# **UCLA UCLA Electronic Theses and Dissertations**

### **Title**

Characterization of Random Telegraph Noise in the Charge Trap Transistor

**Permalink** <https://escholarship.org/uc/item/9904h626>

**Author** Srinivas, Dhruv

**Publication Date** 2023

Peer reviewed|Thesis/dissertation

#### UNIVERSITY OF CALIFORNIA

Los Angeles

Characterization of Random Telegraph Noise in the Charge Trap Transistor

A thesis submitted in partial satisfaction of the requirements for the degree Master of Science in

Electrical and Computer Engineering

by

Dhruv Srinivas

2023

© Copyright by

Dhruv Srinivas

2023

#### ABSTRACT OF THE THESIS

Characterization of Random Telegraph Noise in the Charge Trap Transistor

by

#### Dhruv Srinivas

Master of Science in Electrical and Computer Engineering University of California, Los Angeles, 2023 Professor Subramanian Srikanteswara Iyer, Chair

The charge trap transistor (CTT) is a nonvolatile device used as an analogue neuron for neuromorphic computing. The device used in this thesis is a 22FDX GlobalFoundries device, made with an interfacial layer of silicon oxide and 3.3nm of hafnium oxide. This device operates by applying large gate voltage pulses, a process known as programming (PRG) which traps electrons in the hafnium oxide layer. To reverse this, negative gate voltage pulses are applied in a process known as erase (ERS). This thesis focuses on random telegraph noise (RTN), a type of noise that occurs due to the trapping and detrapping of electrons in this layer and is considered the largest limitation of bit precision in the CTT.

Analysis is performed on the various parameters theorized to affect RTN and as a result, overall CTT stability. Initial experiments were performed using Taguchi's Method, which indicated a correlation between the number of RTN events in a given period and the degree of programming. This experiment also indicated a correlation between the number of RTN events in

ⅱ

a given period and the number of PRG/ERS cycles. After this initial series of experiments were performed, where these individual parameters were studied in further detail. Furthermore, other parameters analyzed in further detail including the magnitude of an ERS event, the duration of measurement, and measurement conditions.

The thesis of Dhruv Srinivas is approved.

Benjamin S. Williams

Sudhakar Pamarti

Subramanian Srikanteswara Iyer, Committee Chair

University of California, Los Angeles

2023

### **TABLE OF CONTENTS**



#### **LIST OF FIGURES**





#### **LIST OF TABLES**



#### ACKNOWLEDGEMENT

I would like to acknowledge my academic advisor, Professor Subramanian Srikanteswara Iyer. Working in the CHIPS lab was a great experience, and Professor Iyer's guidance helped me understand a variety of topics related to this thesis. Over the course of conducting research in the lab, I learned how to approach research systematically. The freedom to approach this problem in a self directed manner helped me grow and mature as a researcher and person.

I would also like to thank both Professor Sudhakar Pamarti and Professor Benjamin S. Williams for agreeing to be committee members, and for the opportunity to attend classes with them. Digital Filter Design and Applied Quantum Mechanics were two of my favorite courses, and I was able to apply some of the knowledge gained in those classes to my research.

I would also like to thank my coworkers and other members of the CHIPS lab. I would like to specifically thank Siyun Qiao for helping me understand the CTT better, Krutikesh Sahoo for helping me better approach experiment design, and Dr. Zhe (Frank) Wan for being my mentor when I first joined the CHIPS lab. Additionally, I would like to extend my gratitude to Steven Moran, Sepideh Nouri, and Dr. Faraz Khan for their work on the CTT. I would also like to thank other student members of CHIPS, Ankit Kuchhangi, Vineeth Harish, Randal Irwin, Guangqi Ouyang, Haoxiang Ren, Golam Sabbir, Henry Sun, and Dr. Yu-Tao Yang.

Lastly, I would like to thank my friends and family. I would like to express my gratitude to my parents Monika Agarwal and Srinivas Sivaprakasm, and sister Ananya Srinivas, for their unconditional love and support. I would also like to thank my friend Dacheng Li for his support and friendship.

ⅸ

# **Chapter 1 Introduction**

# <span id="page-11-1"></span><span id="page-11-0"></span>**1.1 Overview of the Charge Trap Transistor**

The CTT used over the course of this thesis was a 22 nm fully depleted silicon-on-insulator device fabricated by GlobalFoundries. The device's gate oxide is made of an interfacial layer of silicon oxide and a hafnium oxide layer [1] which allows device operation. From measurements performed using a microscope, the thickness of these two layers was determined to be 7 Å for the silicon oxide and 33 Å for the hafnium dioxide, making for an equivalent oxide thickness of 13 Å. There are also other CTT devices, examples include a 14 nm FinFET [2] and a 28 nm FDX device [3].

During device operation, the CTT is meant to function as an analogue neuron for neuromorphic computing [3]. As such, the electrical properties of the device must be adjustable. This is done by applying voltage pulses to the gate, which causes electrons to be trapped within oxygen vacancies in the hafnium oxide [1-4]. The trapped electrons then cause the threshold voltage to increase, which in turn decreases the drain current of the device for a given gate and drain voltage [1-4]. The primary method of PRG is pulse voltage ramp sweep (PVRS) [1-2], a technique where 1.5 V is applied to the gate for 500 μs, after which the gate voltage is increased by 25 mV for every subsequent pulse until the desired inference current  $(I_{inf})$ , the drain current measured when 0.2 V is applied to both the drain and the gate, is reached. Throughout this process, the drain voltage remains constant at 1.2 V.

In order to reverse the effects of programming, ERS is used to partially return to the original state. In this process, a similar PVRS technique is used, but the initial applied gate voltage is -1.5 V and is ramped by -50 mV for each 500 μs pulse. Additionally, the drain voltage is held at 0 V for this process. Unfortunately, returning the device to the original threshold voltage is typically not possible [2], because ERS can sometimes be unable to free electrons from deep traps. This is demonstrated in Figure 1-1, which illustrates the effect of PRG and ERS on the drain current of the device. During PRG, the threshold voltage of the device is increased, leading to a lower drain current. After which ERS is applied, returning the threshold voltage to within  $\sim$ 20 mV of the original value.



Demonstration of PRG and ERS on Drain Current

**Figure 1-1.** Demonstration of PRG and ERS' effect on device current

#### <span id="page-13-0"></span>**1.1.1 Explanation of the Charge Trap Transistor's Use as an Analogue Neuron**

The CTT works due to the increased number of oxygen vacancies in high-K dielectrics when compared to silicon oxide [5-6] [8]. This increase occurs because the melting point of hafnium oxide is much higher than the melting point of silicon oxide. As a result, during the annealing process, more oxygen vacancies remain in the hafnium oxide [5]. This is what allows for the improved trapping properties of high-K dielectrics. The effect of this trapping process is clearest for the device's flatband and threshold voltages [1-7], and it is the threshold voltage shift that allows the CTT to be used as an analogue neuron. Additionally, these trapped charges are usually stable and do not detrap quickly [5] allowing for information to be stored for extended periods of time.

#### <span id="page-13-1"></span>**1.1.2 Factors that Affect Charge Trap Transistor Stability**

There are three main sources of instability in the CTT device which serve to minimize its effectiveness as a memory device. The first cause of instability is 1/f noise, which is a noise characterized by having an amplitude proportional to 1/f [9-11], and is illustrated in Figure 1-2. This noise has two theories behind its cause, the carrier number model  $(ΔN)$  and the mobility fluctuation model  $(\Delta \mu)$ . The former model theorizes that the noise is caused by the rapid trapping and detrapping of interface traps at the silicon-silicon oxide interface [12-13], and the superposition of this creating a 1/f spectrum [12-13]. The mobility fluctuation model, on the other hand, claims that the cause of 1/f noise is the fluctuation of carrier mobility [12]. While 1/f noise was not a focus of this thesis, a further explanation of it is provided in 2.1.1.



**Figure 1-2.** Demonstration of 1/f noise spectra.

The second cause of instability is threshold voltage relaxation, which concerns the device's behavior immediately after PRG and ERS. This behavior is caused by the detrapping of shallow traps in the hafnium oxide shortly after PRG or ERS [4-5]. The detrapping of these traps causes the threshold voltage to rise by a notable amount and causes the current increase shown in Figure 1-3. However, the threshold voltage relaxation only lasts for a finite period of time, and by running the device for this time period and then measuring the drain current, its effect can be reduced. For this reason, the CTT was operated at  $V_{gs} = V_{ds} = 0.6V$  for 20 minutes after PRG and ERS events, after which data was gathered as normal. Additionally, several techniques have been developed to compensate for threshold voltage variation [20], but these were not used in this thesis.



**Figure 1-3.** Simulated demonstration of threshold voltage relaxation.

The last main cause of instability is RTN, the focus of this thesis. This low frequency noise has been measured in silicon-on-insulator devices previously [14-15] and further details are covered in 2.2. This noise can be identified by its characteristic switching behavior where the drain current periodically jumps between different values [14-15] [17]. The low frequency noise is caused by the trapping and detrapping of electrons within traps in the gate oxide [14-15] [17] and can be difficult to model. Additionally, the time constant for trapping an electron can differ from the time constant for detrapping an electron, which can lead to asymmetric noise [14-15] [17]. This noise is demonstrated in Figure 1-4, where there are four simulated RTN events present. The first event begins when an electron detraps and is visually indicated by an increase in current. The second event occurs when an electron traps and is indicated by a decrease in

current. The third event is identical to the second, and the fourth event is identical to the first. It is important to note that determining what trap causes a specific RTN event is impossible, and as a result, Figure 1-4 can represent between one and four distinct traps. It is for this reason that this thesis chose to define one RTN event as a large increase or decrease in current.



**Figure 1-4.** Simulated demonstration of RTN.

# <span id="page-17-0"></span>**1.2 Architecture Using the Charge Trap Transistor**

The CTT has been used in existing neural network architecture [3], as have devices that operate using similar principles [18-19]. The main architecture used when incorporating the CTT has been to form analogue neurons of a convolutional neural network (CNN) [3] [18]. By using a crossbar architecture, the current through each CTT can be modified using the principles described in 1.1.1. The data that passes through this crossbar essentially is treated as if it has performed several multiply-and-accumulate (MAC) operations. At this time, this technology is in the demonstration stages, but it does serve to indicate possible use cases for the CTT and other charge trap architectures.

# <span id="page-18-0"></span>**Chapter 2 Background**

## <span id="page-18-1"></span>**2.1 Noise Sources in the Charge Trap Transistor**

As explained in 1.1.2, two of the main sources of noise in the CTT are 1/f noise and RTN. These noises can be visualized using a histogram of the device's drain current [21] and is demonstrated in Figure 2-1. In this figure, there are three RTN levels, which can be seen in the histogram's three separate peaks. Additionally, 1/f noise causes the spread of the peaks themselves. Figure 2-1 was taken over the course of 1000 seconds and measured on a device which had been programmed to 60 nA. This measurement took place at  $V_{gs} = V_{ds} = 0.2V$ .



**Figure 2-1**. (a) Drain current measurement of a programmed CTT (b) Histogram demonstrating RTN and 1/f Noise

This depiction of RTN behavior is useful for devices which have RTN events which involve multiple electrons. However, it does not directly lend itself to a more quantitative analysis because for certain operational regimes, RTN behavior is not as easily visualized. While in the subthreshold regime  $I_{ds}$  has an exponential dependence on  $V_{th}$  [22], this is not the case in the saturation regime. This makes it much harder to visualize RTN in this regime, because the impact of such events are diminished.

#### <span id="page-19-0"></span>**2.1.1 1/f Noise**

As explained in 1.1.2, the two theories behind the cause of 1/f noise are the  $\Delta N$  and  $\Delta \mu$ models. While both models have evidence supporting them, this thesis makes the assumption that the carrier number model is the more accurate of the two. This is due to irradiation experiments that have been previously performed [24-25]. In these radiation experiments, the number of interface traps increased as did the magnitude of 1/f noise [24-25]. For the ΔN model, 1/f noise is caused by the superposition of the trapping and detrapping of interface traps in the interface between the silicon and silicon oxide [12-13]. These traps are present due to dangling Si-H bonds [23].

#### <span id="page-19-1"></span>**2.1.2 Random Telegraph Noise**

RTN is the primary focus of this thesis and was briefly discussed in 1.1.2. It is caused by some of the limitations of high-K dielectrics. When these oxides were initially being researched, the main limitation for usage was the trappy nature of these materials [6], which is also the main cause for RTN. This creates multiple discrete drain current levels during operation, a fact which can be visualized in Figure 2-1 and in Figure 2-2. The time lag plot in Figure 2-2 is another useful tool of analysis, as it serves an identical purpose to 2-1 and is able to visually indicate the presence of RTN and the number of levels present.



**Figure 2-2.** Time lag plot showing RTN behavior.

The main technique used to analyze RTN in this thesis is covered in Chapter 3. This technique involves modeling the traps as if they are at a fixed point in the hafnium oxide, the reasoning for which will be provided in 2.2.1. From this model, the threshold voltage shift can be calculated and compared to drain current readings. Additionally, by using the filter described in 3.2, RTN events can be directly extracted from the data.

### <span id="page-21-0"></span>**2.2 Physics Behind Random Telegraph Noise**

This section covers the physics behind RTN, specifically delving into three topics which provide an explanation for the filter design covered in Chapter 3. One additional piece of information on RTN is that it can be atemporal, only existing for select periods of time instead of remaining continuous [27]. This necessitated an additional experiment performed in 4.1, which analyzed the effect of the measurement duration on the measured number of RTN events.

#### <span id="page-21-1"></span>**2.2.1 Charge Location**

As explained previously, RTN occurs in the hafnium oxide layer of the gate oxide due to charge trapping and detrapping. For this behavior to occur, electrons must tunnel into and out of these traps. The difficulty of this increases exponentially with distance [28] [38], which means the distribution of charge within the gate oxide is not uniform. Instead, the traps are located close to the substrate of the device [28-29] [38], and this fact is used to create a model of RTN behavior.

#### <span id="page-21-2"></span>**2.2.2 Theorized Parameters Which Affect Random Telegraph Noise**

There are several parameters theorized to affect the measured RTN frequency and amplitude. These include the temperature, duration of measurement, device area, degree of programming, and number of PRG/ERS cycles. However, due to equipment limitations, the temperature and device area were not studied. Over the course of this thesis, additional parameters were also studied, both to ensure the validity of performed experiments, and due to unexpected results obtained during the performance of other experiments. The parameters covered in this section are parameters that had existing evidence for a relationship between themselves and RTN.

Temperature has previously been shown to increase the magnitude of RTN [39], but was not studied in this thesis due to a lack of necessary equipment. Over the course of longer measurement durations, the probability of capturing very low frequency RTN increases [30]. This in turn can cause the measured RTN frequency to increase. Moreover, as the device area decreases, the impact RTN has on the device increases [21] because of the increase in oxide capacitance. Increasing the degree of PRG in a device has been shown to increase the amplitude of RTN [40] and is analyzed in further detail in this thesis. Lastly, the number of PRG/ERS cycles increases RTN frequency through a self-heating effect which creates more defects in the device [30].

## <span id="page-23-0"></span>**2.3 Why Does Random Telegraph Noise Matter?**

The shift in noise due to RTN was believed to be a possible issue for the operation of the CTT. Due to the change in threshold voltage [30] [31] caused by RTN, the device current could change by notable amounts, with analysis at the time showing RTN events changing the drain current by up to 5 nA, which can be seen in Figure 2-1. The main limitation for data storage in the CTT is noise, with reductions in noise allowing for more data to be stored in each CTT. As a result, sources of noise are studied and parameters which may increase this noise are analyzed in order to find methods to decrease said noise.

# <span id="page-24-0"></span>**Chapter 3 Research Methodology**

The initial step to studying RTN in the CTT was to first determine the parameters that notably impacted its frequency and magnitude. In order to do this, Taguchi's Method was used in a process described in 3.1. The design of experiments used in this section included the gate voltage, drain voltage, degree of programming, and number of PRG/ERS cycles. The latter two parameters were expected to affect RTN as explained in 2.2.2, while the gate and drain voltages were expected to have no impact.

## <span id="page-25-0"></span>**3.1 Taguchi's Method**

Taguchi's Method is a method to explore an experiment space quickly when there are many parameters to consider [32-34]. This is done by exploring the space through a number of experiments performed at various levels [32-34]. These levels usually vary based on the resolution desired and the number of parameters explored. For this thesis, RTN was treated as the parameter to optimize in order to denote which parameters notably increased or decreased its frequency or magnitude. Table 3-1 lists all the factors and parameters for these experiments.



**Table 3-1**. List of factors and parameters for the initial experiment group.

As Table 3-1 indicates, there were only four parameters of interest at this point in the study. As the exploration of the design space occurred, other parameters were determined to also have a notable effect on RTN, and this is covered in Chapter 4. The initial Taguchi Table, a table that lists the experiments used, is presented as Table 3-2. This was an L8 table that explored four parameters at two different levels each. These initial experiments required approximately 25 hours of experiments to be performed. Each experiment involved a 20-minute burn in period at

 $V_{gs} = V_{ds} = 0.6V$  to remove threshold voltage relaxation transients, a process briefly described in 1.1.2, and was followed by a 3 hour measurement at the appropriate measurement conditions.



**Table 3-2.** Experiments in the initial L8 Taguchi Table.

The results of these experiments are explained in 3.3. The device used initially had an inference current of 700 nA but was programmed to 580 nA to ensure that ERS could return the device from the lower inference current state to the higher inference current state. As a result, there were some areas of the experiment space that were not covered in this experiment group, but were explored during other experiments.

### <span id="page-27-0"></span>**3.2 Initial Model Design**

In order to quantitatively analyze the number of RTN events, a model for a single RTN event was needed. For this, such an event was defined as a single electron trapping or detrapping within the hafnium oxide. The first assumption made for this model was that the subthreshold current could be modeled using Equation 1 [41] when the device was in subthreshold.

$$
I_d = \mu_{eff} C_{ox} \frac{W}{L} (n-1) \left(\frac{kT}{q}\right)^2 exp\left(\frac{q(V_{gs} - V_{threshold})}{nkT}\right) (1 - exp(-\frac{qV_{ds}}{nkT}))
$$

**Equation 1.** An equation for the subthreshold current of a MOSFET.

For this equation, the values within it were able to be calculated. The electron mobility, oxide capacitance, width, length, and temperature were all known parameters. The only parameter that had to be calculated was the ideality which was subsequently extracted from an  $I_d$ - $V_{gs}$  sweep and an  $I_d$ - $V_{ds}$  sweep.



**Table 3-3.** Parameter values used to model the drain current.

For the initial set of experiments, there were several that were performed in the linear and saturation regimes. As such, those experiments utilized different equations based on the regime of operation. For this thesis, the saturation drain current was modeled using equation 2.

$$
I_d = \mu_{eff} C_{ox} \frac{W}{L} \left[ (V_{gs} - V_{threshold}^2) V_{ds} - \frac{V_{ds}^2}{2} \right]
$$

**Equation 2.** Equation used for the saturation regime.

The next step for modeling RTN was to determine how changes in oxide charge change the threshold voltage. This was a known equation and is Equation 3.

$$
\Delta V_{threshold} = \frac{\Delta Q_{ox}}{C_{ox}}
$$

**Equation 3.** Effect of oxide charge on threshold voltage.

One notable fact about Equation 3, however, is that it assumes the charge is located at the interface between the oxide and substrate. In order to make this a more accurate equation, the charge location must be shifted to within the hafnium oxide. As a result, this charge location was assumed to be 1 nm into the gate oxide, an assumption that is supported by existing theory explained in 2.2.1.



**Figure 3-1.** Location of oxide charges in the CTT.

With the total thickness of the gate oxide for the CTT being 4nm, this resulted in a notable change in the equations. Moreover, in order to use these features as a model, it was necessary to convert from the charge per area, to the total charge. By combining these facts, Equation 4 was able to be obtained.

$$
N = \frac{4}{3} \Delta V_{threshold} \frac{C_{ox} WL}{q}
$$

**Equation 4.** The number of trapped electrons based on the shift in threshold voltage.

From Equation 4, the effect of a single electron trapping event can be determined by inputting  $N = 1$ , which outputs a threshold voltage shift of 0.32 mV. Typically for device operation, the threshold voltage shift is within 100 mV, resulting in a total noise of just 0.32%. Additionally, this also provides us with the number of trapped electrons in the device when it has been programmed by 100mV to be ~310 electrons.

## <span id="page-30-0"></span>**3.3 Initial Filter Design**

With the model now designed, a filter needed to be implemented in order to capture RTN events. This filter was designed to involve a three-step process, with the individual steps being shown in Figure 3-2. Firstly, a low pass filter with a frequency cutoff of 0.1 Hz was applied to remove higher frequency noise components that this thesis was not concerned with. This step used a Chebyshev Type 2 filter due to this filter utilizing a flat passband.

The second step was to convert the drain current to the number of fluctuations that occurred. For this, the threshold voltage was calculated using the appropriate equation for drain current based on the region of operation. Then the threshold voltage was converted to the number of electron fluctuations using Equation 4. The last part of this step was to quantize the electron fluctuations to discrete levels.

The third step of the filter was to remove transients that took place over a period of under 10 seconds. This was to remove artifacts that may have appeared from the initial low pass filter which was applied. Additionally, it served to make the analysis of the data more intuitive from a visual perspective.



**Figure 3-2.** (a) Raw data (b) Filtered data (c) Quantized electron fluctuations (d) Three step filter

output.

## <span id="page-32-0"></span>**3.4 Results from Taguchi's Method**

The results for the L8 table are covered in Figure 3-3 and Figure 3-4. Each of these figures shows the results of the four parameters and it can easily be seen in Figure 3-3 that the inference current and number of PRG/ERS cycles do impact the number of RTN events in three hours.



**Figure 3-3.** (a) Effect of V<sub>ds</sub> on the number of RTN events (b) Effect of V<sub>gs</sub> on the number of RTN events (c) Effect of I<sub>inf</sub> on the number of RTN events (d) Effect of PRG/ERS cycles on the number of RTN events.

These results are in line with the expectations described in 2.2.2. Both increasing the level of programming in the device and the increasing number of PRG/ERS cycles led to an increase in the number of RTN events measured over the course of three-hours. Additionally, there does not appear to be a correlation between the drain or gate voltage and the number of RTN events. The effect of measurement conditions are further analyzed in 4.1, but it is clear that neither of these parameters affect the number of RTN events. However, Figure 3-4 shows that the gate voltage does correlate with the relative magnitude of RTN. The reason for this was explained in previous sections and is due to the difference in Equations 1 and 2, where the drain current's dependence on threshold voltage changes from an exponential dependence to a linear dependence. This is further analyzed in 4.3 where it is shown that RTN events have less of an impact in the linear and saturation regimes.



**Figure 3-4.** (a) Effect of  $V_{ds}$  on the relative magnitude of RTN events (b) Effect of  $V_{gs}$  on the relative magnitude of RTN events (c) Effect of I<sub>inf</sub> on the relative magnitude of RTN events (d)

Effect of PRG/ERS cycles on the relative magnitude of RTN events.

# <span id="page-35-0"></span>**Chapter 4 Results**

## <span id="page-35-1"></span>**4.1 Effect of Measurement Duration on RTN**

Because existing literature did indicate that the measurement duration may affect the measured frequency of RTN [30] , an experiment was conducted to measure this parameter. For this experiment, a device was taken with an initial current of 900 nA and was programmed to 70 nA. This was done in order to increase the number of RTN events, a correlation covered in 3.4 and 4.4.



**Figure 4-1.** Effect of measurement duration on the number of RTN events.

As can be seen, there does appear to be a pattern to RTN in Figure 4-1. This is due to threshold voltage relaxation which occurred over the course of 10 hours. As a result, the degree the device was programmed decreased, leading to the number of RTN events decreasing.

However, there is still considerable variation in the number of RTN events per hour even after accounting for the threshold voltage relaxation. The reason for this is likely due to a combination of statistical variation and atemporal RTN which was explained in 2.2. Due to this five-sigma variation of 14 RTN events per hour, correlations between for RTN will have to exceed a rather large bar. Moreover, if the number of RTN events per 3 hours is used instead, the five-sigma value decreases to 8 events per hour. However, in order to conduct more experiments, the decision was made to use an observation period of one hour.

## <span id="page-37-0"></span>**4.2 Effect of Measurement Conditions on RTN**

#### <span id="page-37-1"></span>**4.2.1 Experiment Design**

While the results in Figure 3-3 do not indicate a correlation between measurement conditions and RTN, an additional analysis was performed in further detail. For this experiment, the device was measured for one hour at a total of 16 different measurement conditions. These experiments were conducted immediately after each other in order to minimize external interference, and the temperature of the device remained constant throughout the experiment.

		Experiment Gate Voltage (V) Drain Voltage (V)
1	0.2	0.2
2	0.2	0.4
3	0.2	0.6
4	0.2	0.8
5	0.4	0.2
6	0.4	0.4
7	0.4	0.6
8	0.4	0.8
9	0.6	0.2
10	0.6	0.4
11	0.6	0.6
12	0.6	0.8
13	0.8	0.2
14	0.8	0.4
15	0.8	0.6
16	0.8	0.8

**Table 4-1.** Experiment table for studying RTN's dependence on measurement conditions.

#### <span id="page-38-0"></span>**4.2.2 Experiment Results**

Once the experiments were conducted, the data was analyzed by averaging the number of RTN events according to the drain or gate voltage. The results of this experiment are shown in Figure 4-2 and Figure 4-3.



**Figure 4-2.** Effect of drain voltage on the number of RTN events.

From Figure 4-2, it is clear that there is no correlation between the drain voltage and number of RTN events. This is as expected based on both Figure 3-3 and no such correlation appearing in existing literature. Additionally, Figure 4-3 shows that the gate voltage also does not affect the number of RTN events. Overall, this validates the decision to conduct later experiments at  $V_{gs} = V_{ds} = 0.2V$ , in order to simplify further experiments so that they may be

performed quicker. This experiment was also repeated on a heavily programmed device, which gave the same conclusion.



**Figure 4-3**. Effect of gate voltage on the number of RTN events.

## <span id="page-40-0"></span>**4.3 RTN in the Saturation Regime**

While 4.2 showed measurement conditions did not have any significant impact on RTN, 3.4 did demonstrate that for devices in saturation, the relative magnitude of RTN decreases substantially. Due to this, an analysis was performed in order to quantify the magnitude decrease. This is demonstrated in Figure 4-4. The experiment that obtained this figure used a device that had been subject to 10 PRG/ERS cycles and had been programmed to an inference current of 65 nA. This device was chosen in order to maximize the number of RTN events. After which a 20-minute burn in period occurred at  $V_{gs} = V_{ds} = 0.6V$ , then the device was measured for one hour at the indicated voltages.



**Figure 4-4.** (a) Raw data in saturation (b) Filtered data in saturation (c) Quantized electron fluctuations in saturation (d) Three step filter output in saturation.

The shift in drain current due to a single RTN event in saturation was determined to be 0.45% for the measurement conditions depicted in Figure 4-4. The same event in the subthreshold regime changes the drain current by 0.81%. The deviation in subthreshold can also be calculated theoretically from Equations 1. By splitting the threshold voltage into a constant and a term due to RTN  $(V_{threshold}(t) = V_{threshold, avg} + \Delta V_{threshold}(t))$ , we can isolate the term that varies with time into Equation 5.

$$
\frac{\Delta I_d}{I_d} = exp(\frac{q\Delta V_{threshold}}{nkT})
$$

**Equation 5.** Equation used to calculate the change in current due to an RTN event.

By using the values in Table 3-3, we find that the drain current change due to a single RTN event is 0.80%. Moreover, because  $\Delta V_{threshold} \ll \frac{nkT}{q}$ , the change in drain current in the  $\overline{q}$ subthreshold regime is approximately linear for all RTN events that appeared during experiments.

## <span id="page-42-0"></span>**4.4 Effect of Programming on RTN**

Following the experiments performed in 4.1 and 4.2, analysis pivoted to better understanding the effect of various parameters based on the findings in 3.4. The first two experiments conducted both covered the effects of programming on RTN.

#### <span id="page-42-1"></span>**4.4.1 Experiment Design**

The initial experiment was performed in order to coarsely understand the effect of PRG, and involved programming the device to 300 nA. While it would have been ideal to study the full range of inference current available, the device at the time was intended to also study the effect of PRG/ERS cycles, which required the device to be at a level which ERS could return to. The series of experiments performed are shown in Table 4-2, and involve programming the device by increments of 75 nA.



**Table 4-2.** Experiment table for the first analysis of PRG's effect on RTN.

After each programming event, a 20-minute burn in period was performed at  $V_{gs} = V_{ds} = 0.6V$  to remove any transients, and the device was measured at inference for three hours. The was chosen was to guide 4.2 in determining the inference current condition needed to maximize the number of RTN events for the experiment. As a result, the observation period for each measurement was set to three hours.

The second experiment sought a more complete understanding of the effect on programming, so it used a virgin device, a device that had never been programmed, so that it could analyze the full range of inference currents possible. The device chosen for this experiment had an inference current of 780 nA and was programmed in increments of 100 nA after which the same burn in period was performed as the first experiment in this section. Following this the device was measured for one hour at inference. Table 4-3 shows the experiment table used for this section.



**Table 4-3.** Experiment table for the second analysis of PRG's effect on RTN.

#### <span id="page-43-0"></span>**4.4.2 Experiment Results**

Figure 4-5 shows the result of the first experiment. It shows a clear increase in the number of RTN events that occurred over the course of three hours as the inference current decreased. The difference in this value amounts to 16 RTN events per three hours. Notably, this is higher than the threshold of 8 RTN events per three hours which was determined in 4.1.



**Figure 4-5.** Coarse analysis of the degree of PRG on RTN

The results of the second set of experiments are shown in Figure 4-6 and agree with the findings of Figure 4-5. Additionally, this experiment extends the full range of possible inference currents that would be utilized by the CTT. Moreover, the difference in RTN events per hour over the course of this experiment exceeds the variation of 14 RTN events per hour determined in 4.1, indicating a statistically significant correlation between the degree of programming and number of RTN events.



**Figure 4-6.** Effect of programming on RTN.

## <span id="page-46-0"></span>**4.5 Effect of Erase on RTN**

The degree of ERS was not initially a parameter of interest in this thesis. However, after performing the first experiment in 4.4.1, an attempt was made to analyze ERS/PRG cycles in order to better guide the experiment in 4.1. This resulted in unexpected results, which led to a dedicated analysis of the effect of ERS on RTN. The analysis was performed on a device which was erased in stages until it reached higher inference current levels. Afterwards, the device was programmed back to  $\sim$ 75 nA. The amount the device was erased was varied which gave the results shown in Figure 4-7.



**Figure 4-7**. Effect of ERS on the number of RTN events.

For larger ERS events, there was a decrease in the number of RTN events in one hour, a result which was expected based on the results of 4.4, which showed that a higher inference

current was correlated with a decrease in RTN events. This change in RTN events for large ERS events is above the threshold found in 4.1, whereas the change in RTN events for small ERS events is below that threshold. This means that if an ERS event is small, then the number of RTN events will remain approximately the same. Due to this result, the experiment performed in 4-6 was modified to make use of a larger difference in the two inference levels.

## <span id="page-48-0"></span>**4.6 Effect of PRG/ERS Cycles on RTN**

The last parameter analyzed over the course of this thesis was the number of PRG/ERS cycles. As explained in 2.2.2, this was expected to increase the number of RTN events due to the increased number of self-heating cycles. The experiment took a device which had a virgin inference current of 800 nA and the device was programmed to 500 nA. This value was chosen to balance between the two limitations, those being requiring a minimum ERS level which 4.5 showed was needed, and the fact that PRG is not fully reversible. As a result, this led to the device alternating between 500 nA and 60 nA. This was extended further in later experiments in order to ensure this pattern held at higher numbers of cycles.



Demonstration of PRG/ERS Cycles on RTN Events

**Figure 4-8.** Effect of PRG/ERS cycles on the number of RTN events.

Figure 4-8 shows the results of this experiment, where a correlation can be seen between the number of cycles and number of RTN events. Both the high and low inference current show a greater number of RTN events in one hour than the five-sigma threshold determined in 4.1, meaning there is statistical support for this correlation. Additionally, a visual analysis of Figure 4-8 shows that the first PRG cycle increased RTN more than subsequent events, indicating there are some notable changes that occur in the oxide.

## <span id="page-50-0"></span>**4.7 Effect of Programming Vds on RTN**

While the research on this thesis was ongoing, another student in the CHIPS lab experimented with applying 2 V to the drain during PRG [20]. During this process, they found that there was seemingly a decrease in noise [20]. Therefore, another experiment was designed to determine if this was the case. Using the data in 4.4.2 as a point of comparison, the experiment consisted of taking a virgin device and programming it repeatedly while applying 2 V to the drain.



**Figure 4-9**. Effect of Programming  $V_{ds}$  on the number of RTN Events in one hour.

The data gathered from this experiment is shown in Figure 4-9. While there was visually a decrease in the number of RTN events in one hour that did remain consistent, this value did not exceed the five-sigma threshold. However, if one compares multiple data points at once, then this claim becomes more accurate. As a result, it does appear that using  $V_{ds} = 2V$  decreasing the number of RTN events per hour.

# <span id="page-52-0"></span>**Chapter 5 Conclusion**

# <span id="page-52-1"></span>**5.1 Optimal Use of the CTT**

Based on the experiments performed in Chapter 4, there are four parameters that influence the frequency of RTN: the level the device has been programmed to, magnitude of ERS if that was performed last, the number of PRG/ERS cycles, and drain voltage during PRG. Additionally, the impact of RTN on the drain current can be reduced by operating the device in saturation. As a result, there are five conditions to meet for device operation in order to reduce the influence of RTN on the device. Firstly, the device should be programmed to a minimal degree and operated at higher I<sub>inf</sub>. Secondly, the device should be subject to the minimum number of PRG/ERS cycles. Thirdly, ERS events should occur over the largest possible step. Fourth, the device should be operated in the saturation regime. And lastly the drain voltage during PRG should be 2 V.

## <span id="page-53-0"></span>**5.2 Quantitative Analysis on the Effect of RTN on the CTT**

The threshold voltage change due to a single RTN event was determined to be 0.32 mV in 3.2. For standard operation, a single RTN event was determined to not have a significant impact on inference. This was briefly explained in 4.6, where the drain current shift for a single event was experimentally determined to be 0.81%. This is a negligible amount, and is usually smaller than other noise sources in the device. For a superposition of RTN events, or RTN events which involve multiple electrons, the conclusion remains the same. During the course of this thesis the largest drain current variation as a result of RTN was  $\pm 6.3$ %. When the device was operated in saturation, the largest drain current deviation due to RTN was  $\pm$ 3.2%. This is a small enough value to conclude that RTN does not notably impact the operation of the CTT.

# <span id="page-54-0"></span>**5.3 Limitations of This Research**

The primary limitation of this research was that while temperature control was present and the device kept at a constant 295 K, changing the temperature to lower or higher values was not possible. As explained in 2.2.2, RTN may have had a correlation with temperature, which would mean an analysis of it would lead to a more complete understanding of this noise. Moreover, temperature plays a significant role in the performance of semiconductors and transistors, so changing the temperature of the transistor may affect some of the correlations covered in Chapter 4. Additionally, the work performed in this thesis was limited by the length of time available to conduct experiments. If the experiments could be run for lengthier periods the threshold for a five-sigma variation would decrease by a moderate amount. Furthermore, studies could also be performed on the stability of RTN over the course of hundreds of hours, rather than the tens of hours used in this thesis. Moreover, increasing the number of devices used would improve the ability to quantify the various observed patterns.

There are also two other sets of experiments that could be performed in order to gain a better understanding of the CTT. Based on the results in 4.5 where ERS only reduced the number of RTN events when above a certain threshold, it may be possible to find the activation energy of oxide traps by repeatedly testing different ERS thresholds. However, this is currently speculative, and further literature review and experiments would need to be conducted to determine if it is possible. The second set of experiments would be related to the voltages used during PRG. The results in 4.7 showed that using a higher drain voltage during PRG can decrease the number of RTN events, likely due to trapping electrons in deeper traps that are less likely to experience RTN behavior. Therefore, attempting to find better methods of PRG that target deeper traps would be a logical extension of the work done in this thesis.

# <span id="page-55-0"></span>**Chapter 6 References**

[1] X. Gu and S. S. Iyer, "Unsupervised Learning Using Charge-Trap Transistors," in *IEEE*

*Electron Device Letters*, vol. 38, no. 9, pp. 1204-1207, Sept. 2017, doi:

10.1109/LED.2017.2723319.

[2] F. Khan, E. Cartier, J. C. S. Woo and S. S. Iyer, "Charge Trap Transistor (CTT): An

Embedded Fully Logic-Compatible Multiple-Time Programmable Non-Volatile Memory

Element for High-*k*-Metal-Gate CMOS Technologies," in *IEEE Electron Device Letters*, vol. 38,

no. 1, pp. 44-47, Jan. 2017, doi: 10.1109/LED.2016.2633490.

[3] Y. Du *et al*., "An Analog Neural Network Computing Engine Using CMOS-Compatible Charge-Trap-Transistor (CTT)," in *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 38, no. 10, pp. 1811-1819, Oct. 2019, doi:

10.1109/TCAD.2018.2859237.

[4] X. Gu, Z. Wan and S. S. Iyer, "Charge-Trap Transistors for CMOS-Only Analog Memory," in *IEEE Transactions on Electron Devices*, vol. 66, no. 10, pp. 4183-4187, Oct. 2019, doi:

10.1109/TED.2019.2933484.

[5] Chowdhury, N. A., and D. Misra. "Charge Trapping at Deep States in HF–Silicate Based High-κ Gate Dielectrics." *Journal of The Electrochemical Society*, vol. 154, no. 2, 19 Dec. 2007, pp. G30–G37., <https://doi.org/10.1149/1.2402989>.

[6] M. Houssa, A. Stesmans, M. Naili, and M. M. Heyns, "Charge trapping in very thin high-permittivity gate dielectric layers," Appl. Phys. Lett., vol. 77, 2000, Art. no. 1381.

[7] S. Zafar, A. Callegari, E. Gusev and M. V. Fischetti, "Charge trapping in high k gate dielectric stacks," *Digest. International Electron Devices Meeting,*, San Francisco, CA, USA, 2002, pp. 517-520, doi: 10.1109/IEDM.2002.1175893.

[8] K. Tse, D. Liu, K. Xiong, J. Robertson, Oxygen vacancies in high-k oxides, Microelectronic Engineering 84 (9–10) (2007.) 2028–2031

[9] E. Simoen, A. Mercha, C. Claeys, N. Lukyanchikova and N. Garbar, "Critical discussion of the front-back gate coupling effect on the low-frequency noise in fully depleted SOI MOSFETs," in *IEEE Transactions on Electron Devices*, vol. 51, no. 6, pp. 1008-1016, June 2004, doi:

10.1109/TED.2004.828159.

[10] Jen-tai Hsu, J. Wang, J. Woo and C. R. Viswanathan, "Flicker noise in thin film fully depleted SOI MOSFETs," *1991 IEEE International SOI Conference Proceedings*, Vail Valley, CO, USA, 1991, pp. 30-31, doi: 10.1109/SOI.1991.162841.

[11] Dieudonné, F., Haendler, S., Jomaah, J., & Balestra, F. (2003). Low frequency noise in 0.12 μm partially and fully depleted SOI technology. *Microelectronics Reliability*, *43*(2), 243–248. https://doi.org/10.1016/s0026-2714(02)00279-2

[12 ]L. K. J. Vandamme, Xiaosong Li and D. Rigaud, "1/f noise in MOS devices, mobility or number fluctuations?," in *IEEE Transactions on Electron Devices*, vol. 41, no. 11, pp. 1936-1945, Nov. 1994, doi: 10.1109/16.333809.

[13] C. H. Suh, "Carrier number fluctuation model of 1/f noise in a semiconductor device," in *IEEE Transactions on Electron Devices*, vol. 34, no. 4, pp. 906-910, April 1987, doi: 10.1109/T-ED.1987.23014.

[14] J. Nishimura, T. Saraya and T. Hiramoto, "Statistical comparison of random telegraph noise (RTN) in bulk and fully depleted SOI MOSFETs," *Ulis 2011 Ultimate Integration on Silicon*, Cork, Ireland, 2011, pp. 1-4, doi: 10.1109/ULIS.2011.5757959.

[15] C. G. Theodorou *et al*., "New LFN and RTN analysis methodology in 28 and 14nm FD-SOI MOSFETs," *2015 IEEE International Reliability Physics Symposium*, Monterey, CA, USA,

2015, pp. XT.1.1-XT.1.6, doi: 10.1109/IRPS.2015.7112833.

[16] Theodorou, C., Ghibaudo, G. (2020). Noise and Fluctuations in Fully Depleted

Silicon-on-Insulator MOSFETs. In: Grasser, T. (eds) Noise in Nanoscale Semiconductor

Devices. Springer, Cham. https://doi.org/10.1007/978-3-030-37500-3\_2

[17] Puglisi, F M. (2020) Noise in Resistive Random Access

Memory Devices. In: Grasser, T. (eds) Noise in Nanoscale Semiconductor Devices. Springer, Cham. [https://doi.org/10.1007/978-3-030-37500-3\\_2](https://doi.org/10.1007/978-3-030-37500-3_2)

[18] Moran, Steven. "Analog in-Memory Multiply-and-Accumulate Engine Fabricated in 22nm FDSOI Technology." EScholarship, 2022.

[19] H. -S. Choi, H. Kim, J. -H. Lee, B. -G. Park and Y. Kim, "AND Flash Array Based on

Charge Trap Flash for Implementation of Convolutional Neural Networks," in *IEEE Electron*

*Device Letters*, vol. 41, no. 11, pp. 1653-1656, Nov. 2020, doi: 10.1109/LED.2020.3025587.

[20] Qiao, Siyun. "Characterization of the Charge-Trap Transistor for Analog in-Memory

Computing." University of California - Los Angeles, 2022.

[21] Frank, Dave. "International Reliability Physics Symposium." *Random Telegraph Noise – Measurement, Analysis, and Consequences*.

[22] Hu, Chenming. *Modern Semiconductor Devices for Integrated Circuits*. Pearson Prentice Hall, 2010.

[23] S. N. Rashkeev, D. M. Fleetwood, R. D. Schrimpf and S. T. Pantelides, "Effects of hydrogen motion on interface trap formation and annealing," in *IEEE Transactions on Nuclear Science*, vol. 51, no. 6, pp. 3158-3165, Dec. 2004, doi: 10.1109/TNS.2004.839202.

[24] Fleetwood, D M, Winokur, P S, Reber, Jr, R A, Meisenheimer, T L, Schwank, J R,

Shaneyfelt, M R, and Riewe, L C. *Effects of oxide traps, interface traps, and border traps'' on metal-oxide-semiconductor devices*. United States: N. p., 1993. Web. doi:10.1063/1.353777.

[25] D. M. Fleetwood, M. J. Johnson, T. L. Meisenheimer, P. S. Winokur, W. L. Warren and S. C. Witczak, "1/f noise, hydrogen transport, and latent interface-trap buildup in irradiated MOS devices," in *IEEE Transactions on Nuclear Science*, vol. 44, no. 6, pp. 1810-1817, Dec. 1997, doi: 10.1109/23.658947.

[26] T. Nagumo, K. Takeuchi, S. Yokogawa, K. Imai and Y. Hayashi, "New analysis methods for comprehensive understanding of Random Telegraph Noise," *2009 IEEE International Electron Devices Meeting (IEDM)*, Baltimore, MD, USA, 2009, pp. 1-4, doi:

10.1109/IEDM.2009.5424230.

[27] Tibor Grasser, Stochastic charge trapping in oxides: From random telegraph noise to bias temperature instabilities, Microelectronics Reliability, Volume 52, Issue 1, 2012, Pages 39-70, ISSN 0026-2714, https://doi.org/10.1016/j.microrel.2011.09.002.

[28] G. Zhang, X. -P. Wang, W. J. Yoo and M. -F. Li, "Spatial Distribution of Charge Traps in a SONOS-Type Flash Memory Using a High-*k* Trapping Layer," in *IEEE Transactions on Electron Devices*, vol. 54, no. 12, pp. 3317-3324, Dec. 2007, doi: 10.1109/TED.2007.908888.

[29] Assaf Shappir, David Levy, Yosi Shacham-Diamand, Eli Lusky, Ilan Bloom, Boaz Eitan, Spatial characterization of localized charge trapping and charge redistribution in the NROM device, Solid-State Electronics, Volume 48, Issue 9, 2004, Pages 1489-1495, ISSN 0038-1101, https://doi.org/10.1016/j.sse.2004.03.013.

[30] K. Fukuda, Y. Shimizu, K. Amemiya, M. Kamoshida and C. Hu, "Random telegraph noise in flash memories - model and technology scaling," *2007 IEEE International Electron Devices Meeting*, Washington, DC, USA, 2007, pp. 169-172, doi: 10.1109/IEDM.2007.4418893.

[31] Naoki Tega *et al*., "Impact of threshold voltage fluctuation due to random telegraph noise on scaled-down SRAM," *2008 IEEE International Reliability Physics Symposium*, Phoenix, AZ, USA, 2008, pp. 541-546, doi: 10.1109/RELPHY.2008.4558943.

[32] Karna, Dr. Shyam & Ph, Karna & Scholar, A. (2023). An Overview on Taguchi Method.

[33] Roy, Ranjit K. *A Primer on the Taguchi Method*. 2nd ed., Society of Manufacturing Engineers, 2010.

[34] J.A Ghani, I.A Choudhury, H.H Hassan, Application of Taguchi method in the optimization of end milling parameters, Journal of Materials Processing Technology, Volume 145, Issue 1, 2004, Pages 84-92, ISSN 0924-0136, https://doi.org/10.1016/S0924-0136(03)00865-3.

[35] Florescu, Octavian. "Lecture 13." EE40 Introduction to Microelectronic Circuits. EE40 Summer 2005: Lecture 13, Berkeley, CA, UC Berkeley.

[36] C. -Y. Chen *et al*., "Correlation of Id- and Ig-random telegraph noise to positive bias temperature instability in scaled high-κ/metal gate n-type MOSFETs," *2011 International Reliability Physics Symposium*, Monterey, CA, 2011, pp. 3A.2.1-3A.2.6, doi: 10.1109/IRPS.2011.5784475.

[37] G. X. Duan *et al*., "Activation Energies for Oxide- and Interface-Trap Charge Generation Due to Negative-Bias Temperature Stress of Si-Capped SiGe-pMOSFETs," in *IEEE Transactions on Device and Materials Reliability*, vol. 15, no. 3, pp. 352-358, Sept. 2015, doi: 10.1109/TDMR.2015.2442152.

[38] Chen H, He L. The spatial and energy distribution of oxide trap responsible for 1/f noise in 4H-SiC MOSFETs. *Journal of Physics Communications*. 2021;5(3).

https://www.proquest.com/scholarly-journals/spatial-energy-distribution-oxide-trap/docview/251 2974355/se-2. doi: https://doi.org/10.1088/2399-6528/abe7d8.

[39] G. Nicosia *et al*., "Characterization and Modeling of Temperature Effects in 3-D NAND

Flash Arrays—Part II: Random Telegraph Noise," in *IEEE Transactions on Electron Devices*,

vol. 65, no. 8, pp. 3207-3213, Aug. 2018, doi: 10.1109/TED.2018.2839904.

.

[40] H. C. Ma *et al*., "Program Trapped-Charge Effect on Random Telegraph-Noise Amplitude in a Planar SONOS Flash Memory Cell," in *IEEE Electron Device Letters*, vol. 30, no. 11, pp. 1188-1190, Nov. 2009, doi: 10.1109/LED.2009.2030589.

[41] Fonstad, C. (n.d.). *Lecture 12 - Sub-threshold Mosfet Operation*. *6.012 Microelectronic Devices and Circuits*. Retrieved 2022, from

https://ocw.mit.edu/courses/6-012-microelectronic-devices-and-circuits-fall-2009/pages/syllabus/