with no radiation, or $I_{D(post\text{-}rad)}$ / $I_{D(pre\text{-}rad)}$.

Table III shows the output drain current and current gain for several radiation-induced V_{th} shifts for 22 nm devices operating in the linear region. This gain factor was then plotted as a function of the V_{th} shifts in Fig. 54. As can be seen from the semi-log graph, there is an exponential relationship, and to first order for this linear region, the equation that models this behavior is

$$(gain\ factor, 22\ nm) = e^{0.0252*(\Delta V_{th,rad})}.$$

Table III. Data and calculations resulting from varying the parameter r from Fig. 52 to model TID-induced ΔV_{th} for 22 nm FDSOI devices.

Modeled ΔV_{th} from Radiation (mV)	Output Drain Current (nA)	Drain Current Gain Factor
0	2.38	1
50	8.72	3.7
100	31.19	13
150	107.3	45
200	347.1	146

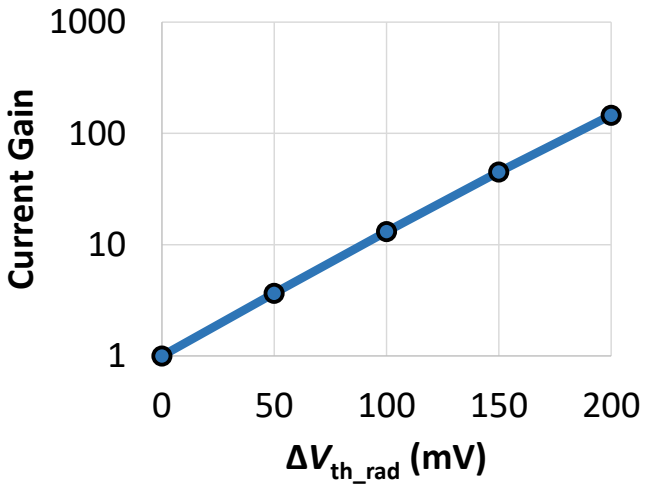


Fig. 54. Output neuron current gain as a function of V_{th} shift resultant from TID for 22 nm FDSOI devices.

An equivalent form of this equation can be derived from the definition of subthreshold slope. Since subthreshold slope (SS) is the slope of a semi-log I_D-V_G plot, it is defined by

$$SS = \frac{\Delta V_G}{\Delta(logI_D)}.$$

Using the logarithmic identity *log(a) – log(b) = log(a/b)*, the fact that the current gain is I_{D2} / I_{D1}, and $\Delta V_G = \Delta V_{th}$, the above equation can be rewritten as

$$(gain\ factor) = I_{D2}/I_{D1} = 10^{\left(\Delta Vth/SS\right)}.$$

For the simulated 22 nm FDSOI devices, the subthreshold slope was extracted to be 88 mV/decade around the operating gate voltage. Plugging this into the above equation and converting to natural logarithms results in the following equation:

$$(gain\ factor, 22\ nm) = e^{ln(10)*\Delta Vth/88} = e^{0.026*\Delta Vth}$$

This is almost exactly the same as the relationship between gain factor and ΔV_{th} found above from the circuit simulations, demonstrating the relationship between current gain of the neuron and ΔV_{th} depends on the subthreshold slope.

Simulating the 14 nm bulk FinFET devices in a similar fashion, a current gain for these devices is also found. The equation for these devices is

$$(gain\ factor, 14\ nm) = e^{0.0256*(\Delta V_{th,rad})}.$$

Neural Network Modeling

The equations found in the previous section show that the drain current gain factor of the neuron scales as a result of irradiation which was modeled in millivolts of threshold voltage shift. Since the weight of the neuron is determined by a linear operation on the drain current, the neuron weight also scales as a result of irradiation. Therefore, scaling the neuron weights in a neural network can be used to model the effect of radiation in the neural network. Using this information in a neural network simulator can therefore give a first look into the possible effects of TID on classification accuracy for 22 nm FDSOI and 14 nm FinFET CTTs implemented in neural networks.

Neural Network Modeling Setup

Keras, written in Python, is a high-level neural network application programming interface which was used to create the simulations [111]. Keras was used in conjunction with TensorFlow [112], an open-source framework for developing machine learning applications. Keras and TensorFlow were used to construct and train a perceptron-based neural network. The neural network was trained to classify handwritten digits between *0* and *9* from the Modified National Institute of Standards and Technology (MNIST) database [57]. Each handwritten digit is a grayscale image composed of 28x28 pixels. Of the 70,000 total digits in the database, 60,000

images are designated for the training data set. The remaining 10,000 digits are for classification using the trained network, and there are approximately 1000 samples of each digit.

The perceptron neural network model created is a multi-layer, feedforward neural network. Each model created has 784 inputs, one for each pixel in the 28x28 image, and 10 outputs, corresponding to the ten possible classifications of *0, 1, 2, … 9*. The number of hidden layers is a tunable parameter, and in these simulations, it ranges from 16 layers to 512 layers in powers of 2. In these simulations, the effect of radiation is modeled by scaling all the neuron values in the trained network to model the current gain. This is from the results of the previous section showing that radiation causes the drain current (which corresponds to the neuron output) to scale as a function of dose. The results of scaling all the neuron values by a factor two in the neural network is shown in Fig. 55. The distribution of neuron values becomes flatter and more spread out with the scaled values. In addition, clipping of the neuron values occurs at the two extrema as the neuron values are scaled more and more. In the simulations that follow, the average of at least five runs is displayed to reduce the effect of variation in training on the results. Each data point collected is the trained neural network's classification accuracy, which is the percentage of total number of correct classifications of the 10,000 handwritten digits.

Modeling Results

Fig. 56 shows the simulation results of classification accuracy as a function of gain factor for different numbers of layers in the perceptron network. A gain factor of 1 indicates the classification accuracy of the neural network before radiation. In general, as the gain factor increases, corresponding to an increase in radiation and therefore threshold voltage shift, the overall classification accuracy decreases before leveling off at higher gain factors. As the number

of layers increases, so does the classification accuracy for any given gain factor. This indicates that the architecture of the neural network plays an important role in determining the accuracy of the neural network.

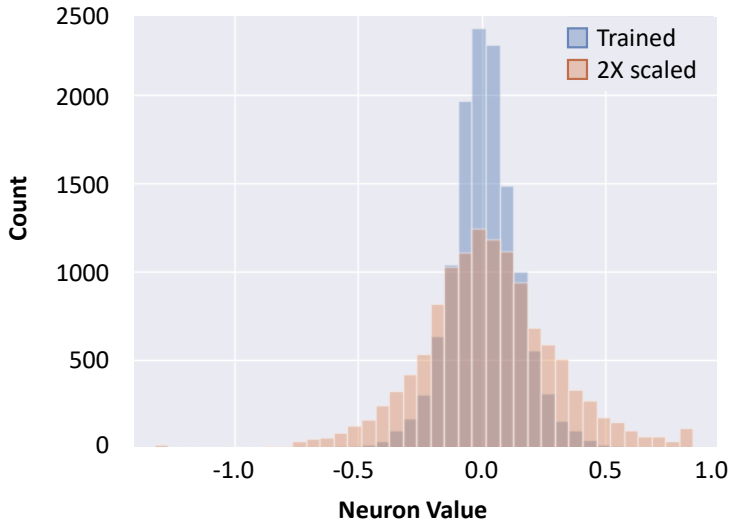


Fig. 55. Neuron value histogram before and after scaling all neuron values by a factor of 2. The original trained network values are shown in blue, while the scaled values (simulating radiation) are shown in orange.

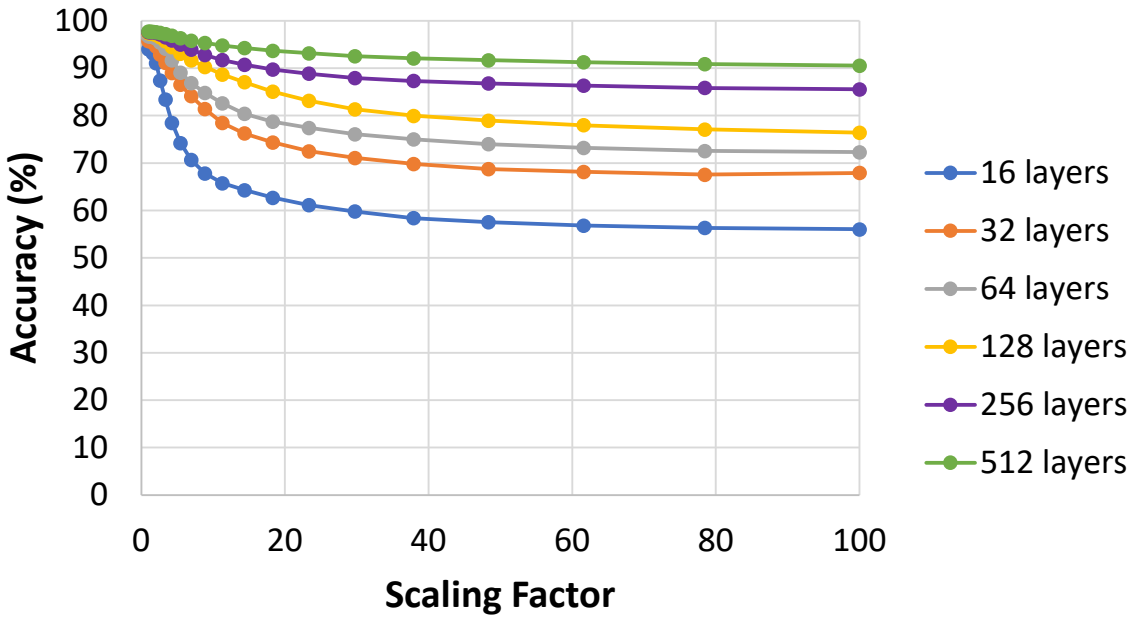


Fig. 56. Accuracy as a function of current gain and number of layers in the perceptron-based neural network simulation.

By using the conversion between current gain and ΔV_{th} as found in the circuit simulation section, Fig. 57 was produced showing the relationship between accuracy and the amount of threshold voltage shift due to TID in CTTs in the neural network simulation. As the number of layers decreases or threshold voltage shift increases, the classification accuracy of the network decreases.

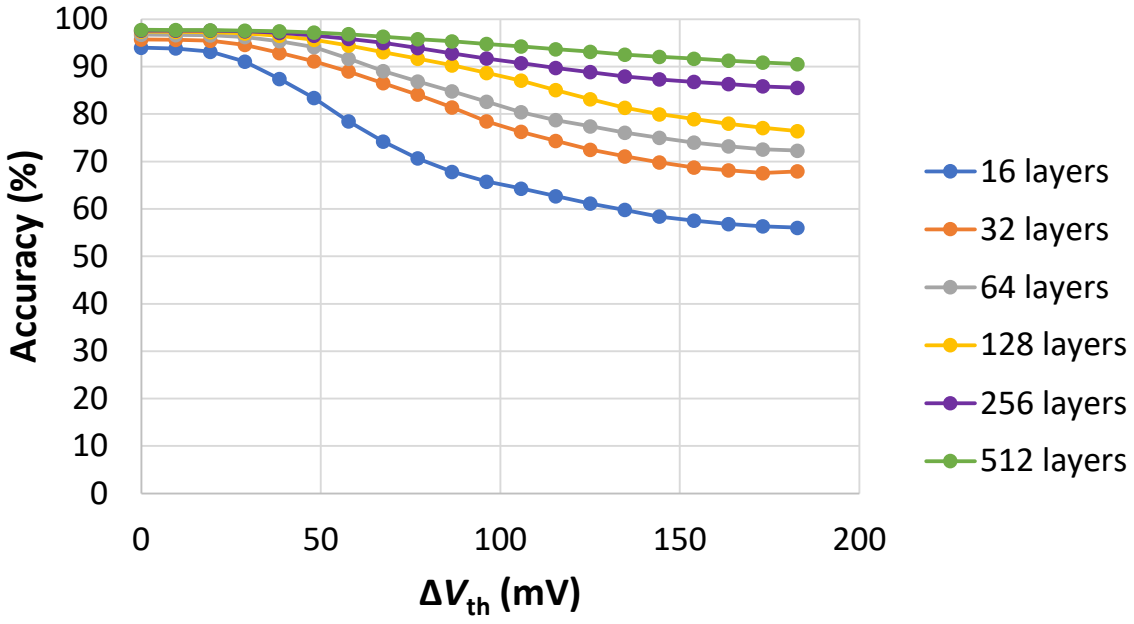


Fig. 57. Accuracy as a function of threshold voltage shift as calculated by the relation between gain factor and ΔV_{th} as determined in the previous section.

Modeling Results: 22 nm FDSOI CTTs

Utilizing the experimental data on 22 nm FDSOI CTTs irradiated without programming (Fig. 29 and Fig. 30), the effect of dose on classification accuracy can be predicted. Fig. 30 shows a decrease in threshold voltage until approximately 200 krad(SiO_2) where $\Delta V_{th} \approx$ -120 mV, and at higher doses, the threshold voltage shift saturates. From these data, a conversion between threshold voltage and dose can be determined. Fig. 58 shows the relationship between dose and classification

accuracy using the results shown in Fig. 56 and Fig. 57. As TID increases until 200 krad(SiO_2), the classification accuracy decreases. However, above 200 krad(SiO_2), an increase in dose does not result in additional threshold voltage shifts; therefore, the accuracy remains constant above 200 krad(SiO_2). In addition, increasing the number of layers in the neural network increases the classification accuracy for the different dose levels. Finally, the initial increase in TID results in a gradual decrease in classification accuracy, and none of the curves result in a classification accuracy less than 60%.

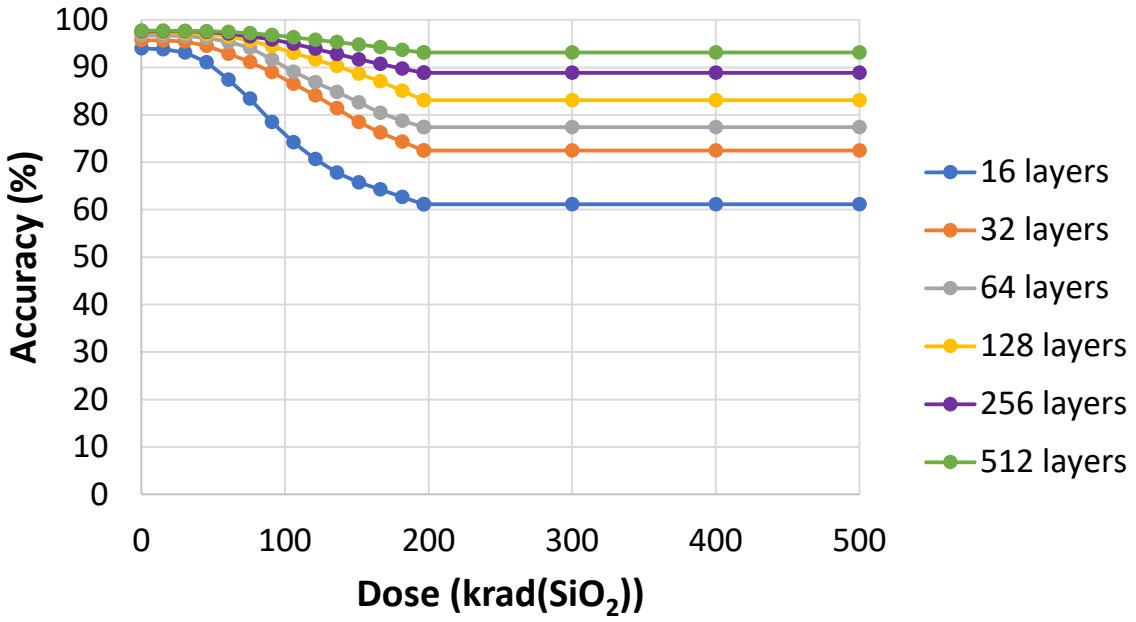


Fig. 58. Accuracy as a function of TID for several numbers of layers in 22 nm FDSOI CTTs in a twin-cell structure.

Modeling Results: 14 nm Bulk FinFET CTTs

A similar method is used to determine the effect of dose on classification accuracy for 14 nm bulk FinFETs, with focus on lower values of gain factor due to the small ΔV_{th} for these devices. The data used for relating threshold voltage shift and dose come from the 40-fin devices of Fig. 36 and Fig. 37, where $\Delta V_{th} \approx$ -10 mV at 500 krad(SiO_2). Fig. 59 shows the resulting accuracy with increasing TID for 14 nm twin-cell CTTs. As the dose increases, the classification accuracy

remains approximately constant. The possible exception is the simulations for 16 layers in which the accuracy decreases slightly. Additionally, as the number of layers increases, so does the accuracy, once again indicating that the architecture of the neural network is an important consideration.

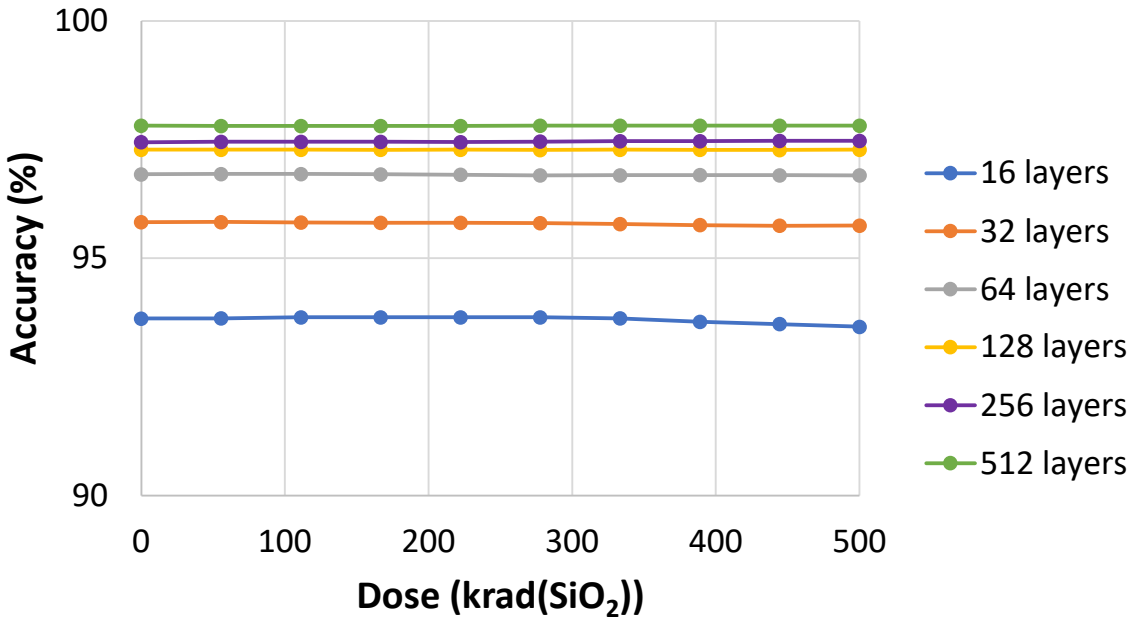


Fig. 59. Accuracy as a function of TID for several numbers of layers in 14 nm bulk FinFET CTTs in a twin-cell structure.

Discussion of Modeling Results

The results of the neural network simulation section as a whole strongly suggest that the TID response of neural networks constructed from CTTs can be architecture-dependent. For the TID-sensitive 22 nm FDSOI CTTs, a neural architecture with 16 layers results in the classification accuracy dropping below 90% after approximately 50 krad(SiO_2). However, increasing the number of layers to 512 results in classification accuracies of ~90% for 500 krad(SiO_2). For all numbers of layers, the classification accuracy degrades with TID and remains higher than 60%. The degradation with TID is gradual as opposed to a sudden failure of the network. As for the TID-

resistant 14 nm bulk FinFET CTTs, the classification accuracy at 500 krad(SiO_2) is dominated by the number of layers in the neural network with essentially no change in accuracy with increasing dose. Therefore, as a possible application, it may be possible to mitigate TID-induced ΔV_{th} degradation by careful consideration and selection of the neural network architecture and structure.

CHAPTER VIII

CONCLUSIONS

This dissertation has evaluated the charge trapping properties and TID responses of 22 nm FDSOI and 14 nm bulk FinFET charge-trap memory transistors. Significant positive threshold-voltage shifts due to electron trapping in the HfO_2 gate dielectric leads to memory windows that are large enough to support non-volatile memory applications in each case. To first order, radiation-induced trapped charge does not interact strongly with programming-induced charges in the gate dielectric in either device type. Hole trapping in the buried oxide leads to a collapse of the memory window in the 22 nm FDSOI devices at doses below 200 krad(SiO_2). 14 nm bulk devices with 2 fins and total effective fin width of 150 nm are minimally affected by TID, but 14 nm bulk devices with 40 fins and total effective fin width of 3 μm show increasing STI-related leakage with increasing TID.

Significant increases in transistor subthreshold stretchout are observed when FDSOI devices are programmed after being irradiated to 500 krad(SiO_2). This is attributed to hot-carrier-induced border-trap generation at the Si/SiO_2 interface and/or charge lateral non-uniformities in the buried oxide. For bulk 14 nm FinFETs, interface-trap generation may occur during continued programming after available electron traps are filled in the HfO_2. The release of hydrogen from the dielectric layer and/or from passivated dopants in the drain may play roles in the observed trap generation, as may multi-carrier-scattering induced multi-vibrational release of hydrogen from passivated Si dangling bonds at the Si/SiO_2 interface. This dissertation shows that CTTs based on HfO_2 dielectrics are promising for use as nonvolatile memory elements in space and other high

radiation environments, but detailed charge trapping and neutralization mechanisms require further investigation, and process optimization and/or refinement of erase cycles are required for 14 nm bulk FinFET CTTs.

A preliminary study of the effects of TID on CTTs implemented in neural networks is also examined. The TID-sensitive 22 nm FDSOI CTTs have a gradual decrease in classification accuracy with increasing TID until 200 krad(SiO_2) after which the classification accuracy remains constant. In contrast, the 14 nm bulk FinFETs, which have minimal V_{th} shifts due to TID, show no change in classification accuracy with increasing dose. However, for both technologies, modifying the neural network architecture by increasing the number of layers in the network results in improved classification accuracy across all doses. These results indicate that it may be possible to mitigate TID-degradation by careful selection of architecture and related parameters in the neural network.

In summary, total-ionizing-dose on charge-trap transistors does not directly alter the memory element of the CTTs. Instead, the response of CTTs to TID is mainly contingent on the nearby insulators such as the BOX and STI and their properties. In addition, enough electron trapping occurs in CTTs for use as memory elements, and TID does not affect the programmability of 14 nm bulk FinFET CTTs. However, in 22 nm FDSOI CTTs, programming does neutralize some of the trapped holes from TID resulting in higher threshold voltage shifts. Finally, in looking ahead to potential implementation of CTTs in neural networks, it is found that the underlying neuromorphic architecture plays a critical role in the accuracy of the neural network after being subjected to TID.

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," *IEEE Electron Device Lett.*, vol. 38, no. 1, pp. 44–47, 2017.

[2] A. Kerber, S. A. Krishnan, and E. A. Cartier, "Voltage ramp stress for bias temperature instability testing of metal-gate/high-*k* stacks," *IEEE Electron Device Lett.*, vol. 30, no. 12, pp. 1347–1349, 2009.

[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," *IEEE Electron Device Lett.*, vol. 37, no. 1, pp. 88–91, 2016.

[4] C.-M. Zhang *et al.*, "GigaRad total ionizing dose and post-irradiation effects on 28 nm bulk MOSFETs," in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, Oct. 2016, pp. 1–4, doi: 10.1109/NSSMIC.2016.8069869.

[5] C.-M. Zhang, F. Jazaeri, G. Borghello, S. Mattiazzo, A. Baschirotto, and C. Enz, "Bias dependence of total ionizing dose effects on 28-nm bulk MOSFETs," in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, Nov. 2018, pp. 1–3, doi: 10.1109/NSSMIC.2018.8824379.

[6] L. Yang *et al.*, "Total ionizing dose response and annealing behavior of bulk nFinFETs with ON-state bias irradiation," *IEEE Trans. Nucl. Sci.*, vol. 65, no. 8, pp. 1503–1510, Aug. 2018, doi: 10.1109/TNS.2018.2827675.

[7] I. Chatterjee *et al.*, "Bias dependence of total-dose effects in bulk FinFETs," *IEEE Trans. Nucl. Sci.*, vol. 60, no. 6, pp. 4476–4482, Dec. 2013, doi: 10.1109/TNS.2013.2287872.

[8] I. Chatterjee *et al.*, "Geometry dependence of total-dose effects in bulk FinFETs," *IEEE Trans. Nucl. Sci.*, vol. 61, no. 6, pp. 2951–2958, Dec. 2014, doi: 10.1109/TNS.2014.2367157.

[9] C.-M. Zhang, F. Jazaeri, G. Borghello, S. Mattiazzo, A. Baschirotto, and C. Enz, "Mobility degradation of 28-nm bulk MOSFETs irradiated to ultrahigh total ionizing doses," in *2018 IEEE International Conference on Integrated Circuits, Technologies and Applications (ICTA)*, Nov. 2018, pp. 162–163, doi: 10.1109/CICTA.2018.8705713.

[10] Y. Kuang *et al.*, "Total dose effects of 28nm FD-SOI CMOS transistors," in *2018 IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, Oct. 2018, pp. 67–69, doi: 10.1109/S3S.2018.8640197.

[11] P. Paillet *et al.*, "Total ionizing dose effects on deca-nanometer fully depleted SOI devices," *IEEE Trans. Nucl. Sci.*, vol. 52, no. 6, pp. 2345–2352, Dec. 2005, doi: 10.1109/TNS.2005.860699.

[12] J. R. Schwank, M. R. Shaneyfelt, P. E. Dodd, J. A. Burns, C. L. Keast, and P. W. Wyatt, "New insights into fully-depleted SOI transistor response after total-dose irradiation," *IEEE Trans. Nucl. Sci.*, vol. 47, no. 3, pp. 604–612, Jun. 2000, doi: 10.1109/23.856487.

[13] M. Gorchichko *et al.*, “Total-ionizing-dose effects and low-frequency noise in 30-nm gate-length bulk and SOI FinFETs sith SiO_2 /HfO_2 gate dielectrics,” *IEEE Trans. Nucl. Sci.*, vol. 67, no. 1, pp. 245–252, Jan. 2020, doi: 10.1109/TNS.2019.2960815.

[14] T. D. Haeffner *et al.*, “Comparison of total-ionizing-dose effects in bulk and SOI FinFETs at 90 and 295 K,” *IEEE Trans. Nucl. Sci.*, vol. 66, no. 6, pp. 911–917, Jun. 2019, doi: 10.1109/TNS.2019.2909720.

[15] M. G. Esposito *et al.*, “Investigating heavy-ion effects on 14-nm process FinFETs: displacement damage versus total ionizing dose,” *IEEE Trans. Nucl. Sci.*, vol. 68, no. 5, pp. 724–732, May 2021, doi: 10.1109/TNS.2021.3072886.

[16] A. Privat *et al.*, “Evidence of interface trap build-up in irradiated 14-nm bulk FinFET technologies,” *IEEE Trans. Nucl. Sci.*, vol. 68, no. 5, pp. 671–676, May 2021, doi: 10.1109/TNS.2021.3065267.

[17] P. J. McWhorter, S. L. Miller, and T. A. Dellin, “Radiation response of SNOS nonvolatile memory transistors,” *IEEE Trans. Nucl. Sci.*, vol. 33, no. 6, pp. 1414–1419, 1986.

[18] E. S. Snyder, P. J. McWhorter, T. A. Dellin, and J. Sweetman, “Radiation response of floating gate EEPROM memory cells,” *IEEE Trans. Nucl. Sci.*, vol. 36, no. 6, pp. 2131–2139, 1989.

[19] S. Gerardin *et al.*, “Radiation effects in Flash memories,” *IEEE Trans. Nucl. Sci.*, vol. 60, no. 3, pp. 1953–1969, 2013.

[20] “22FDX product brief,” *GlobalFoundries*, 2020. https://www.globalfoundries.com/sites/default/files/22fdx-product-brief.pdf (accessed Apr. 08, 2021).

[21] “14LPP product brief,” *GlobalFoundries*, 2020. https://www.globalfoundries.com/sites/default/files/product-briefs/pb-14lpp.pdf (accessed Apr. 08, 2021).

[22] B. Jayaraman *et al.*, “80-kb logic embedded high-*k* charge trap transistor-based multi-time-programmable memory with no added process complexity,” *IEEE J. Solid-State Circuits*, vol. 53, no. 3, pp. 949–960, Mar. 2018, doi: 10.1109/JSSC.2017.2784760.

[23] C. Kothandaraman *et al.*, “Oxygen vacancy traps in hi-*k*/metal gate technologies and their potential for embedded memory applications,” in *2015 IEEE International Reliability Physics Symposium*, Apr. 2015, p. MY.2.1-MY.2.4, doi: 10.1109/IRPS.2015.7112816.

[24] F. Khan *et al.*, “Turning logic transistors into secure, multi-time programmable, embedded non-volatile memory elements for 14 nm FINFET technologies and beyond,” in *2019 Symposium on VLSI Technology*, Jun. 2019, pp. T116–T117, doi: 10.23919/VLSIT.2019.8776510.

[25] H. Hamamura *et al.*, “Electron trapping characteristics and scalability of HfO_2 as a trapping layer in SONOS-type flash memories,” in *2008 IEEE International Reliability Physics Symposium*, Apr. 2008, pp. 412–416, doi: 10.1109/RELPHY.2008.4558920.

[26] A. Kerber and E. A. Cartier, “Reliability challenges for CMOS technology qualifications with hafnium oxide/titanium nitride gate stacks,” *IEEE Trans. Device Mater. Reliab.*, vol. 9, no. 2, pp. 147–162, Jun. 2009, doi: 10.1109/TDMR.2009.2016954.

[27] X. Gu and S. S. Iyer, "Unsupervised learning using charge-trap transistors," *IEEE Electron Device Lett.*, vol. 38, no. 9, pp. 1204–1207, Sep. 2017, doi: 10.1109/LED.2017.2723319.

[28] X. Gu, Z. Wan, and S. S. Iyer, "Charge-trap transistors for CMOS-only analog memory," *IEEE Trans. Electron Devices*, vol. 66, no. 10, pp. 4183–4187, Oct. 2019, doi: 10.1109/TED.2019.2933484.

[29] E. Hunt-Schroeder *et al.*, "14-nm FinFET 1.5 Mb embedded high-*k* charge trap transistor one time programmable memory using differential current sensing," *IEEE Solid-State Circuits Lett.*, vol. 1, no. 12, pp. 233–236, Dec. 2018, doi: 10.1109/LSSC.2019.2899519.

[30] J. Viraraghavan *et al.*, "80Kb 10ns read cycle logic embedded high-k charge trap multi-time-programmable memory scalable to 14nm FIN with no added process complexity," in *2016 IEEE Symposium on VLSI Circuits (VLSI-Circuits)*, Jun. 2016, pp. 14–15, doi: 10.1109/VLSIC.2016.7573462.

[31] F. Khan *et al.*, "Design optimization and modeling of charge trap transistors (CTTs) in 14 nm FinFET technologies," *IEEE Electron Device Lett.*, vol. 40, no. 7, pp. 1100–1103, Jul. 2019, doi: 10.1109/LED.2019.2919871.

[32] Y. Du *et al.*, "An analog neural network computing engine using CMOS-compatible charge-trap-transistor (CTT)," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 38, no. 10, pp. 1811–1819, Oct. 2019, doi: 10.1109/TCAD.2018.2859237.

[33] M. Donato, B. Reagen, L. Pentecost, U. Gupta, D. Brooks, and G.-Y. Wei, "On-chip deep neural network storage with multi-level eNVM," in *2018 55th ACM/ESDA/IEEE Design Automation Conference (DAC)*, Jun. 2018, pp. 77.2:1–6, doi: 10.1109/DAC.2018.8465818.

[34] K. Noda, "Using hot carrier injection for embedded non-volatile memory," *NSCore White Paper*, 2008. http://www.nscore.com/images/WhitePaper_081002.pdf (accessed Apr. 16, 2021).

[35] E. Hunt-Schroeder *et al.*, "Behavioral modeling of a charge trap transistor one time programmable memory," in *2019 IEEE 28th North Atlantic Test Workshop (NATW)*, May 2019, pp. 206–211, doi: 10.1109/NATW.2019.8758743.

[36] S. S. Iyer, "A study of the charge trap transistor (CTT) for post-fab modification of wafers," Los Angeles, CA, 2018. [Online]. Available: https://apps.dtic.mil/sti/pdfs/AD1050743.pdf.

[37] X. Gu, "Charge-trap transistors for neurmorphic computing," University of California, Los Angeles, 2018.

[38] K. Kerris, "Practical dosimetry for radiation hardness testing," *IEEE NSREC Short Course*, pp. I1–I50, 1992.

[39] P. S. Winokur, "Total-dose radiation effects (from the perspective of the experimentalist)," *IEEE NSREC Short Course*, pp. II1–II87, 1992.

[40] A. J. Lelis, T. R. Oldham, and W. M. DeLancey, "Response of interface traps during high-temperature anneals (MOSFETs)," *IEEE Trans. Nucl. Sci.*, vol. 38, no. 6, pp. 1590–1597, 1991, doi: 10.1109/23.124150.

[41] P. J. McWhorter, S. L. Miller, and W. M. Miller, "Modeling the anneal of radiation-induced trapped holes in a varying thermal environment," *IEEE Trans. Nucl. Sci.*, vol. 37, no. 6, pp.

1682–1689, 1990, doi: 10.1109/23.101177.

[42] P. S. Winokur, J. R. Schwank, P. J. McWhorter, P. V. Dressendorfer, and D. C. Turpin, "Correlating the radiation response of MOS capacitors and transistors," *IEEE Trans. Nucl. Sci.*, vol. 31, no. 6, pp. 1453–1460, 1984, doi: 10.1109/TNS.1984.4333529.

[43] P. J. McWhorter and P. S. Winokur, "Simple technique for separating the effects of interface traps and trapped-oxide charge in metal-oxide-semiconductor transistors," *Appl. Phys. Lett.*, vol. 48, no. 2, pp. 133–135, Jan. 1986, doi: 10.1063/1.96974.

[44] P. C. Lacaze and J.-C. Lacroix, *Non-Volatile Memories*. London: ISTE Ltd, 2014.

[45] P. Dimitrakis, *Charge-Trapping Non-Volatile Memories*. Athens: Springer, 2015.

[46] M. H. White, D. A. Adams, and J. Bu, "On the go with SONOS," *IEEE Circuits Devices Mag.*, vol. 16, no. 4, pp. 22–31, Jul. 2000, doi: 10.1109/101.857747.

[47] H. Puchner, P. Ruths, V. Prabhakar, I. Kouznetsov, and S. Geha, "Impact of total ionizing dose on the data retention of a 65 nm SONOS-based NOR Flash," *IEEE Trans. Nucl. Sci.*, vol. 61, no. 6, pp. 3005–3009, Dec. 2014, doi: 10.1109/TNS.2014.2367451.

[48] X. Zhang, A. Huang, Q. Hu, Z. Xiao, and P. K. Chu, "Neuromorphic computing with memristor crossbar," *Phys. status solidi*, vol. 215, no. 13, p. 1700875, Jul. 2018, doi: 10.1002/pssa.201700875.

[49] W. S. Zhao, G. Agnus, V. Derycke, A. Filoramo, J.-P. Bourgoin, and C. Gamrat, "Nanotube devices based crossbar architecture: toward neuromorphic computing," *Nanotechnology*, vol. 21, no. 17, pp. 175–202, Apr. 2010, doi: 10.1088/0957-4484/21/17/175202.

[50] C. Yakopcic, T. M. Taha, G. Subramanyam, R. E. Pino, and S. Rogers, "Analysis of a memristor based 1T1M crossbar architecture," in *The 2011 International Joint Conference on Neural Networks*, Jul. 2011, pp. 3243–3247, doi: 10.1109/IJCNN.2011.6033651.

[51] P. Yao *et al.*, "Face classification using electronic synapses," *Nat. Commun.*, vol. 8, no. 1, Aug. 2017, doi: 10.1038/ncomms15199.

[52] M. Prezioso, F. Merrikh-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, and D. B. Strukov, "Training and operation of an integrated neuromorphic network based on metal-oxide memristors," *Nature*, vol. 521, no. 7550, pp. 61–64, May 2015, doi: 10.1038/nature14441.

[53] S. Agarwal *et al.*, "Resistive memory device requirements for a neural algorithm accelerator," in *2016 International Joint Conference on Neural Networks (IJCNN)*, Jul. 2016, pp. 929–938, doi: 10.1109/IJCNN.2016.7727298.

[54] I. Jung, E. R. Kim, and M. Choi, "Learning nanotechnology through crossbar-based architecture and carbon nanotube(CNT) FETs," in *2011 IEEE International Conference on Microelectronic Systems Education*, Jun. 2011, pp. 60–63, doi: 10.1109/MSE.2011.5937093.

[55] Y. H. Song, S. Y. Park, J. M. Lee, H. J. Yang, and G. H. Kil, "Bidirectional two-terminal switching device for crossbar array architecture," *IEEE Electron Device Lett.*, vol. 32, no. 8, pp. 1023–1025, Aug. 2011, doi: 10.1109/LED.2011.2157452.

[56] R. B. Jacobs-Gedrim *et al.*, "Impact of linearity and write noise of analog resistive memory

devices in a neural algorithm accelerator," in *2017 IEEE International Conference on Rebooting Computing (ICRC)*, Nov. 2017, pp. 160–169, doi: 10.1109/ICRC.2017.8123657.

[57] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[58] C. Xavier, "Fully depleted silicon technology to underlie energy-efficient designs at 28 nm and beyond," *Electrical Design News*, 2013. https://www.edn.com/fully-depleted-silicon-technology-to-underlie-energy-efficient-designs-at-28-nm-and-beyond/ (accessed Apr. 09, 2021).

[59] E. X. Zhang, D. M. Fleetwood, N. D. Pate, R. A. Reed, A. F. Witulski, and R. D. Schrimpf, "Time-domain reflectometry measurements of total-ionizing-dose degradation of *n*MOSFETs," *IEEE Trans. Nucl. Sci.*, vol. 60, no. 6, pp. 4470–4475, Dec. 2013, doi: 10.1109/TNS.2013.2285129.

[60] L. J. Palkuti and J. J. LePage, "X-ray wafer probe for total dose testing," *IEEE Trans. Nucl. Sci.*, vol. 29, no. 6, pp. 1832–1837, 1982, doi: 10.1109/TNS.1982.4336456.

[61] D. M. Fleetwood, "Total ionizing dose effects in MOS and low-dose-rate-sensitive linear-bipolar devices," *IEEE Trans. Nucl. Sci.*, vol. 60, no. 3, pp. 1706–1730, Jun. 2013, doi: 10.1109/TNS.2013.2259260.

[62] R. M. Brewer *et al.*, "Total ionizing dose responses of 22-nm FDSOI and 14-nm bulk FinFET charge-trap transistors," *IEEE Trans. Nucl. Sci.*, vol. 68, no. 5, pp. 677–686, May 2021, doi: 10.1109/TNS.2021.3059594.

[63] I. S. Esqueda *et al.*, "Modeling the radiation response of fully-depleted SOI nchannel MOSFETs," *IEEE Trans. Nucl. Sci.*, vol. 56, no. 4, pp. 2247–2250, Aug. 2009, doi: 10.1109/TNS.2009.2012709.

[64] N. N. Mahatme *et al.*, "Impact of back-gate bias and device geometry on the total ionizing dose response of 1-transistor floating body RAMs," *IEEE Trans. Nucl. Sci.*, vol. 59, no. 6, pp. 2966–2973, Dec. 2012, doi: 10.1109/TNS.2012.2223828.

[65] E. Simoen *et al.*, "Radiation effects in advanced multiple gate and silicon-on-insulator transistors," *IEEE Trans. Nucl. Sci.*, vol. 60, no. 3, pp. 1970–1991, Jun. 2013, doi: 10.1109/TNS.2013.2255313.

[66] D. M. Fleetwood, "Radiation-induced charge neutralization and interface-trap buildup in metal-oxide-semiconductor devices," *J. Appl. Phys.*, vol. 67, no. 1, pp. 580–583, 1990.

[67] J. A. Felix, J. R. Schwank, D. M. Fleetwood, M. R. Shaneyfelt, and E. P. Gusev, "Effects of radiation and charge trapping on the reliability of high- κ gate dielectrics," *Microelectron. Reliab.*, vol. 44, no. 4, pp. 563–575, Apr. 2004, doi: 10.1016/j.microrel.2003.12.005.

[68] X. J. Zhou, D. M. Fleetwood, J. A. Felix, E. P. Gusev, and C. D'Emic, "Bias-temperature instabilities and radiation effects in MOS devices," *IEEE Trans. Nucl. Sci.*, vol. 52, no. 6, pp. 2231–2238, Dec. 2005, doi: 10.1109/TNS.2005.860667.

[69] S. K. Dixit *et al.*, "Radiation induced charge trapping in ultrathin HfO_2-based MOSFETs," *IEEE Trans. Nucl. Sci.*, vol. 54, no. 6, pp. 1883–1890, Dec. 2007, doi: 10.1109/TNS.2007.911423.

[70] G. X. Duan *et al.*, "Bias dependence of total ionizing dose effects in SiGe-SiO_2/HfO_2 pMOS FinFETs," *IEEE Trans. Nucl. Sci.*, vol. 61, no. 6, pp. 2834–2838, Dec. 2014, doi: 10.1109/TNS.2014.2362918.

[71] V. V. Afanas'ev and A. Stesmans, "Stable trapping of electrons and holes in deposited insulating oxides: Al_2O_3, ZrO_2, and HfO_2," *J. Appl. Phys.*, vol. 95, no. 5, pp. 2518–2526, Mar. 2004, doi: 10.1063/1.1641521.

[72] W. L. Warren, J. Kanicki, J. Robertson, E. H. Poindexter, and P. J. McWhorter, "Electron paramagnetic resonance investigation of charge trapping centers in amorphous silicon nitride films," *J. Appl. Phys.*, vol. 74, no. 6, pp. 4034–4046, Sep. 1993, doi: 10.1063/1.355315.

[73] M. Li *et al.*, "Total ionizing dose effects of 55-nm silicon-oxide-nitride-oxide-silicon charge trapping memory in pulse and DC modes," *Chinese Phys. Lett.*, vol. 35, no. 7, p. 078502, Jul. 2018, doi: 10.1088/0256-307X/35/7/078502.

[74] F. El Mamouni *et al.*, "Fin-width dependence of ionizing radiation-induced subthreshold-swing degradation in 100-nm-gate-length FinFETs," *IEEE Trans. Nucl. Sci.*, vol. 56, no. 6, pp. 3250–3255, Dec. 2009, doi: 10.1109/TNS.2009.2034155.

[75] C.-M. Zhang *et al.*, "Characterization and modeling of gigarad-TID-induced drain leakage current of 28-nm bulk MOSFETs," *IEEE Trans. Nucl. Sci.*, vol. 66, no. 1, pp. 38–47, Jan. 2019, doi: 10.1109/TNS.2018.2878105.

[76] M. P. King *et al.*, "Analysis of TID process, geometry, and bias condition dependence in 14-nm FinFETs and implications for RF and SRAM performance," *IEEE Trans. Nucl. Sci.*, vol. 64, no. 1, pp. 285–292, Jan. 2017, doi: 10.1109/TNS.2016.2634538.

[77] D. M. Fleetwood *et al.*, "New insights into radiation-induced oxide-trap charge through thermally-stimulated-current measurement and analysis," *IEEE Trans. Nucl. Sci.*, vol. 39, no. 6, pp. 2192–2203, 1992, doi: 10.1109/23.211421.

[78] D. M. Fleetwood, P. S. Winokur, O. Flament, and J. L. Leray, "Stability of trapped electrons in SiO_2," *Appl. Phys. Lett.*, vol. 74, no. 20, pp. 2969–2971, May 1999, doi: 10.1063/1.123982.

[79] D. M. Fleetwood, "Radiation-induced charge neutralization and interface-trap buildup in metal-oxide-semiconductor devices," *J. Appl. Phys.*, vol. 67, no. 1, pp. 580–583, Jan. 1990, doi: 10.1063/1.345199.

[80] N. S. Saks and M. G. Ancona, "Generation of interface states by ionizing radiation at 80K measured by charge pumping and subthreshold slope techniques," *IEEE Trans. Nucl. Sci.*, vol. 34, no. 6, pp. 1347–1354, 1987, doi: 10.1109/TNS.1987.4337478.

[81] M. A. Xapsos *et al.*, "Separation of radiation induced and process induced lateral non-uniformities MOSFET," *IEEE Trans. Nucl. Sci.*, vol. 37, no. 6, pp. 1677–1681, 1990, doi: 10.1109/23.101176.

[82] D. M. Fleetwood, "Total-ionizing-dose effects, border traps, and 1/ f noise in emerging MOS technologies," *IEEE Trans. Nucl. Sci.*, vol. 67, no. 7, pp. 1216–1240, Jul. 2020, doi: 10.1109/TNS.2020.2971861.

[83] F. B. McLean, "A framework for understanding radiation-induced interface states in SiO_2

MOS structures," *IEEE Trans. Nucl. Sci.*, vol. 27, no. 6, pp. 1651–1657, 1980, doi: 10.1109/TNS.1980.4331084.

[84] D. J. DiMaria, E. Cartier, and D. Arnold, "Impact ionization, trap creation, degradation, and breakdown in silicon dioxide films on silicon," *J. Appl. Phys.*, vol. 73, no. 7, pp. 3367–3384, Apr. 1993, doi: 10.1063/1.352936.

[85] S. N. Rashkeev, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides, "Proton-induced defect generation at the Si-SiO_2 interface," *IEEE Trans. Nucl. Sci.*, vol. 48, no. 6, pp. 2086–2092, 2001, doi: 10.1109/23.983177.

[86] E. Cartier, "Characterization of the hot-electron-induced degradation in thin SiO_2 gate oxides," *Microelectron. Reliab.*, vol. 38, no. 2, pp. 201–211, Feb. 1998, doi: 10.1016/S0026-2714(97)00168-6.

[87] A. G. Marinopoulos, I. Batyrev, X. J. Zhou, R. D. Schrimpf, D. M. Fleetwood, and S. T. Pantelides, "Hydrogen shuttling near Hf-defect complexes in Si/SiO_2/HfO_2 structures," *Appl. Phys. Lett.*, vol. 91, no. 23, p. 233503, Dec. 2007, doi: 10.1063/1.2820380.

[88] K. Hess, B. Tuttle, F. Register, and D. K. Ferry, "Magnitude of the threshold energy for hot electron damage in metal–oxide–semiconductor field effect transistors by hydrogen desorption," *Appl. Phys. Lett.*, vol. 75, no. 20, pp. 3147–3149, Nov. 1999, doi: 10.1063/1.125259.

[89] D. Ferry, "Energy exchange in single-particle electron–electron scattering," *Phys. B Condens. Matter*, vol. 272, no. 1–4, pp. 538–541, Dec. 1999, doi: 10.1016/S0921-4526(99)00335-X.

[90] K. Hess, "Theory of channel hot-carrier degradation in MOSFETs," *Phys. B Condens. Matter*, vol. 272, no. 1–4, pp. 527–531, Dec. 1999, doi: 10.1016/S0921-4526(99)00363-4.

[91] M. Reaz *et al.*, "3-D full-band monte carlo simulation of hot-electron energy distributions in gate-all-around Si nanowire MOSFETs," *IEEE Trans. Electron Devices*, vol. 68, no. 5, pp. 2556–2563, May 2021, doi: 10.1109/TED.2021.3068328.

[92] L. Tsetseris, X. J. Zhou, D. M. Fleetwood, R. D. Schrimpf, and S. T. Pantelides, "Physical mechanisms of negative-bias temperature instability," *Appl. Phys. Lett.*, vol. 86, no. 14, p. 142103, Apr. 2005, doi: 10.1063/1.1897075.

[93] P. J. H. Denteneer, C. G. Van de Walle, and S. T. Pantelides, "Microscopic structure of the hydrogen-phosphorus complex in crystalline silicon," *Phys. Rev. B*, vol. 41, no. 6, pp. 3885–3888, Feb. 1990, doi: 10.1103/PhysRevB.41.3885.

[94] R. E. Paulsen, R. R. Siergiej, M. L. French, and M. H. White, "Observation of near-interface oxide traps with the charge-pumping technique," *IEEE Electron Device Lett.*, vol. 13, no. 12, pp. 627–629, Dec. 1992, doi: 10.1109/55.192866.

[95] S. Habermehl, R. D. Nasby, and M. J. Rightley, "Cycling endurance of silicon–oxide–nitride–oxide–silicon nonvolatile memory stacks prepared with nitrided SiO_2/Si(100) interfaces," *Appl. Phys. Lett.*, vol. 75, no. 8, pp. 1122–1124, Aug. 1999, doi: 10.1063/1.124616.

[96] J. Sawada *et al.*, "TrueNorth ecosystem for brain-inspired computing: scalable systems, software, and applications," in *SC16: International Conference for High Performance*

Computing, Networking, Storage and Analysis, Nov. 2016, pp. 130–141, doi: 10.1109/SC.2016.11.

[97] R. M. Brewer *et al.*, "The impact of proton-induced single events on image classification in a neuromorphic computing architecture," *IEEE Trans. Nucl. Sci.*, vol. 67, no. 1, pp. 108–115, Jan. 2020, doi: 10.1109/TNS.2019.2957477.

[98] I. A. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *J. Microbiol. Methods*, vol. 43, no. 1, pp. 3–31, Dec. 2000, doi: 10.1016/S0167-7012(00)00201-3.

[99] F. Libano *et al.*, "Selective hardening for neural networks in FPGAs," *IEEE Trans. Nucl. Sci.*, vol. 66, no. 1, pp. 216–222, Jan. 2019, doi: 10.1109/TNS.2018.2884460.

[100] G. Li *et al.*, "Understanding error propagation in deep learning neural network (DNN) accelerators and applications," in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, Nov. 2017, pp. 8:1-8:12, doi: 10.1145/3126908.3126964.

[101] C. Torres-Huitzil and B. Girau, "Fault tolerance in neural networks: neural design and hardware implementation," in *2017 International Conference on ReConFigurable Computing and FPGAs (ReConFig)*, Dec. 2017, pp. 58–63, doi: 10.1109/RECONFIG.2017.8279793.

[102] G. W. Burr *et al.*, "Neuromorphic devices and architectures for next-generation cognitive computing," in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2017, pp. 23–26, doi: 10.1109/ISCAS.2017.8050222.

[103] R. Velazco, A. Assoum, N. E. Radi, R. Ecoffet, and X. Botey, "SEU fault tolerance in artificial neural networks," *IEEE Trans. Nucl. Sci.*, vol. 42, no. 6, pp. 1856–1862, 1995, doi: 10.1109/23.489227.

[104] Y. Jiang, C. Yang, J. Na, G. Li, Y. Li, and J. Zhong, "A brief review of neural networks based learning and control and their applications for robots," *Complexity*, vol. 2017, pp. 185–198, 2017, doi: 10.1155/2017/1895897.

[105] R. Reed and R. Marks II, *Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks*. Cambridge, MA: MIT Press, 1999.

[106] M. Engelhardt, "LTSpice." Analog Devices, Wilmington, MA, 2020, [Online]. Available: https://www.analog.com/en/design-center/design-tools-and-calculators/ltspice-simulator.html.

[107] "Latest models," *Nanoscale Integration and Modeling (NIMO) Group*, 2012. http://ptm.asu.edu/ (accessed Apr. 21, 2021).

[108] S. Sinha, G. Yeric, V. Chandra, B. Cline, and Y. Cao, "Exploring sub-20nm FinFET design with predictive technology models," in *Proceedings of the 49th Annual Design Automation Conference on - DAC '12*, 2012, pp. 283–288, doi: 10.1145/2228360.2228414.

[109] Wei Zhao and Yu Cao, "New generation of predictive technology model for sub-45nm design exploration," in *7th International Symposium on Quality Electronic Design (ISQED'06)*, pp. 585–590, doi: 10.1109/ISQED.2006.91.

[110] S. Moran *et al.*, “Impacts of perturbation on a charge trap transistor analog neural network,” in *GOMACTech*, 2020, pp. 1–6.

[111] F. Chollet, “Keras,” 2015. https://keras.io.

[112] M. Abad *et al.*, “TensorFlow: large-scale machine learning on heterogeneous systems,” *TensorFlow White Paper*, 2015. https://www.tensorflow.org/ (accessed Apr. 26, 2021).