Characterisation and modelling of Random Telegraph Noise in nanometre devices

Mehzabeen Mehedi

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Doctor of Philosophy

November 2021

I dedicate this Thesis with love to the reason of my existence and the wind beneath my wings-

*My mother **HALIMA HASSAN***

&

*My father **MEHEDI HASSAN***

Acknowledgements

All praise be to the Almighty for giving me the health, knowledge, opportunity and patience to perform the experiments and complete the dissertation in time.

The next hundred pages of this document contain the work I consider to be my contribution to the field of engineering that I love. But what I take with me is much more than the technical expertise. Pursuing my doctoral degree has been a wonderful journey with guidance, encouragement and support from so many people. I am very thankful to Liverpool John Moores University for awarding me the Faculty Scholarship to undertake my doctoral study.

I would like to express my sincere gratitude to my director of study, Professor Jian Fu Zhang. This study would have been impossible without your kind supervision, inspiration, direction and support throughout the years. Your knowledge, expertise, advice, and patience have been a motivating force for me during this journey. I am extremely grateful to you for putting your faith in me and my potential and guiding me through this challenging path and urging me towards the most rewarding direction. Your prompt responses to my emails and detailed feedback were of the utmost help. I cannot thank you enough for allowing me to do the part time jobs for my survival in the UK and trying your best to get me any possible funding whenever available. I would especially like to thank you for helping me in the submission and acceptance of my publications.

I would like to show my heartiest appreciation to my second supervisor Dr. Zhigang Ji for his extreme skills and knowledge in overcoming the hurdles during the first year of my PhD journey and thoroughly helping me out to get familiar with this research field while encouraging me with the future opportunities afterwards. It has been a real privilege to work with you all over the past several years.

I am very thankful to my third supervisor and also the Head of the Electronics and Electrical Engineering Department, Dr. John Marshland for always being so supportive and allowing me not to participate in voluntary teaching as I am already working 20 hours part time for my survival. I will always be grateful for this help.

I would like to thank Dr. Rui Gao, who has been more like my mentor during the start of my PhD. A big thank you for your consistent help during all the experiments in the lab and also for your support later in the study. I extend my gratitude to Dr. Zheng Chai and Dr. Firas Hatem for your kind suggestions, discussions and so many unforgettable memories.

A special thanks to Mr. Kean Tok for helping me in data analysing and all the group discussion on MATLAB simulations, new ideas and suggestions. Your outstanding hard work and contribution to my project is highly appreciated.

Many thanks to Dr. James Brown and Mr. Pedro Freitas for making my PhD journey more lively and memorable. Thank you for all the fun, coffee times, lunch breaks and most needed counselling during my hard times. I am also extremely grateful to Dr. James Brown for proof reading my thesis. I could not have asked for a better team than this.

I express my gratitude to the administrative team of the university. This includes the engineering administration, international advice team, career council, financial support team, student union and explicitly Natasha Walden-Jones from the engineering admin who has always been available to provide detailed suggestions and assistance whenever I had any issues.

Last but NEVER the least, my heartiest thanks to my family back home in Bangladesh, especially to my parents. Thank you for trusting me and allowing me to follow my dreams. I hope I have made you both proud and hope that all your sacrifices were worth it. Special thanks to my lovely Grandmother for her blessings and immense love. Thank you my beloved sisters, Mehnaz, Mehrin and Mehkan for all the things which cannot be expressed in words. I know how eagerly you all have been waiting for the completion of my PhD. And of course, the little members of the family: Aahil and Arhaam whom I have been missing most during this abroad life. I am extremely fortunate to have you all in my Life.

Abstract

The power consumption of digital circuits is proportional to the square of operation voltage and the demand for low power circuits reduces the operation voltage towards the threshold of MOSFETs. A weak voltage signal makes circuits vulnerable to noise and the optimization of circuit design requires an accurate noise model. RTN is the dominant noise for modern CMOS technologies. This research focuses on the instability induced by Random Telegraph Noise (RTN) in nano-devices for low power applications, such as the Internet of Things (IoT). RTN is a stochastic noise that can be observed in the drain/gate current of a device when traps capture and emit electrons or holes. The impact of RTN instabilities in devices has been widely investigated. Although progress has been made, the understanding of RTN instabilities remains incomplete and many issues are unresolved. This work focuses on developing a statistical model for characterising, modelling and analysing of the impact of RTN on MOSFET performance, as well as to study the prediction for long-term RTN impact on real circuits.

As transistor sizes are downscaled, a single trapped charge has a larger impact and RTN becomes increasingly important. To optimize circuit design, one needs to assess the impact of RTN on circuits, which can only be accomplished if there is an accurate statistical model of RTN. The dynamic Monte Carlo modelling requires the statistical distribution functions of both the amplitude and the capture/emission time (CET) of traps. Early works were focused on the amplitude distribution and the experimental data of CETs has been too limited to establish their statistical distribution reliably. In particular, the time window used has often been small, e.g. 10 sec or less, so that there is little data on slow traps. It is not known whether the CET distribution extracted from such a limited time window can be used to predict the RTN beyond the test time window. The first contribution of this work is three-fold: to provide long-term RTN data and use it to test the CET distributions proposed by early works; to propose a methodology for characterising the CET distribution for a fabrication process efficiently; and, for the first time, to verify the long-term prediction capability of a CET distribution beyond the time window used for its extraction.

On the statistical distributions of RTN amplitude, three different distributions were proposed by early works: Lognormal, Exponential, and Gumbel distributions. They give substantially different RTN predictions and agreement has not been reached on which distribution should be used, calling

the modelling accuracy into question. The second contribution of this work is to assess the accuracy of these three distributions and to explore other distributions for better accuracy. A novel criterion has been proposed for selecting distributions, which requires a monotonic reduction of modelling errors with increasing number of traps. The three existing distributions do not meet this criterion and thirteen other distributions are explored. It is found that the Generalized Extreme Value (GEV) distribution has the lowest error and meets the new criterion. Moreover, to reduce modelling errors, early works used bimodal Lognormal and Exponential distributions, which have more fitting parameters. Their errors, however, are still higher than those of the monomodal GEV distribution. GEV has a long distribution tail and predicts substantially worse RTN impact. The project highlights the uncertainty in predicting the RTN distribution tail by different statistical models.

The last contribution of the project is studying the impact of different gate biases on RTN distributions. At two different gate voltage conditions: one close to threshold voltage |Vth| and the other under operating conditions, it is found that the RTN amplitude follows different distributions. At operating voltage condition, Lognormal distribution has the lowest error for RTN amplitude distribution in comparison with other distributions. The amplitude distribution at close to |Vth| has a longer tail compared with the distribution tail at operating voltage. However, RTN capture/emission time distribution is not impacted by gate bias and follows Log-uniform distribution.

List of Abbreviations

Abbreviation	Definition
AC	Alternating Current
AG model	As-grown Generation model
AHT	As-grown (Hole) Traps
AT	As-grown Traps
BCDF	Bimodal Cumulative Distribution Function
BD	Break-Down
BTI	Bias Temperature Instability
CDF	Cumulative Distribution Function
CET	Capture Emission Time
CMOS	Complementary Metal Oxide Semiconductor
CSR wafer	Cambridge Silicon Radio wafer
DAQ	Data Acquisition
DC	Direct Current
DRAM	Dynamic Random Access Memory
DUT	Device Under Test
EAD	Energy Alternating Defects
EM	Estimation Maximization
Env	Envelope
FHMM	Factorial Hidden Markov Model
GD	Generated Defects
GEV	Generalized Extreme Value
GND	Ground (signal-system)
GS	Giga-Sample points
HCI	Hot Carrier Injection
HK/MG	High-κ Dielectric and Metal Gate
HMM	Hidden Markov Model
IV	Drain current (Id)~Gate voltage (Vg)
JFET	Junction Field Effect Transistor
LSRCC	L-shaped RC circuit
MCMC	Markov Chain Monte Carlo
MLE	Maximum Likelihood Estimation
MOSEFT	Metal Oxide Semiconductor Field Effect Transistor
MS	Mega-Sample points
MSM	Measure-Stress-Measure
NBTI	Negative Bias Temperature Instability
NMOS	Negative Metal Oxide Semiconductor
OSC	Oscilloscope
PC	Personal Computer

PP	Percolation Path
PSD	Power Spectrum Density
RC circuit	Resistor–Capacitor Circuit
RDF	Random Dopant Fluctuation
RF	Radio-Frequency
RRAM	Resistive Random-Access Memory
RTN	Random Telegraph Noise
RTS	Random Telegraph Signal
SMU	Source Measurement Unit
SRAM	Static Random-Access Memory
SSE	Sum-Square-Error
TC	Transfer Characteristics
TDDS	Time-Dependent Defect Spectroscopy
TLP	Time Lag Plots
TSMC	Taiwan Semiconductor Manufacturing Company
TWC	Trigger-When-Charged
UE	Upper Envelope
UE/LE	Upper/Lower Envelope
VLSI	Component of very Large Scale Integrated
WDF	Within Device Fluctuation
WTLP	Weighted Time Lag Plot

List of Symbols

Symbol	Description	Unit
$\overline{\Delta N}$	change in number of carriers	m^{-3}
μ	Mean/average value	
μ_0	carrier mobility	
A_0	plateau amplitude at low frequency	
Ef/EF	Fermi level	eV
E_T	trap energy level	
f	Frequency	Hz
f_0	corner frequency	
Fs	sampling frequency	Hz
f_t	Trap occupancy function	
g_m	transconductance of an Id-Vg curve	S
H	Hooge constant	
Id	Drain current	A
I_{dcc}	constant Id value for Vth extraction	A
k	Boltzmann's constant	
N	number of carriers per unit area	m^{-3}
Nt	number of occupied oxide traps	
$P_{H,L}$ (t)	probability of the transition state at high or low	
q	charge carrier	m^{-3}
R	Resistance	Ω
Si	Silicon	
S_I	noise current spectral density	
Si02	Silicon Dioxide	
Sv/Si	Power Spectral Density	
T	Temperature	ºC
t	time	sec
tw	time window	sec
Vbs	Substrate bias	V
Vd	Drain Voltage	V
Vdd	drain supply voltage /Circuit operating voltage	V
Vg	Gate voltage	V
Vgs	gate to source voltage	V
W/L	MOSFET channel width/length	m
α_0	material constant	
β_0	suppression factor	

γ	exponent/ material constant	
ΔI_D	Drain current fluctuation	A
δVth	Single trap amplitude	V
ΔV_{TH}	Threshold Voltage shift	V
Δt	Time intervals	sec
λ	Coefficient of electron wave attenuation	
π	pi	
σ	Standard deviation	
τ	Average time constant	sec
τ_C	Average capture time	sec
τ_e	Average emission time	sec
τ_H	Time constant at high state	sec
τ_L	Time constant at low state	sec
η	Average single trap impact	

List of Figures

List of Tables

Table of Contents

Chapter 1 Introduction

1.1 Preface and Motivation

In the microelectronics industry, metal oxide semiconductor field effect transistors (MOSFETs) are one of the backbones of technological advancements. This is because aggressive scaling in MOSFET's dimensions has allowed billions of devices to be integrated on one chip, which enables low power consumption, higher computational speed and functionality while also reducing the price per device. The success of MOSFETs in analog and digital applications relies on the remarkable stability of silicon (Si) and silicon dioxide (SiO_2) materials. However, highly scaled transistors at nanometre scales have major reliability issues that threaten integrated circuit performance and stability. These include random telegraph noise (RTN) and several gate oxide ageing phenomena such as hot carrier injection (HCI), gate oxide breakdown (BD), and bias temperature instability (BTI) These instabilities cause the MOS transistor parameters, such as threshold voltage shifting, fluctuating and/or drifting of current and mobility degradation [1]. Random Telegraph Noise (RTN), caused by discrete charge trapping-detrapping in the gate stack, has become a significant contributor to transistor variability [2]. It affects circuits and systems over their lifetime and needs to be taken into account in circuit design and modelling.

This research project will focus on RTN which gives rise to a discrete switching within the drain current of MOSFETs. The RTN was reported by Ralls *et al* in 1984 [3]. It was then investigated and modelled by Kirton and M. Uren in 1989 [4]. This phenomenon becomes more important in deeply scaled transistors [5] when even scaled down below 45-nm [6][7]. After the first processor chips of the 45-nm technology node were released in 2007, great attention was paid towards the impact of RTN on static random access memory (SRAM). SRAM is a critical component of very large scale integrated (VLSI) systems as it provides the fastest random access time to stored data, and is used for lower-level caches (L1-L3) and registers [8]. In order to increase the size of the cache on a chip, it is desirable and economical to fit as many cells into an SRAM array as possible. As a result, SRAM uses the smallest MOSFETs available from each technological node. For the first time in 2009, the International Technology Roadmap for Semiconductors (ITRS) raised the issue of SRAM noise margins, as the threshold voltage variability induced by RTN increases [9]. Nonetheless, there is continuing research and development to overcome transistors' short channel effects due to scaling when using the high-κ dielectric and metal gate (HK/MG) technique. In 2014, Intel Corporation commercially released the multi-gate transistor named Tri-gate on the 22-nm technology node [10]. Figure 1.1.1 shows a chart of these two technology nodes' characteristics and their challenges [11]. RTN has become a severe reliability challenge in the age of tri-gate transistors. The traps in gate dielectric are considered a main source of RTN. Traps are mainly defects within microelectronic devices that can capture mobile charge carriers and degrade the performance of the devices [12].

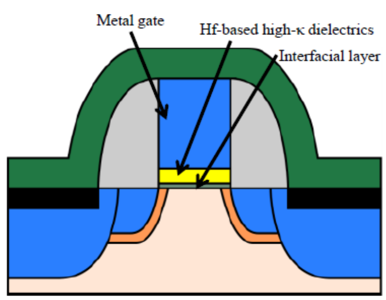
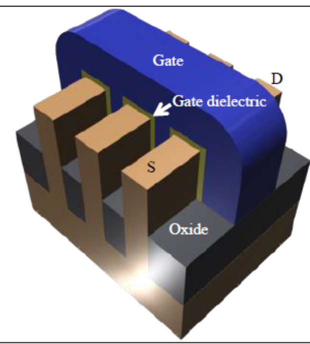
Technology node	45 nm ~	22 nm ~
Device structure		
Characteristics	Hf-based high-κ dielectric Metal gate	Multiple gate 3D structure
Challenges	SCE NBTI Variability	NBTI Variability Random telegraph noise

Figure 1.1.1 Trend of MOSEFTs structure and challenges [11].

Earlier work [13] has improved people's understanding of defects and instabilities in MOSFETs. There are, however, questions remaining to be answered. The ageing and instability of devices under real use conditions is typically small and difficult to measure. To overcome this difficulty, previous researchers [14] accelerated the ageing and increased the instability by stressing devices under higher biases. This is acceptable for negative bias temperature instability since the generated defects under raised biases are the same as those under real use low biases. For the instability induced by the random telegraph noise (RTN), however, the dominating defects are different under different biases. Consequently, the results obtained under raised biases are not applicable to real operation. In addition, primary works on RTN also suffer from some further shortcomings:

One of them is that devices have to be selected for their analysable RTN signals. In reality, many devices do not have analysable RTN signals for extracting mean capture/emission time and their drain current exhibits a complex ‘within-a-device-fluctuation (WDF)’ (details in section 2.5). This selection of devices does not allow obtaining the true statistical properties of device instabilities. Therefore, devices should not be selected to evaluate real device-to-device variation. Also, the test window used for RTN measurement is typically very limited as longer time windows will result in unfeasibly large amounts of data. These are the main deficits of models for the assessment of statistical distribution of RTN time constant and amplitude. Another shortcoming of earlier work is that researchers either do not report the threshold voltage fluctuation, ΔV_{TH}, or simply evaluate it from the current fluctuation by dividing the trans-conductance under the operation bias. As the operation bias is typically well above the threshold voltage, the accuracy of ΔV_{TH} evaluated in this way is not known. Two potential sources of errors are: the neglect of mobility degradation and the localization of current path under threshold condition. To assess the impact of instability on a circuit, an accurate ΔV_{TH} is essential, but not yet available in previous works.

The Internet of Things (IoT) will become increasingly important in the future and one of its key requirements is low power. To achieve low power, the operation bias must be reduced. For lower bias, the same ΔV_{TH} will lead to higher variation of driving current, because of the increase of trans-conductance. There is a lack of systematic data on the instabilities for lower biases. Because of these mentioned shortcomings, there is no well accepted model available for assessing the impact of RTN instabilities on circuits. This prevents the designers from optimizing their designs. At present, the designers have to use an estimated margin to cover the effect of instabilities. RTN also has severe impact on analog and RF applications in advanced complementary metal oxide semiconductor (CMOS) technologies [15] such as CMOS image sensors [16] and flash memories [17]. The flash memory devices are especially sensitive to RTN due to its reliance on charge trapping as a mechanism for data storage.

The motivation of this project is to develop new techniques for characterising instability induced by RTN in nano scale CMOS devices and to model it for low power applications. Low power applications, such as IoT devices, require using low operation biases and the impact of instabilities is larger under lower biases. There are no well accepted physics-based statistical device models available for the above-mentioned instabilities induced by RTN that can be used to optimize the circuit design. In addition, most of the current studies for RTN characterisation were conducted in the frequency domain [3] [18] [19]. Moreover, there are some concerns about using RTN simulation tools such as Hidden Markov Model (HMM) [20], Time Lag Plot (TLP) (detail in section 1.4.1) [21] and two stage L-shaped RC circuit (LSRCC) [22] in characterising RTN signal. For instance, HMM simulator is limited to three active traps in a particular device. The RTN waveform or 1/f noise spectrum generated by using LSRCC does not relate to the physical mechanism of traps but is only a lookalike of the random waveform instead. In the case of TLP, if the background noise (i.e. the noise from the measurement system or from the surroundings) becomes comparable with the RTN current/voltage steps then the detection of precise defects and parameter extraction become difficult. This project will tackle these challenges. Starting from an in-depth understanding of RTN characteristics, the correct measurement conditions are identified to estimate the statistical parameters of RTN from multiple devices. Finally, this work develops a statistical model supported by extensive experimental data as well as simulation tool to predict long-term RTN on real circuits in future.

1.2 Electronic Noise in Semiconductor Devices

1.2.1 Noise Mechanisms

In an electronic circuit, the currents and voltages fluctuate randomly around their average levels due to the fluctuations of electronic transitions. This causes difficulty in distinguishing the required signal from the noise when the noise power becomes significant in relation to the signal power which typically referred as signal to noise ratio. In general, the signal to noise ratio in large devices is better than in small area devices. Noise cannot be eliminated completely and hence it has been a fundamental problem. Thus, ultimately the industry has to make adjustments by limiting the accuracy of measurements and settling at a limit for detecting and processing small signals in electronic devices. Earlier works have shown that noise is not only a problem but also has some importance in evaluating and getting an insight to the properties of a particular system [23]. The characterisation and classification of noise at low frequencies can provide important information of the device quality, properties, physics and reliabilities such as traps or defects and scattering processes.

The noise in an electronic device is random and it is categorised into two groups: external noise and internal noise. The sources of external noise in an electronic system are electrostatic and electromagnetic signals, cross-talk between adjacent circuits, vibration from surroundings, light and radio signals [24]. These disturbances can be minimised by protecting or shielding the experimental environment, using filters, vacuum chambers, and by changing the layouts. The second group is internal noise whose sources are defects or faults in the devices. Internal noise is a random phenomenon and defects can exist anywhere in the bulk of the semiconductors, dielectric, or at the interface between two materials. There are several noise mechanisms for internal noise sources that can generate random fluctuations, such as thermal noise, shot noise, flicker or 1/f noise, $1/f^2$ noise, generation recombination noise, random telegraph noise and avalanche noise. This thesis is mainly covering one type of internal noise, i.e. RTN. Practically, most current and voltage fluctuations in electronic devices are considered to follow Gaussian processes due to the central-limit theorem which states that the sum of a large number of independent random variables shows Gaussian distribution. The one exception is random telegraph noise, that depicts the switching of the signal between two levels and typically does not follow the Gaussian distribution [4]. The different types of noise are briefly discussed below along with their Power Spectrum Density (PSD).

1.2.1.1 Thermal Noise

Thermal noise is the most commonly discussed noise source in electronic circuits and is also referred to as Johnson noise. Thermal noise caused by thermal random fluctuations of electric charge carriers (electrons and holes) in a material when the temperature is above zero Kelvin. As electrons are lightweight they are moving rapidly. This movement combined with their charged nature induces fluctuating electric currents. These fluctuations in electric current are a form of noise. This was first observed and studied by J. B. Johnson in 1928 and further explained by H. Nyquist later in the same year [25][26]. Thermal noise follows Gaussian distribution and a constant power spectral density (PSD) across a wide range of frequencies. Each time a charge carrier is scattered, the velocity of the electron is randomized. The thermal noise even exists at zero bias condition. Hence, this noise is considered as a major hindrance to extending Moore’s law into very deep sub-micron technologies [27]. When gate to source voltage (Vgs) is increased, it causes an increase of charge carrier density in the channel of MOS transistors that simultaneously reduces the channel resistance. This in turn leads to the reduction of thermal noise in a particular device. Statistically, thermal noise increases with temperature. If a piece of conducting material (metal or semiconductor), having resistance R, the PSD of the thermal noise at a temperature T can be evaluated by:

$$Sv = 4kTR \tag{1}$$

where, k is Boltzmann's constant [23][25]. This noise is also widely known as white noise in electronic systems.

1.2.1.2 Shot Noise

In electronics, shot noise originates from the discreteness of charge carriers in the presence of an electric field, also known as Schottky noise. The current flowing across the potential barrier, such as a pn-junction, is not continuous and therefore shot noise is generated when the charge carriers arrive or cross the barrier independently and at random. This type of noise was first observed in vacuum tubes by Walker Schottky in 1918 [24][28][29]. The physics behind shot noise and thermal noise are very closely related. However, the shot noise occurs only under DC bias and in conducting materials that have a potential barrier. In a MOSFET, when each charge carrier 'q' crosses the barrier at a certain time then the current 'I' fluctuates with a PSD:

$$Si = 2qI\beta_0 \tag{2}$$

where β_0 is known as suppression factor whose typical value is less than 1 [24][28]. Both thermal and shot noises are independent of frequencies and shot noise is also temperature independent unlike thermal noise. In MOS transistors, high gate leakage current due to ultra-thin gate oxides can also generate this kind of noise [30][31].

1.2.1.3 Generation Recombination Noise

Generation recombination noise, known as g-r noise, is the type of electrical signal noise that originates from the statistical fluctuation of the current transport due to the random and spontaneous capture and emission of charge carriers. When charge carriers get trapped by a defect in the bulk or at the interface of the device, it can induce fluctuations even in the electric field, mobility, space charge region width, barrier height, diffusion coefficient etc. [24][29][32]. In electronic materials, the defects present within the forbidden bandgap are referred to as traps and if this trap is close to the middle of Si band gap then it is considered to stem g-r noise. These traps exist in devices due to the presence of impurities, defects or faults in the semiconductor materials or around the surface after manufacture. The power spectral density of the charged carrier fluctuations can be derived as [33]:

$$Sn\,(f) = \frac{4\overline{\Delta N^2}\tau}{1 + (2\pi f)^2\tau^2} \tag{3}$$

Here, τ is the average time constant for random switching carrier, f is the frequency and $\overline{\Delta \mathrm{N}}$ is the change in the number of carriers due to the trapping and detrapping phenomenon. The spectrum shape generated by equation (3) is referred to as Lorentzian. This type of noise is considered as a function of both temperature and biasing conditions.

1.2.1.4 Flicker or 1/f noise

Flicker noise is a dominant electronic noise in the low frequency range with a power spectral density proportional to 1/f. It is therefore often referred to as 1/f noise or pink noise though these terms have wider definitions as it occurs in almost all electronic devices. It can show up with a variety of other effects such as impurities in a conductive channel, generation and recombination noise in a transistor. The origin of 1/f noise is related to traps in the gate dielectric hence it varies from system to system and its physical mechanism has not been completely identified yet. Several models have been proposed for the physical mechanism of 1/f noise, but the two major theories are: the number fluctuation theory and the mobility theory. The carrier number fluctuation theory states that the flicker noise occurs due to the random charge trapping and de-trapping processes in the oxide traps near the Si/SiO2 interface. This fluctuation results in fluctuation in the channel carrier density which in turn modulates the drain current [34]. For example, in order to characterise each trap by the relaxation time (τ), and the occupation function *N(t)* can be defined as : *N* = 1 when the trap is occupied, *N* = 0 when the trap is empty. Then the power spectral density of *N(t)* can be written as equation (3). In the case of several traps present with time constant (τ), then the superposition of RTN signal gives a 1/f noise with the overall PSD equation express as [35] :

$$S\,(f) = \frac{K_B T q^2 N_t(E_F)}{\lambda W L f^{\gamma}} \tag{4}$$

where, N_t is the is the distribution of the traps over the energy and space, λ is the coefficient of electron wave attenuation and γ is the exponent. In contrast, the mobility fluctuation theory, also known as Hooge's Model, states that the current increases or decreases due to the fluctuation in the carrier mobility [36]. The noise spectral density for the Hooge's model is [37]:

$$S(f) = \frac{\alpha_0 H I^{\alpha_0}}{f^{\gamma}.N} \tag{5}$$

where, H = 2×10^{-3} is the Hooge constant [37], α_0 and γ are material constant, and N is the number of carriers. Flicker noise becomes a real problem for devices in nanoscale range as it increases with the decrease of device dimensions. The level of 1/f noise is often considered as a measure of the quality of semiconductor devices and its reliability.

1.2.1.5 Random Telegraph Noise

The study of RTN has provided new insights into the nature of defects within devices. Theoretically, for a two-state simple current signal with amplitude ΔI_D and the average time constant (Poisson distributed time duration of staying at low state τ_L and higher state τ_H) can generate the following RTN signal shown in Figure 1.2.1, both in time and frequency domain respectively. From such a signal the PSD of the current fluctuation of the device can be derived as [38] :

$$Si\,(f) = \frac{4(\Delta I)^2}{(\tau_L + \tau_H)\left[\left(\frac{1}{\tau_L} + \frac{1}{\tau_H}\right)^2 + (2\pi f)^2\right]} \tag{6}$$

The types of PSD behaviour shown below are the Lorentzian type. RTN is an interesting study from the electronic characterisation point of view as the random switching process can be acquired from just one trap in the time domain. This thesis is mainly focused on this particular noise characterisation of CMOS transistors and more details, methods and theories are further discussed in the following sections and chapters.

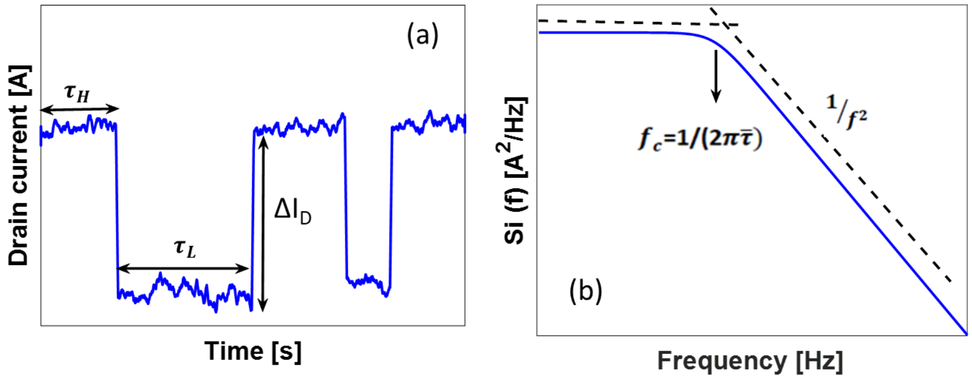


Figure 1.2.1 (a) Illustration of the RTS given by the impact of a single defect in the oxide in time domain (b) Illustration of a RTN spectrum in log-log frequency domain for a single trap impact where fc is the corner frequency and 1/f^2 shows the Lorentzian characterisation.

1.3 RTN in small-area MOSFETs

1.3.1 Overview of RTN phenomenon

The discrete switching behaviour in the current and performance degradation of semiconductor devices due to the presence of traps has been investigated ever since the first generation of CMOS technology was developed in the 1960s [39]. The single switching event induced by RTN was observed in 1978 in a junction field effect transistor (JFET) by Kandiah and co-workers [40]. Later, the cause of RTN fluctuation due to interface trap and bulk trap in gate oxide of very small area MOSFETs was observed by Ralls et al. and reported in 1984 [3]. However, there was a suspicion about the origin of defects from 1/f noise, until K. S. Ralls et al. [3] reported that the 1/f noise of MOSFETs most likely consists of many RTN signals. This observation was reported on the basis of the work of McWhorter who suggested the number fluctuation model that states 1/f noise is composed of a large number of RTNs by using Ge filament in 1957 [35]. An extensive amount of research has focussed on RTN in electronic devices ever since. For the last few years, RTN has been intensively studied in scaled memory cells like SRAMs, DRAMs, RRAMs [41], [42][43] and also flash memories [17].

The typical behaviour of a simple RTN, caused by a single trap at the gate dielectric of small area MOSFETs, results in a two-level signal as shown in Figure 1.3.1 (a). The drain current on this nano-scale device is fluctuating under a constant gate voltage shown in Figure 1.3.1 (b) [44]. RTN studies use this to characterise electron traps in nMOSFETs and hole traps in pMOSFETs. These trapping events in MOSFETs lead to changes in the threshold voltage of the device [45]. The induced charges cause changes in the potential of the device which lead to a shift of the Id-Vg transfer characteristic of the probed device, as shown in Figure 1.3.1 (c) [46].

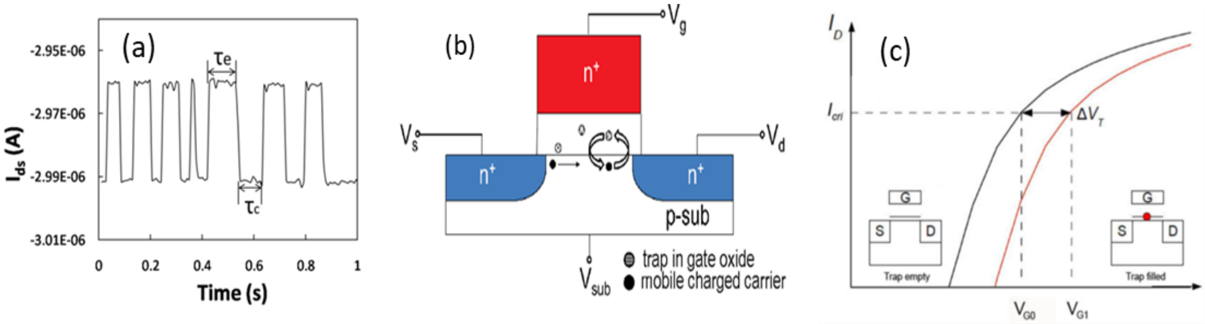


Figure 1.3.1 (a) One typical measured RTN signal from a 90nm ×50nm device, τc and τe are the capture and emission times, and measurement was done with Vg=-0.6V [45]. (b) Random Telegraph Noise [44] at constant Vg bias, oxide defects are charged and discharged frequently. (c) Definition of threshold voltage shift at a specific current criterion by RTN [46].

The impact of an RTN trap and its energy information can be extracted from the RTN amplitude, τc (the average capture time) and τe (the average emission time) according to Equation (7) [4] :

$$\frac{\tau_C}{\tau_e} = \exp(\frac{E_T - E_F}{KT}) \tag{7}$$

where, k is Boltzmann constant, T is temperature, and E_T is RTN trap energy level.

According to Pelgrom's law, the fluctuation in a device by process variation is inversely proportional to the square root of the device size, whereas the fluctuation due to RTN is inversely proportional to the device size [44]. As mentioned, the impact of RTN can be expressed in the form of threshold voltage shift (ΔV_{TH}) at a fixed drain current as illustrated in Figure 1.3.1 (c) by Equation (8): [47]

$$\Delta V_{TH} = \frac{q}{CoxWL} \tag{8}$$

where q is the electron charge, C_{OX} is the gate oxide capacitance per unit area and W and L are the channel width and channel length of the transistor respectively. The impact of RTN increases when the transistor dimensions are scaled down due to the increase in ΔV_{TH} which is inversely proportional to the channel length and width as shown in eq. (8). Recent studies show that ΔV_{TH}

caused by RTN increases more rapidly than the threshold variation caused by random dopant fluctuation (RDF). It is reported in [44] that RTN-induced ΔV_{TH} may exceed RDF-induced threshold variation at the 3σ level in 22 nm technology, thus requiring more attention in future.

RTN signal characterisation includes three parameters: the switching amplitude of the drain current ΔId between the low and high states, the average time of the higher level usually associated as capture time (τc), and the average time of the lower level normally denoted as emission time (τe). If plotted as a histogram the drain current amplitude will exhibit two peaks at a distance that will equal to ΔId. It has been reported that the high- and low-level times of the histogram are normally exponentially distributed due to the Poisson process governing the dynamic trapping events [48].

1.3.2 Physical Origin of RTN

Random telegraph noise (RTN) signal originates from the discrete switching behaviour in the device channel resistance. This is due to the trapping of a charge carrier in the inversion layer that causes change in the carriers of both the number and mobility fluctuation of the inversion layer. The importance of this switching event in a scaled device increases due to the large impact of a single defect on the device performance. Hence, RTN signals have become a major issue in nano range devices and the measurements are usually suggested to be carried out with gate area < 1μm^2 [5]. The study of RTN can also be used as a powerful tool for investigating the capture and emission kinetics of a single defect in the gate oxide, as there is a visible abrupt change in the transistor current [49]. Moreover, in the deeply scaled submicron FETs there is more than one type of defect present, which are responsible for the RTS noise in advanced transistors.

1.3.2.1 Generated Defects

Generated defects (GD) refer to the defects that do not exist in fresh devices. In other words, it is a kind of trap that is created by electrical stresses [50]. It is reported in [50] that during a long stress test, the generated defects near a percolation path in submicron FETs induce larger RTN amplitude and a significant tail in its amplitude distribution. R. Gao *et al.* in 2016 [51] reported that generated defects have two components depending on the applied measurement conditions: one can repetitively charge and discharge and another one is very difficult to discharge. This kind of generated defect has been widely observed [52] but usually ignored [53] due to the lack of a proper characterisation process. It is also reported in earlier works that GD can be determined after carrier discharging [54]. As the discharging does not complete, it makes the extraction of GD sensitive to the discharging time. R. Gao *et al.* proposed a procedure to determine GD which fully captures the GD[51]. In addition to that, GD kinetics are also reported to follow classical power law under different overdrive voltages, which were used for long-term prediction of ageing. Although the majority of generated defects are stably charged and do not induce random telegraph noises in electronic devices, it has been reported that some GD can cause RTN [55]. This project focuses on as-grown traps, as they dominate RTN typically. As-grown traps will be discussed in the following section.

1.3.2.2 Pre-existing Defects

As-grown traps are also called pre-existing traps. These types of traps are already present in the fresh devices after fabrication and before electrical stresses. At present, the understanding of these traps is incomplete and there is a lack of a model which can predict the long-term RTN induced by them.

R. Gao *et al*. investigated pre-existing defect kinetics on heavily stressed devices where further defect generation is supressed [52]. Pre-Existing defects' Device-to-Device Variability can be extracted and analysed from the charging and discharging kinetics on stressed devices [51]. They reported two types of pre-existing defects: As-grown traps (AT) and Energy Alternating Defects (EAD) [51]. It is observed that the energy of as-grown traps does not change during the charging and discharging of charge carriers whereas in the case of EAD the energy level is pulled down after charging and goes back to the original level after discharging. The filling time (charging and discharging time) of EAD is much longer than ATs, which is just a few seconds. Hence, the charging kinetics of ATs and EADs can be separated by using the differences in their energy behaviour during charging and discharging [56], [57]. It is also reported that EADs follow a power law relationship like GD but with different exponents, thus needing separate modelling. From these investigations, it is found that the short term RTN noise is mainly dominated by as-grown traps and energy alternating defects do not induce RTN. Hence, RTN fluctuation is dominated by ATs. Generally, if as-grown traps can be successfully modelled, the long-term RTN of nano scaled devices can be predicted by an extrapolation method. The following chapters of this research work are focused on characterising and modelling the long-term impact of RTN induced by as-grown traps.

1.3.3 RTN Parameters

As already mentioned above, the RTN is characterised by three parameters: the amplitude associated with individual traps as well as the average time to capture and emission of charge carriers. These parameters are essential for the physical characterisation of the defects with the aim of extracting key attributes such as, energy level and the location of the traps in the oxide/dielectric interface.

1.3.3.1 RTN time constant behaviour

The time constant distribution of an RTN signal can be obtained from a simple two-state measured RTN signal over some period. The mean or average time of tc and te are referred to as τc and τe respectively. In the time domain, the time of the current fluctuation stays at high or low level are stochastically distributed and the probability distribution of an individual trap follows an exponential distribution [4]:

$$P_{H,L}(t) = \frac{1}{\tau_{H,L}} \exp\left(-\frac{t}{\tau_{H,L}}\right) \qquad (9)$$

where, $P_{H,L}$ (t) is the probability of the transition state at high or low with the corresponding average time constant $\tau_{H,L}$ respectively and t is the time of interval. The capture and emission phenomenon of a charge carrier can occur anytime and is thus a random phenomenon. As the number of states of a single trap RTN is 2, the total number of possible levels in a device is 2^n, where n is the number of traps in a device. For example, for 3 RTN traps the number of states is 8. There are occasions when the number of levels is smaller than 2^n, as some traps can be co-dependent.

RTN parameters are known to have a strong dependence on the operation conditions, i.e. the gate voltage (Vg), drain voltage (Vd), substrate bias (Vbs) and temperature. Earlier works have mentioned that this dependence can be used to reveal the underlying information of the trap parameters such as trap density, capture cross section for electrons and holes, energy of the trap site and the distance of traps from the Si/SiO_2 interface [4]. Figure 1.3.2 shows the energy band diagram of an nMOS transistor. The traps below the Fermi level (E_F) with black circles represent the traps that are filled, and above E_F (the white circles) are empty. The trap energy level is denoted as E_T and the traps (grey circles) around the E_F are acting as switching traps.

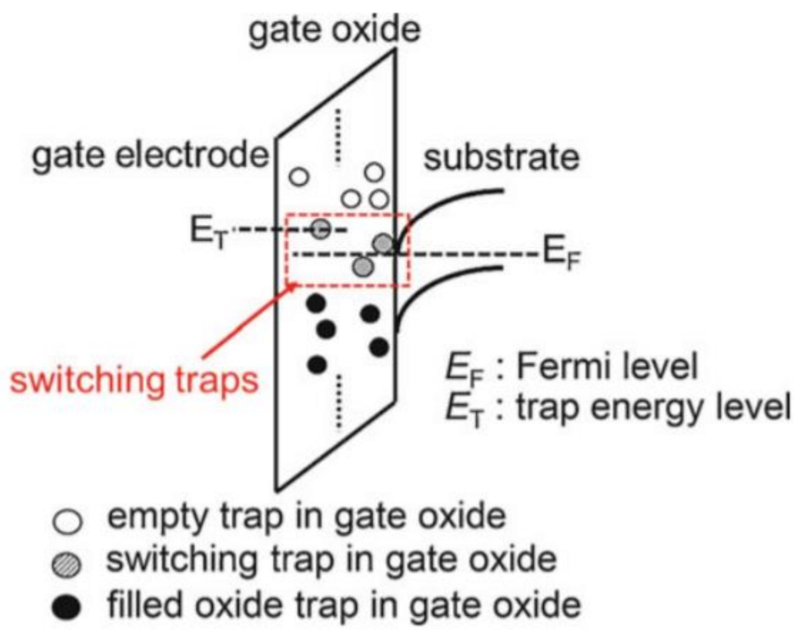


Figure 1.3.2 Energy band diagram of an nMOS transistor [58].

Furthermore, the nature of the centre, that is whether the trap is acceptor-like or donor-like, can be studied from dependencies that were mentioned above. For instance, depending on the type of the trap involved (acceptor or donor) and the type of the carrier, high and low states can be identified as either capture (t_c) or emission (t_e) time of the carrier from and to the relevant band [59]. Figure 1.3.3 shows a schematic diagram to explain the RTN trap type in time domain analysis. The lower level in the Figure 1.3.3 represents the state when a single trap is neutral (state 0), and the upper level (state 1) is the state when the trap is full i.e. either negatively or positively charged. For an n-channel MOSFET, an acceptor-like trap is negatively charged (when in state 1) after capturing an electron and becoming neutral (when in state 0) after emitting it, whereas a donor-like trap is neutral when capturing an electron (i.e. at state 0) and becoming positively charged (at state 1) after emitting it. In both cases, the higher current state corresponds to the neutral state [60]. The trap occupancy increases with applied Vg. Therefore, if the time constant at a lower current state (or, equivalently the high voltage state) increases with Vg then the trap is acceptor-like and state 1 is equivalent to τe and vice versa in the case of donor-like [59][61], as demonstrated in the schematic in Figure 1.3.3 (b) for acceptor-like trap and Figure 1.3.3 (c) for donor-like trap.

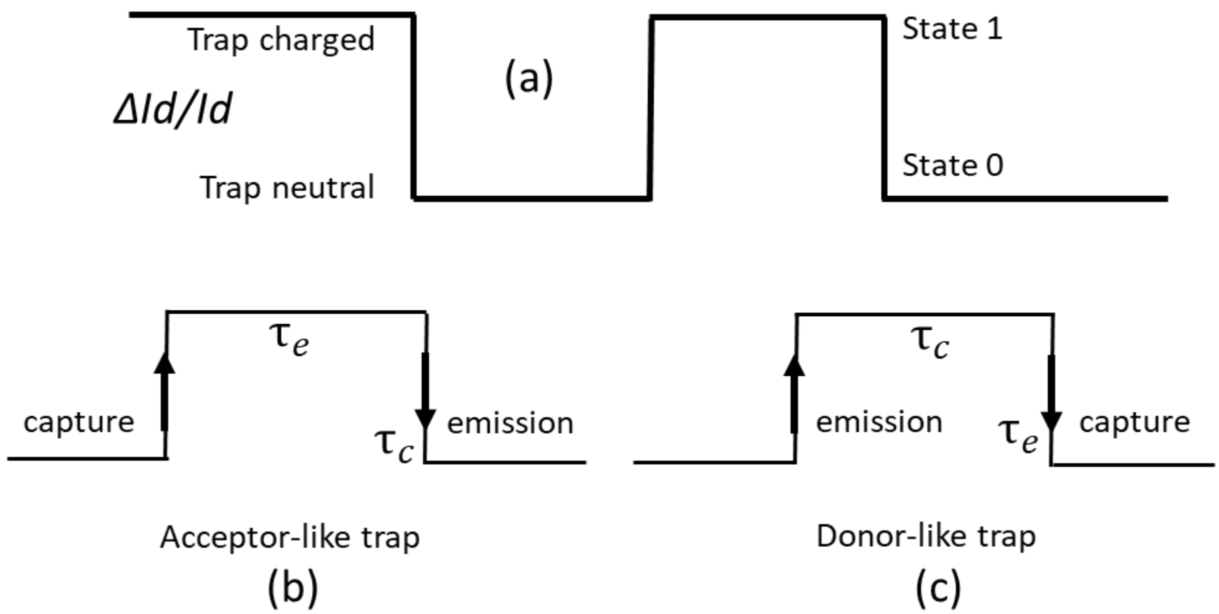


Figure 1.3.3 Schematic identification of donor and acceptor traps for electron trapping de-trapping. (a) State 0 is where the trap is neutral and state 1 is after the trap is charged. Capture and emission process and corresponding mean capture and emission times are shown for an (b) acceptor trap and (c) donor type trap.

Additionally, for an nMOSFET where the channel carrier is electrons, the trap occupancy function (f_t) can be written mathematically:

$$f_t = \frac{\tau_e}{\tau_e + \tau_c} \tag{10}$$

where f_t corresponds to the probability that the trap is filled with an electron [4]. From eq. (10), the ratio of the mean capture and emission time can be derived as:

$$\frac{\tau_c}{\tau_e} = \frac{1 - f_t}{f_t} \tag{11}$$

Here, the probability that the trap is filled with an electron increases with an increase in Vg and therefore, the ratio of τ_c/τ_e decreases with increase of Vg.

Moreover, there are two factors that need to be considered in the study of time constant of RTN measurement within the time domain: the number of transitions of t_c and t_e and the resolution of the observed data. High resolution is needed for RTN measurement so that no transitions between states are missed. To obtain accurate τ_c and τ_e there must be a sufficient number of transitions between states. According to Uren and Kirton [4], there should be a minimum 200 transitions within at least 20,000 data points in order to have error less than 10%. The difficulty arises when high resolution results in a large amount of data, which limits the measurement time window. Hence, the proper average time constant distribution is rarely reported in the literature based on experimental data and there is a lack of a convincing statistical distribution of τ_c and τ_e. One of the main objectives of this research is to develop a technique that increases the time window by keeping the fast sample rate without generating an unmanageable amount of data. This will be possible by making the storage rate slower than the sampling rate. Further details will be discussed in Chapter 2 .

1.3.3.2 RTN Amplitude behaviour

The amplitude of the noise signal is another important parameter. This is because downscaling device sizes causes a rapid increase in RTN amplitude, which can lead to transient probabilistic failure of digital CMOS circuits [62][63][64]. RTN amplitude is the current fluctuation obtained as the difference between the discrete levels of RTN. Practically, the RTN amplitude is more difficult to model due to its wide range observed from sample to sample and even for different RTN traps in the same transistors. According to Kiron and Uren [4], there are two mechanisms which contribute to RTN amplitude: the change in the density number of the carriers and the change in the mobility due to the charge carriers. Mathematically, the fluctuation change of the drain current can be represented in terms of number and mobility fluctuations [65]:

$$\frac{\Delta Id}{Id} = \frac{1}{W \times L}\left(\frac{1}{N}\frac{\partial N}{\partial N_t} \pm \frac{1}{\mu_0}\frac{\partial \mu_0}{\partial N_t}\right) \tag{12}$$

where, N is the number of carriers per unit area, N_t is the number of occupied oxide traps, W is the channel width, L is the channel length and μ_0 is the carrier mobility. The sign in front of the mobility term in equation (12) depends on the type of trap either acceptor or donor. In case of nMOSFETs, usually if the 'mobility term' holds a '-' sign then the centre is neutral when occupied i.e. a donor-type trap present whereas the '+' sign holds for a charged trap when filled i.e. acceptor trap [65]. However, some authors assumed that for nMOSFETs, Id is higher when trap is neutral for both donor-like and acceptor-like, although little evidence has been provided [61]. This means that, for donors, the impact of mobility reduction on Id is larger than the impact of an increase of electron number in the channel. Donor has smaller amplitude due to the compensation effect between number increase and mobility reduction. Some other groups also assumed that when a donor is positively charged, it will not increase the number of electrons in the channel i.e., the ‘number term’ in eq. (12) equals 0. It was proposed that a dipole will be formed between the ‘+’ donor and ‘-’ electron, so that this electron is not mobile as shown in Figure 1.3.4. The ‘-’ immobile electron in the channel causes Coulombic scattering and reduces drain current (Id) [66].

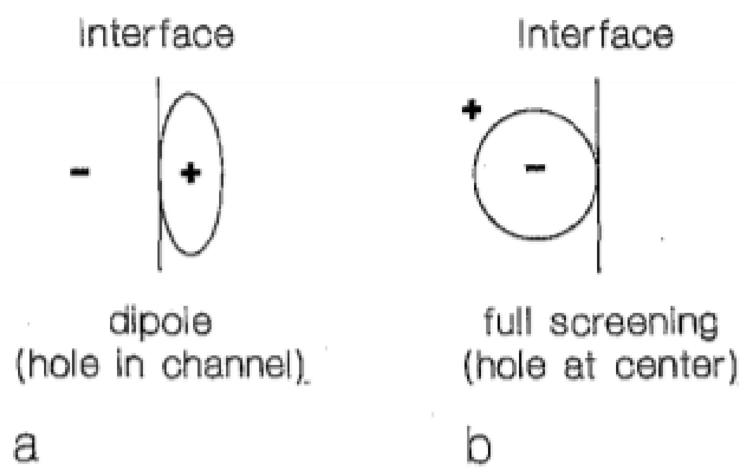


Figure 1.3.4 Schematic drawings of the space charge distribution for a Coulomb-attractive defect center near the SiO_2 -Si interface of a p-channel MOSFET. (a)Hole bound in the potential well of the channel, (b) Hole bound at the Coulomb-attractive defect center in the oxide [66].

RTN amplitude also has a dependency with applied Vg. It has been reported that on average, the RTN amplitude decreases with increasing Vg [21], [67]. RTN amplitude can be expressed as either normalized drain current fluctuation $\Delta I_d/I_d$ or the equivalent ΔV_{TH} shift. Recently, RTN amplitude distribution has attracted much attention due to the lack of proper statistical characterisation. The main uncertainly in RTN amplitude distribution is the exact number of traps present in a device. If there is more than one trap in a device, then they might couple with each other and result in a form of multi-state, complex RTN signal. It has been widely reported that RTN amplitude distribution mostly follows either an exponential [21], [67], [68], [69] or log-normal [6], [21], [70] distribution but without providing the underlying mechanism. Zhang *et al* [70] provide new insights into the RTN amplitude distribution by considering the impacts of trap coupling. They mentioned that this amplitude distribution changes from exponential-like to lognormal-like with the increase of the coupling effect i.e. the trap coupling effect can change the form of amplitude distribution functions [70]. Wang *et al.* [67] further claim that an increase in Vg changes the RTN amplitude distribution function in FinFETs. The distribution function is important for the prediction of RTN impacts on circuits towards high-σ tail [67], where σ is the standard deviation. Therefore, the long tail of the statistical distribution of RTN amplitudes has become one of the major issues in circuit designing which can only be predicted with a correct distribution model.

1.4 Existing RTN characterisation methods

RTN plays an important role in the field of miniaturizing transistor dimensions. The trapping and de-trapping phenomenon of charged carriers is one of the origins of noise in electronic devices [11], [71]. There is a need for a compact model to precisely predict and simulate the impact of random telegraph noise in advanced integrated circuits. The following sections of this chapter will introduce and discuss some of the typical existing methods that are used to study RTN signals for its parameter extraction and characterisation. Their underlying advantages and drawbacks and some open issues will be discussed.

1.4.1 Time Lag Plots (TLPs)

The main aim of characterising RTN is to extract the statistical features of the noise signal, starting from the unprocessed data. A time lag plot is also referred to as a lag scatter plot, which is a tool for analysing the autocorrelation in time series data that can also be used as a tool for analysing time domain RTN measurement [21]. TLPs are a graphical representation of measured data. It is plotted by taking data at the i^{th} time interval, t_i, against the data at i^{th}+1 time interval, t_{i+1}. Figure 1.4.1 shows an example of TLPs for a signal trap RTN taken from the experimental data of this work. The details of the experimental measurement and extracted data are discussed in Chapter 2 and Chapter 3 .

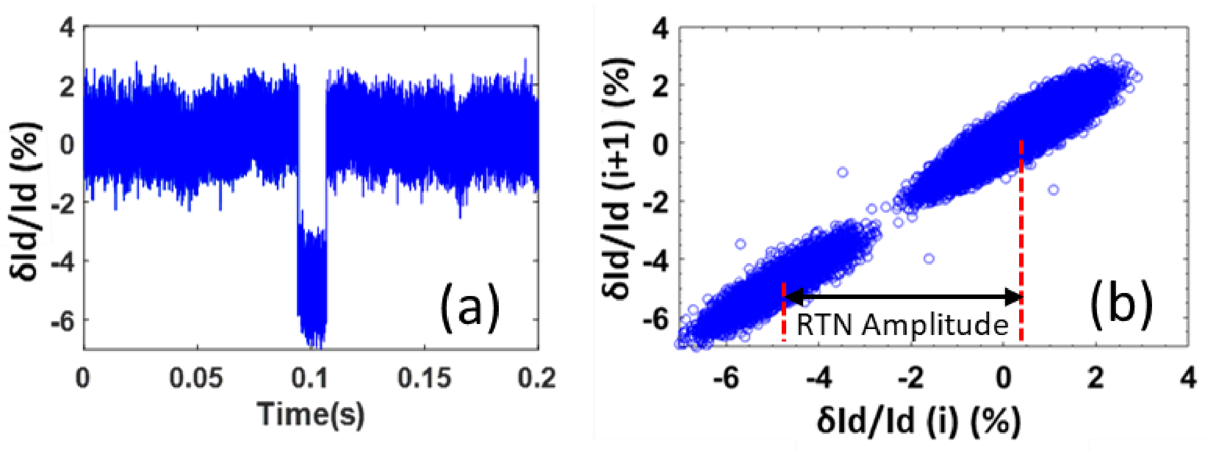


Figure 1.4.1 (a) A simple two state RTN single in time domain taken from experimental data and (b) corresponding simple TLP, showing only two clusters of data points for a single RTN trap.

From TLPs, the amplitude of RTN signal can be extracted. As each cluster in the plot represents each state of a single RTN trap, the amplitude of the RTN can be determined by taking the distance between the centre points of the two clusters as marked with red dash lines in Figure 1.4.1 (b). Hence, TLP can be used to extract the RTN amplitude only and for the time constant, the ratio of the high and low time constants follows from the ratio of the counts in both clouds [71].This method is difficult to use for analysing long-term RTN data where there is more than one trap, as shown in Figure 1.4.2. This is due to the lack of an unambiguous way of identifying RTN levels specifically in the presence of system noise or white and 1/f noise in the measured data [6].

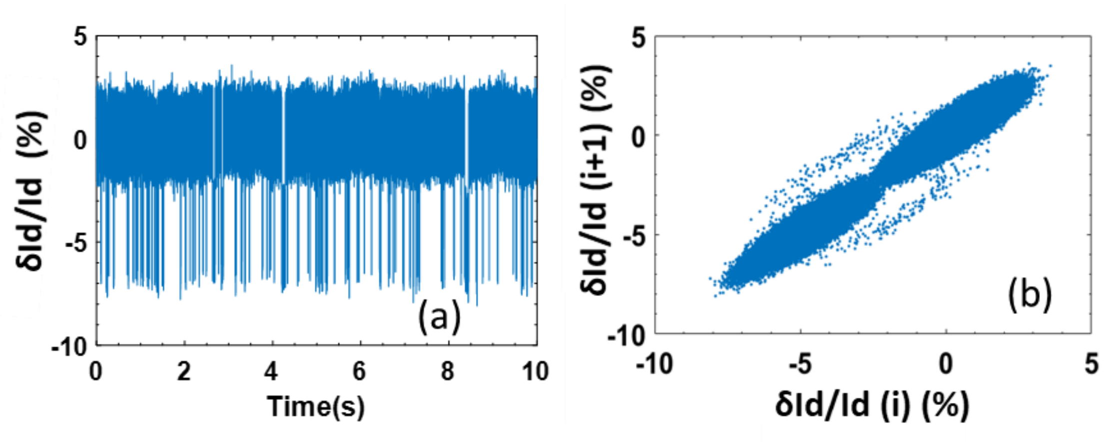


Figure 1.4.2 (a) Long term RTN measurement in time domain and (b) corresponding TLP showing ambiguous results to extract the correct RTN parameters by following this method.

1.4.2 Weighted Time Lag Plot (w-TLP)

w-TLP is a refined version of TLPs that is also referred to as coloured TLP. It has been designed to overcome the limitations of TLPs by minimizing the effect of the background noise in the RTN data, which impedes the detection of the accurate RTN states. This method is implementable and robust even with a lot of background noise [71][72]. The plot has the same starting point as TLP with coordinates [I_D (t_i), I_D (t_{i+1})]. In addition, this method can be used to extract two traps i.e. up to 4 RTN states by applying different weights to different data points in the clouds as shown in Figure 1.4.3 (d). In comparison to Figure 1.4.3 (b) which is a TLP of RTN signal shown in Figure 1.4.3 (a), a different colour or weight is applied to the data set which makes it possible to distribute 4 different current levels as shown in the histogram Figure 1.4.3 (c) and the coloured TLP of (d). The data clouds become more visible based on the density of the data as shown in Figure 1.4.3 (d). For the parameter extraction, the centre of gravity of each cloud is then calculated by taking the magnitude of the data points around the clouds, the distance between points and the grid in the cloud [71].

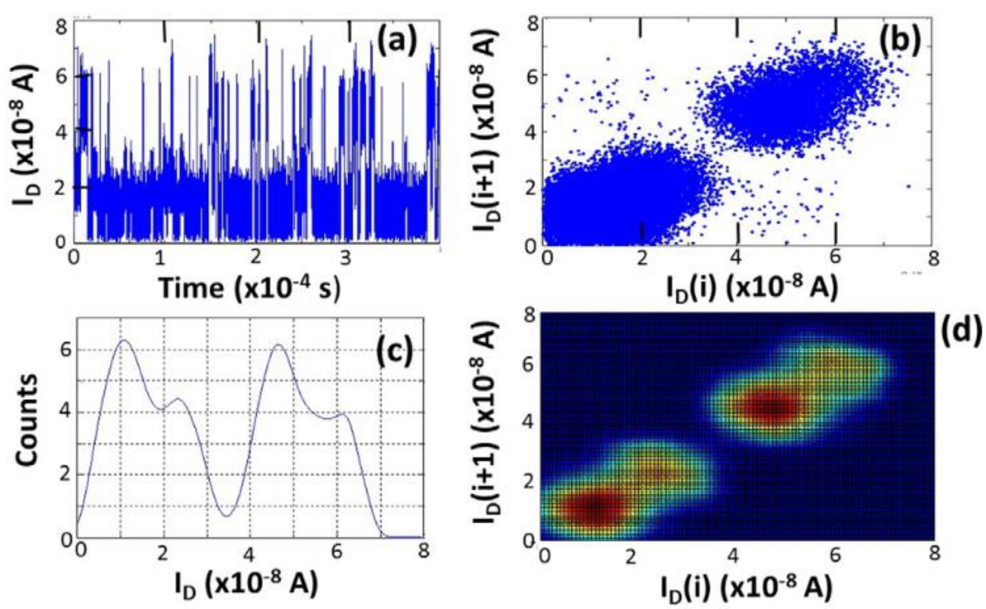


Figure 1.4.3 (a) Time series generated by the hidden Markov routine in Matlab and containing a small and a large amplitude RTS; (b) corresponding simple TLP, showing only two clouds of datapoints; (c) amplitude histogram derived from the time domain signal and (d) colored TLP, showing the presence of the two RTSs. [71].

This procedure yields the weighted magnitude for one grid point and in this way the other current levels can be determined. One more advantage of this method is that it can be used to extract the number of states of unknown RTN signal according to Puglisi's report in 2017 [73]. The limitation of this model is the accuracy of the extracted values as the trap location is estimated and the selected average data cluster size from the centre point is an assumption from the centre of gravity of the cloud.

1.4.3 Gaussian Method

In the literature, one of the most frequently used approaches for extracting RTN parameters from the simple case of two state RTN signal is by fitting the histogram of the measured RTN data as shown in Figure 1.4.4 [73]. This approach is known as the Gaussian method. By fitting the data shown in Figure 1.4.4 (a) with bi-Gaussian distribution model, it is possible to extract the two discrete levels of the RTN trap which will be the difference in the average of the two Gaussian distributions. Afterwards, the time constant values can be extracted by the combination of the amplitude histogram with a noise spectrum in frequency domain. This will provide the necessary information to determine the capture and emission time constant [71]. The PSD spectrum can be calculated by equation (13) :

$$S_I = \frac{A_0}{\left[1 + \left(\frac{f}{f_0}\right)^2\right]} \tag{13}$$

where, S_I represents the noise current spectral density, A_0 is the plateau amplitude at low frequency and f_0 equals the corner frequency of the generation-recombination (g-r) level [71]. Afterwards, the time constant (τ_c, τ_e) values are extracted by deriving equation (13) with the following equation (14) :

$$2\pi f_0 = \frac{1}{\tau_c} + \frac{1}{\tau_e} \tag{14}$$

However, the accuracy of the extracted value of τ becomes lower when there is more than a single trap RTN.

The main assumption of the Gaussian distribution method is that the background noise such as the noise from the instruments and surroundings are considered as Gaussian. The Gaussian distribution is also used for extracting multiple states of RTN from the plotted histogram, known as the

Gaussian mixture as shown in Figure 1.4.4 (b). However, this method is difficult to extract in multiple states of RTN signals. The limitation of this method is when the RTN levels of different states are close to each other, the plotted distributions in the histogram will be close and overlap with each other, so that the precision of the histogram fitting degrades. In addition, the Gaussian method is also not reliable for the prediction of the RTN amplitude distribution. Therefore, a quantitative analysis of complex RTN signals needs more advanced techniques and tools.

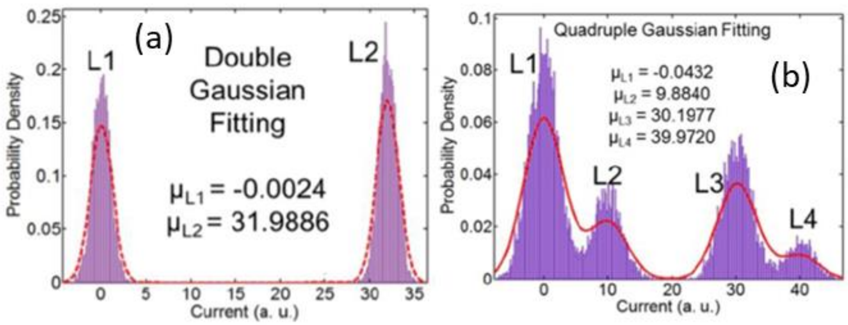


Figure 1.4.4 Puglisi et al., 2017. [73]

1.4.4 Hidden Markov Model (HMM)

HMM is an advanced and the most popular and versatile model available in MATLAB based software. It is considered a powerful tool in the field of pattern recognition and statistical signal analysis. It is widely used to extract the statistical properties of RTN signal caused by traps. HMM model assumes a Markov (memoryless) process characteristic with hidden states i.e. the discrete RTN levels. Generally, the Baum-Welch algorithm [74] is used to estimate the parameters of the hidden Markov chain along with a log-likelihood estimation maximization (EM) approach. Another approach is using the Viterbi algorithm [75] that extracts the most likely path based on the measured data and extracted HMM parameters. Figure 1.4.5 demonstrates some simple noisy RTN test data which is fitted with an HMM waveform from which RTN parameters are extracted. The probability distribution of capture and emission times extracted is expected to have exponential distribution and their average gives τ_C and τ_e values. The amplitude fluctuation is obtained as the difference between the discrete RTN current levels.

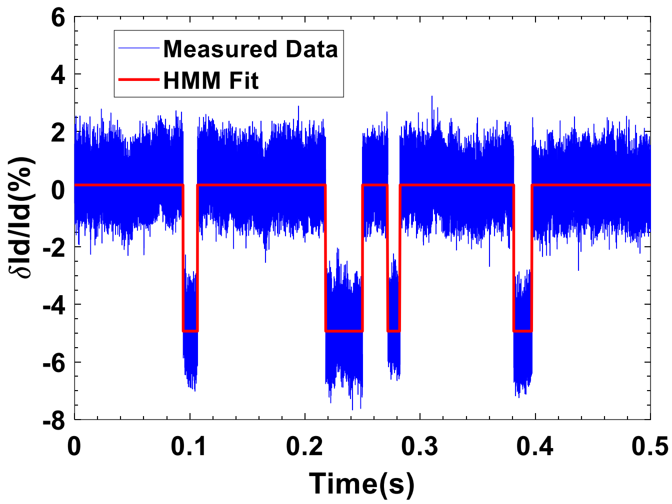


Figure 1.4.5 Measured RTN data and extracted idea RTN waveform using HMM simulation.

The limitation of HMM approach is same as those affecting TLPs and histograms fitting methods i.e. facing challenges in characterising multilevel RTN signal. When using the HMM algorithm, it is required to make assumptions on the number of hidden states. In the case of incorrect assumption of hidden states from the original observed RTN signal, applying HMM can end up with incorrect extracted values.

1.4.5 Factorial Hidden Markov Model (FHMM)

FHMM is a refined version of HMM, proposed for the statistical analysis of complex or multi-level RTN from multiple traps. The measured signal is assumed to be a superposition of a number of independent two-level RTNs, with each of them from one trap and modelled by a Markov chain [72][77][78]. Figure 1.4.6 shows an example of such a model used to extract the parameters of multi-level RTN [73] . Although this raises the number of RTN traps analysable from one device, it becomes increasingly difficult to apply as the number of traps in a device increases with the time window.

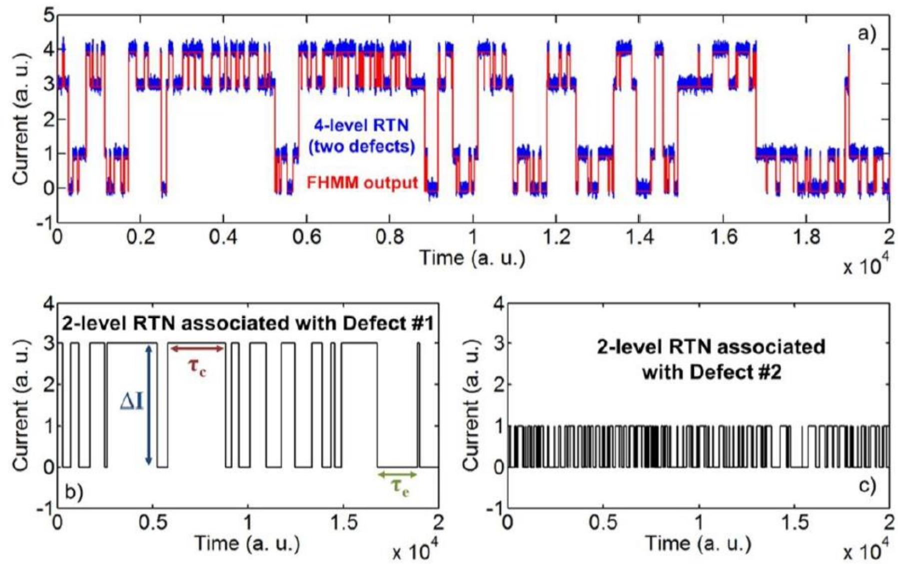


Figure 1.4.6 An example showing the use of FHMM for RTN parameter extraction by Puglisi in 2017. [73]

F. Puglisi in 2017 [73] claimed that among all the mentioned models for RTN parameter extraction, HMM and FHMM provide better accuracy compared to TLP, WTLP and Gaussian fittings. Besides these models, there are other approaches as well which have been used in the literature. One of them is called the Markov Chain Monte Carlo (MCMC) model which applies the Gibbs sampling algorithm [79]. Currently, all these methods (HMM, FHMM and MCMC) face some common issues such as:

- Limitation of accuracy when the superimposed RTN is strong i.e. the accuracy of the extracted parameters decreases as the number of RTN states increases.
- The difficulty of dealing with instabilities and temporary random phenomena.
- Limited measurement time window, as large data sets are difficult to process and handle.
- Typically, all these models assumed a Gaussian distribution of the superimposed noise which may not always be true.

1.5 Thesis Outlines

This thesis focuses on characterising random telegraph signals induced by as-grown traps by introducing a new RTN measurement technique for long-term prediction and modelling of RTN impact. To study and obtain such information, the main problem of the limited measurement window will be tackled first in this work. This is followed by an assessment of the statistical distribution of RTN parameters.

This thesis is divided into seven chapters. Chapter 1 is the literature review. A brief introduction and background information of random telegraph noise is discussed along with its phenomenon, origin, and some existing characterisation methods.

Chapter 2 describes the devices and test facilities used in this research and also some basic electrical characterisation techniques by means of current-voltage (Id~Vg) measurement. The recently developed Within-Device-Fluctuation (WDF) measurement technique is presented.

Chapter 3 the methodology and details of the new RTN overnight measurement technique is provided in steps.

Chapter 4 investigates the statistical distribution of RTN time constants. A methodology for efficiently characterising the capture and emission time distribution for a fabrication process is proposed and for the first time, the long-term prediction capability of capture/emission time distribution beyond the time window used for its extraction is verified.

In Chapter 5 , the accuracy in modelling the statistical distribution of random telegraph noise (RTN) amplitude has been studied. A novel criterion for selecting RTN amplitude distribution has been proposed. The uncertainty in predicting the RTN amplitude distribution by different statistical models is highlighted.

Chapter 6 represents the gate voltage impact on the statistical distribution of RTN amplitude and time constant and discusses its unique features that make it suitable for each statistical case.

Finally, Chapter 7 draws the main conclusion of the research and provide some remarks, suggestions and proposals for future research in this field.

Chapter 2 Devices and Test Facilities

2.1 Introduction

This chapter will represent the devices and test facilities used in this research work, along with an overview of basic electrical characterisation methods of CMOS devices. Firstly, the available test facilities required to conduct the experiments is described. Next, the wafer used for testing and the device specifications are discussed. Thirdly, the basic methodology widely used for characterising noise in CMOS devices is briefly presented. Finally, the details of the recently developed Within-Device-Fluctuation (WDF) measurement technique is explained.

2.2 Test Facilities

Noise measurement is a sensitive process and needs careful handling. The equipment used in this research to conduct the measurements includes a personal computer for programming, a pulse generator (Agilent 81160A), a Femto DHPCA-100 amplifier, a home-made electrical circuit, a parametric measurement mainframe (Agilent E5270A) with 4 source measure units (SMUs), and two 4-channel oscilloscopes (Agilent infiniium MSO8104A / MSO9404A). The tests are conducted by using the Cascade Summit Probe Station with a controllable heating plate to maintain a pre-set temperature.

The measurement starts by conducting the conventional transfer characteristic (IV) sweep (details in the section 2.4.1) to quickly check if the connections and devices are working as expected before proceeding further. The available control program was written in Visual Basic (C+) language. The pulse generator Agilent 81160A is used to generate the input voltage which is connected to the gate terminal of the Device Under Test (DUT). The constant drain voltage is provided via the home-made circuit which is powered by Keysight E3632A DC power supply. The output drain current (Id) of the device is then amplified and converted to an output voltage that is monitored and recorded by the oscilloscopes. The data is then transferred to the computer to manage the storage limitations of the oscilloscope. The overview of the system is illustrated in Figure 2.2.1.

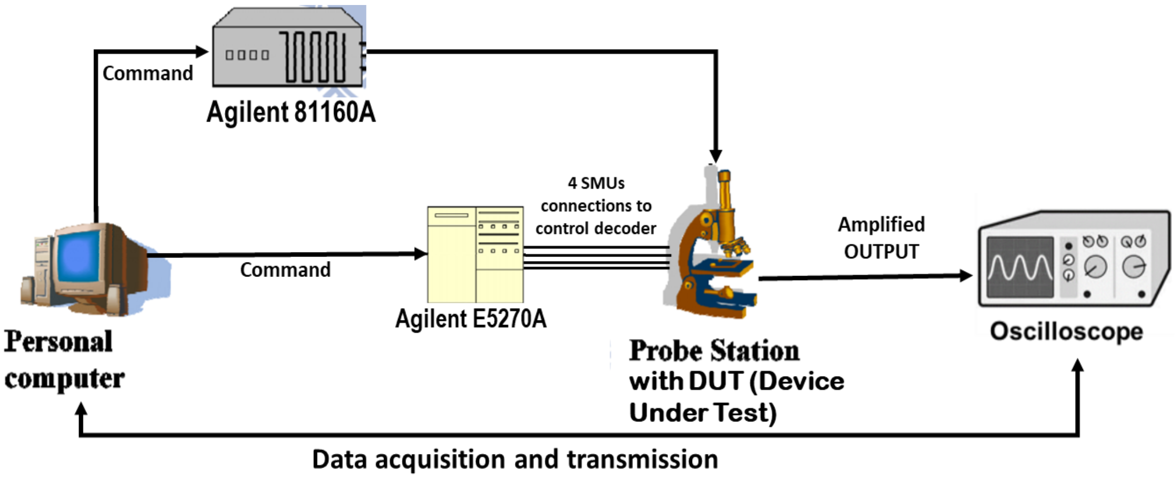


Figure 2.2.1 Illustration of measurement set-up.

2.3 Wafer and Device specification

The sample used for this work is from CSR plc (formerly Cambridge Silicon Radio, acquired by Qualcomm in 2015). This wafer is fabricated by Taiwan Semiconductor Manufacturing Company (TSMC) with a 28nm semiconductor technology node. This research work uses nMOSFETs from the CSR wafer which has a metal gate and a N28HPL high-k/SiON stack. The channel length and width used are 27×90 nm, respectively. The equivalent oxide thickness is 1.2 nm. The maximum and minimum temperature used for the test are 125°C and 28°C, respectively. Figure 2.3.1 shows the sample of CSR wafer that is used in this work.

However, the kind of the devices in CSR wafer were not mentioned in its datasheet but the devices were not FinFETs. FinFET device are more challenging than for planar devices predominantly due to reliability and performance concerns associated with the complex geometric structure. For FinFET, the influence of RTN depends on the relative distance between the charged trap and dominant current path. For example, the trap located in the middle region between source/drain and around the bottom of sidewall interface would exhibits the highest impact. Unlike planar devices, FinFET actually has intrinsic percolation path (PP) even for the uniform device due to its nonplanar structure nature. The intrinsic PP location changes with increasing Vg in FinFET. This issue has not been tackled in this work.

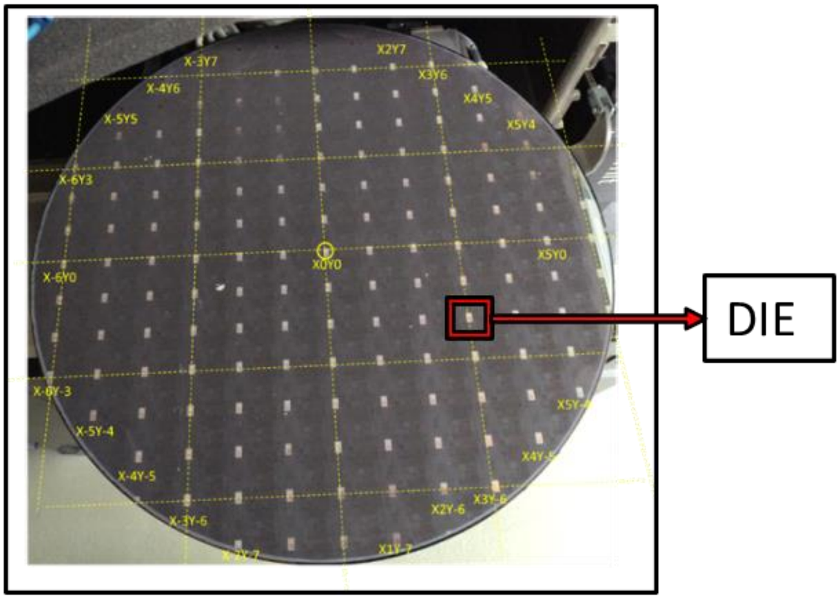


Figure 2.3.1 Sample of CSR wafer.

This wafer contains two groups of devices: individual devices and array devices. These groups of devices can be accessed by two probe cards 78662 (Cantilever Probe Card) specifically designed to probe these embedded standalone transistors. The right side (green part) of the probe card in Figure 2.3.2 (a) is used to select a device size by make the gate, source and drain connections with the system. One is designed for the individual devices and another one is for array devices only. The differences between these two probe cards are: the array probe card has the decoder connections to automatically handle a set of devices within a DIE. In contrast, the individual probe card is only designed to manually access each transistor. Figure 2.3.2(a) shows the design of the probe card used in this work. In this research, NMOS array devices have been used. A single DIE has 16 identical transistors in total, where an individual transistor is at the centre of the die, as shown in the schematic diagram [80] in Figure 2.3.2 (b). These array transistors are accessed by a 4-bit decoder. The array devices and the decoder are connected according to the layout of the CSR devices and probe card. The decoder is controlled by 6-SMUs where 2-SMUs are used to supply Vdd and GND to the decoder and 4-SMUs from the Agilent E5270A are used to supply the voltage to the four bits of the decoder. The overview of the Agilent E5270A 4-SMUs connections is shown in Figure 2.2.1. The decoder is activated by assigning Vdd = + 0.9V. Once the decoder is active, it is possible to access one of the 16 array DUTs according to the bit1~bit4 assignment from 0000 to 1111. Here, 0 means 0V and 1 means + 0.9V.

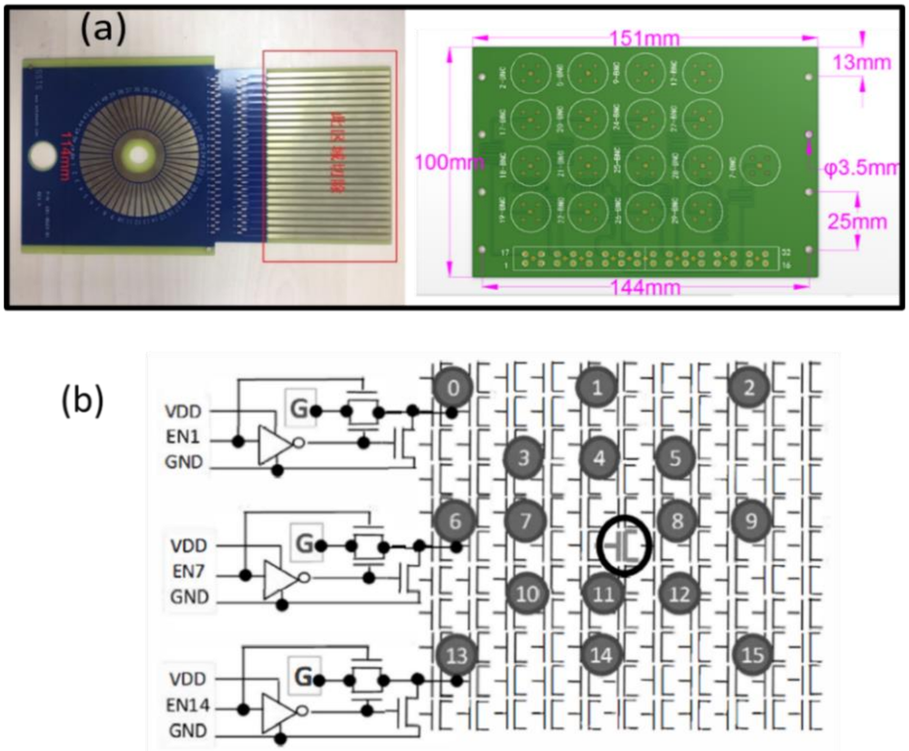


Figure 2.3.2 (a) The probe-card and (b) the schematic representation of Decoder Enable lines connecting to transistors (numbered 0-15) and a center transistor (circled) [80].

2.4 Electrical transistors characterisation

This section gives a brief introduction about the experimental methodology that is commonly used for monitoring the charge trapping dynamics in a reliability study of MOS transistors. Device level electrical characterisation is a fundamental tool that is used to study and verify the basic characteristics of a transistor. It is important in the modelling of a device in order to calibrate the model. The basic elements for characterisation of instability are I_d-V_g measurement and V_{TH} extraction. Here, a brief review of these two methods is presented.

2.4.1 Conventional I_d-V_g characterisation

Conventional I_d-V_g characterisation is usually the first step to evaluate a device by measuring the currents at the terminals versus the applied voltages. The instrument that allows this kind of measurements is known as a parameter analyser. The IV sweep measurement is the most conventional characterisation method and can be performed using commercial Source Measure Units (SMUs) [81]. SMUs are used to measure current and voltages by applying a V_g sweep at the gate of the MOSFET under a constant drain bias (V_d). The typical waveform and the pattern are shown in Figure 2.4.1. The usual speed of this kind of I_d measurement is around one to ten seconds for a full IV sweep depending on the step size used. Thus, this conventional measurement procedure is also called quasi-static or slow DC measurement. This kind of measurement can accurately measure small currents down to 10^{-15} A [57]. Many properties and parameters of a device can be extracted from IV characterisation such as, mobility, threshold voltage etc.

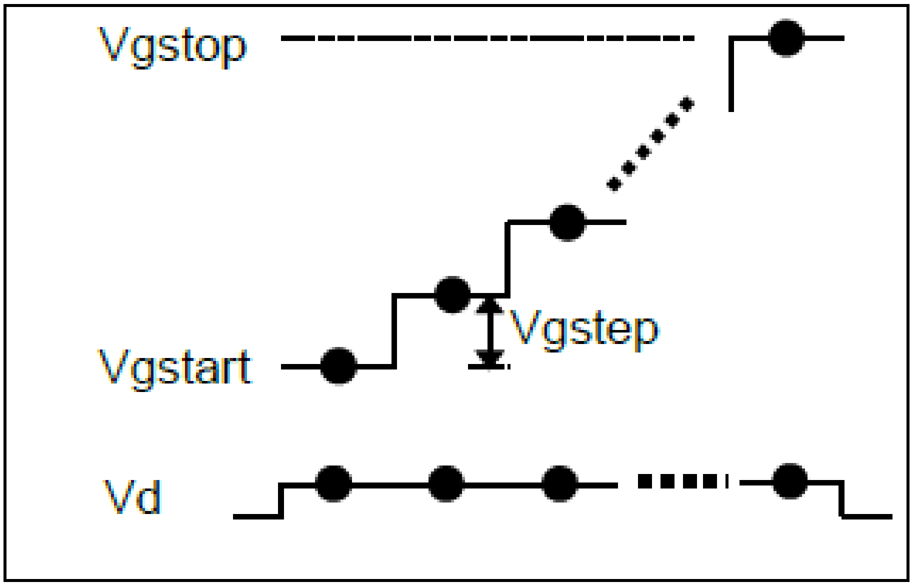


Figure 2.4.1 Typical IV test waveform. A constant drain voltage Vd is applied using SMU and another SMU used to provide the Vg sweep and voltage and current are simultaneously recorded for each step from several to hundreds of milliseconds [57].

2.4.2 Threshold Voltage extraction

As mentioned in the previous section, IV measurement is a useful tool in the case of parameter extraction of MOS devices. One of the most important parameters is the threshold voltage shift (ΔV_{TH}). It is a key parameter when considering RTN characterisation, as ΔV_{TH} fluctuation is caused by the trapping and de-trapping of charge carriers. There are many different methods of extracting the threshold voltage of MOS transistors from IV measurements [82]**.** The two most powerful and easy methods are the maximum g_m and constant current methods which are briefly described below.

The max-g_m method is the most popular way of extracting V_{TH}, proposed by Ghibaudo [83]. The procedure for extracting V_{TH} is shown in Figure 2.4.2. The figure is plotted in MATLAB. First, the g_m (transconductance) of the IV curve is calculated by finding the first derivative (blue curve in Figure 2.4.2) of Id~Vg curve and plotted. At the maximum point of g_m, a tangent is drawn from the max-gm Vg value as shown in the Figure 2.4.2 (the blue dash line). The intercept of the tangent line on the x-axis of Vg is the extracted V_{TH} (blue square box in Figure 2.4.2). On the other hand, for constant current method a line is drawn on the IV curve (black dash line) at I_d = constant drain current (I_{dcc}). A typical I_{dcc} is 100nA×W/L. The corresponding V_g value at the crossover point of this line is the extracted V_{TH}.

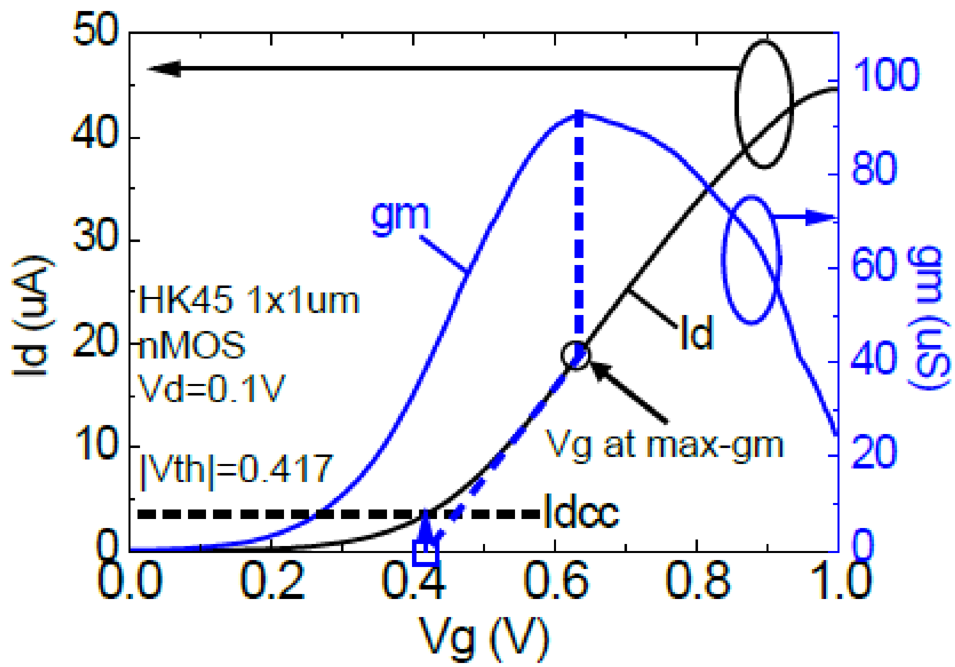


Figure 2.4.2 Max-gm (blue dash lines) and constant current (black dash line) method extracted Vth. Here Idcc is properly selected to intercept at the same Vth as max-gm method on a fresh device [57].

2.5 Within Device Fluctuation (WDF) technique

As the impact of a single defect increases with the scaling down of device size, the I_d fluctuation due to an individual defect on nano-scale devices cannot be averaged out as the number of defects in a single nano device is decreasing. This phenomenon results in Within Device Fluctuation (WDF), a measurement technique proposed by M. Duan [84], [85]**.** Duan *et al.* [86] proposed that WDF is a quick and concise analysing method for characterising RTN signals. The authors also reveal that WDF mainly originates from the stochastic charging-discharging of defects that are located near Fermi level (E_F) at the interface as presented in Figure 2.5.1. For a given Vg, the E_F is below the silicon valence band (E_V) where the traps present are only as-grown traps [86], [87]**.**

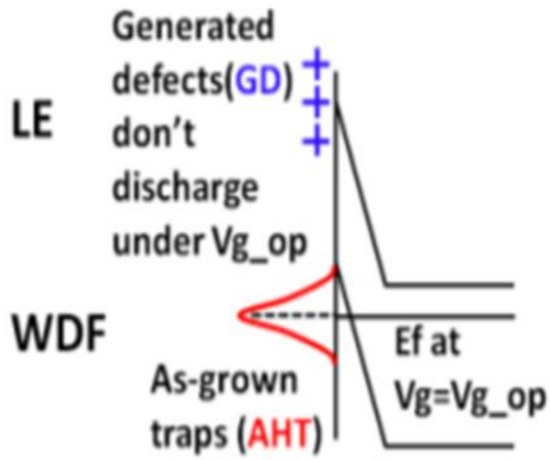


Figure 2.5.1 A schematic illustration of the energy location of two groups of defects: AHTs and GDs [87]..

When there are more than 3 traps in a device, then the step-like Id fluctuation induced by RTN is replaced by a complex within-a-device-fluctuation as shown in Figure 2.5.2 [87]. To analyse the I_d fluctuation as well as to reduce the huge data storage, the 'Envelope' (Env) of the fluctuation is used, which represents the worst case of the I_d fluctuation. That is, the maximum is the upper envelope, UE (worst case), and the minimum current fluctuation is the lower envelope, LE, and the difference between them is the WDF. The worst-case UE is the sum of LE and WDF (WDF = UE – LE). A typical WDF waveform is shown in Figure 2.5.2 (c) and (d). In Figure 2.5.2 (c) it is demonstrated that WDF behaviour with upper and lower envelope increases with stress time. Figure 2.5.2 (d) shows that an increase of measurement time window enhances WDF [18]. This is due to a large time window capturing slower traps that lead to an increase in WDF. The averaged WDF results are found to follow a logarithmic relationship against the time window. As reported by these earlier works [84]-[87]**,** WDF/RTN is caused by only as-grown traps (AT). Hence, this project deals with only AT traps and understanding the WDF phenomenon and its measurement is one important part of this work.

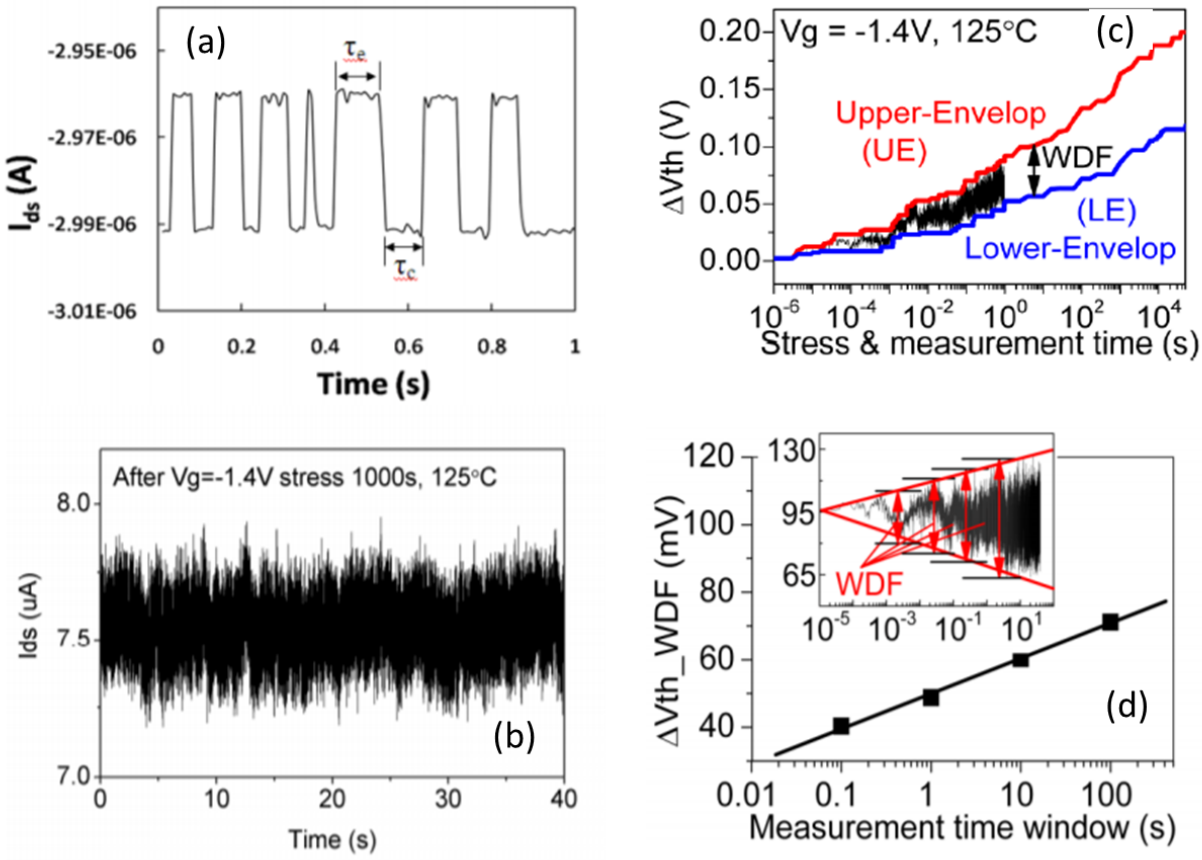


Figure 2.5.2 Examples of (a) RTN and (b) within-a-device-fluctuation (WDF). (c) The typical WDF behavior with upper and lower envelope and (d) A Increase of WDF with the time window (tw). The insight is taken from (c) which shows the upper limit of the envelope is around 0.2 V and Lower envelope is nearly 0.1 V [87].

Chapter 3 Development of new characterisation method

3.1 Introduction

In electronics, measuring noise is a complicated task, as the required measurement signal strength is very weak, nearly ~1 pA. Hence, the measurement set-up must be cautiously designed with appropriate shielding to avoid or minimise any unwanted disturbance from the surroundings. This research is focused on RTN analysis in the time domain and its test is designed with the help of an oscilloscope. The overview of the measurement set-up is presented in section 2.2. This section will follow up the explanation and discussion of the measurement set-up along with the purpose of developing the new technique that is used to conduct the test and collect data. In short, the objective is to present a clear visualization of the measurement methodology.

3.2 Methodology of new characterisation technique

The characterisation of RTN is mostly important for low power application circuits, where the operation voltage is slightly higher than the threshold voltage. The primary shortcomings of existing RTN characterisation methods are: not all devices have observable RTN i.e. no clear time constant which results in a need for careful device selection. The measurement time window used is very limited, e.g. 10 sec or less [88],[20] so there are few reports on slow traps. Thus, the purpose of the new technique is to provide the long-term RTN data by developing a continuous data acquisition process to extract more information for the assessment of RTN parameters statistically within the time domain. The procedure of the newly developed technique is described in the following sections.

3.3 Experimental Steps

This section describes the experimental steps for the RTN test. The aim is to successfully establish an overnight RTN test i.e. long time window ~10^5 sec (nearly a day) and to manage and collect the huge amount of data. Once the systematic test is ready, overnight RTN data is recorded from multiple devices for statistical analysis. The objectives are as follows:

- First, check and compare the system noise with the signal strength.
- Select a sampling rate by scoping test.

- Check transfer characteristics to make sure the connection and devices are working properly.
- Using different rates for data sampling and data recording to overcome hardware limitations with the help of two oscilloscopes.
- Systematic RTN overnight test for statistical results.

3.3.1 System Noise level

One of the main issues while conducting these measurements is the system noise. At a given DC bias, the current flowing through a device can be described as $I_D(t) = I_{DC} + i(t)$, where the I_{DC} is the desired signal at the applied bias and i(t) is the random fluctuating current due to external noise sources and fundamental physical processes. As mentioned earlier in section 1.2.1, this external noise is introduced from the surroundings of the system such as, cross-talk between circuits, any vibrations, electrostatics and electromagnetic coupling from power line etc. These disturbances cannot fully be eliminated but can be minimized by properly protecting the test environment by shielding, filtering, using vacuum chambers, and optimising layout etc. It is important to find out and reduce the system noise to a level that will not mask the desired signal. In this work, the noise level has been minimized by using aluminium foil paper shielding, copper wires, coaxial cables and an ACR Line-R 1200 noise canceller. Figure 3.3.1 represents a few blocks of the test set-up for visualization.

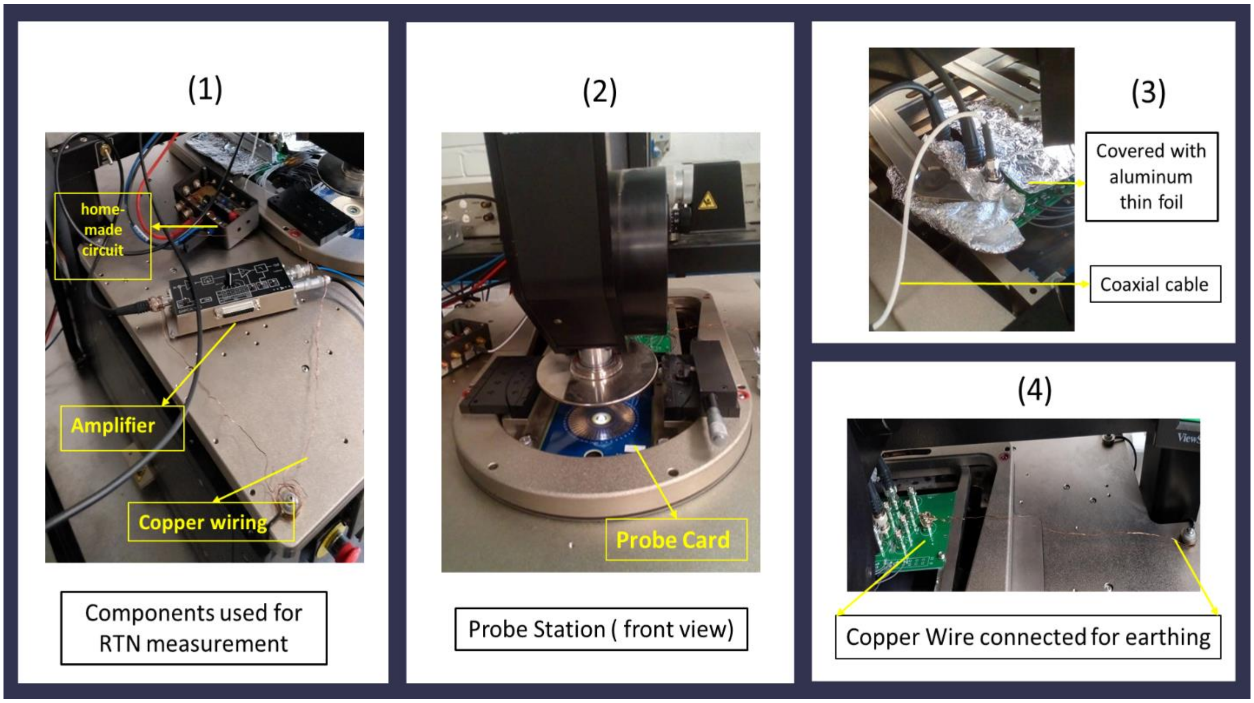


Figure 3.3.1 Experimental set-up representing in blocks.

Here, block (1) shows the Femto DHPCA-100 amplifier and the home-made circuit with batteries, to provide constant drain voltage, and its adjacent points are connected with the probe station using copper wires for earthing. Block (2) represents the probe card connection from the front view. Block (3) shows the use of aluminium foil. Since the system was not vacuumed, aluminium foil is used to cover some of the adjacent connections on the probe card to reduce disturbance from surroundings. Coaxial cables are recommended for noise reduction and the system was powered via an ACR Line-R 1200 noise canceller. The probe card is further shielded with copper wires as shown in block (4). Altogether, the amplified system noise for this work has been successfully maintained below ±1mV. Figure 3.3.2 is an example of the system noise (green traces) along with the expected RTN signal (black line). The measurement condition used is: Vg = 0.5V, V_D = 0.1V, temperature is 125 ºC under sampling rate of 1 MSamples/sec. As already stated, this work is focused on investigating the distribution of charging and discharging of as-grown traps only, hence low bias of 0.5V is chosen at gate for the tests to minimise the interference from any trap generation process [89]**.** Such near threshold operation lowers the power consumption and can be useful for low power applications, such as the edge units of the IoT, which suffer acutely from noise issues.

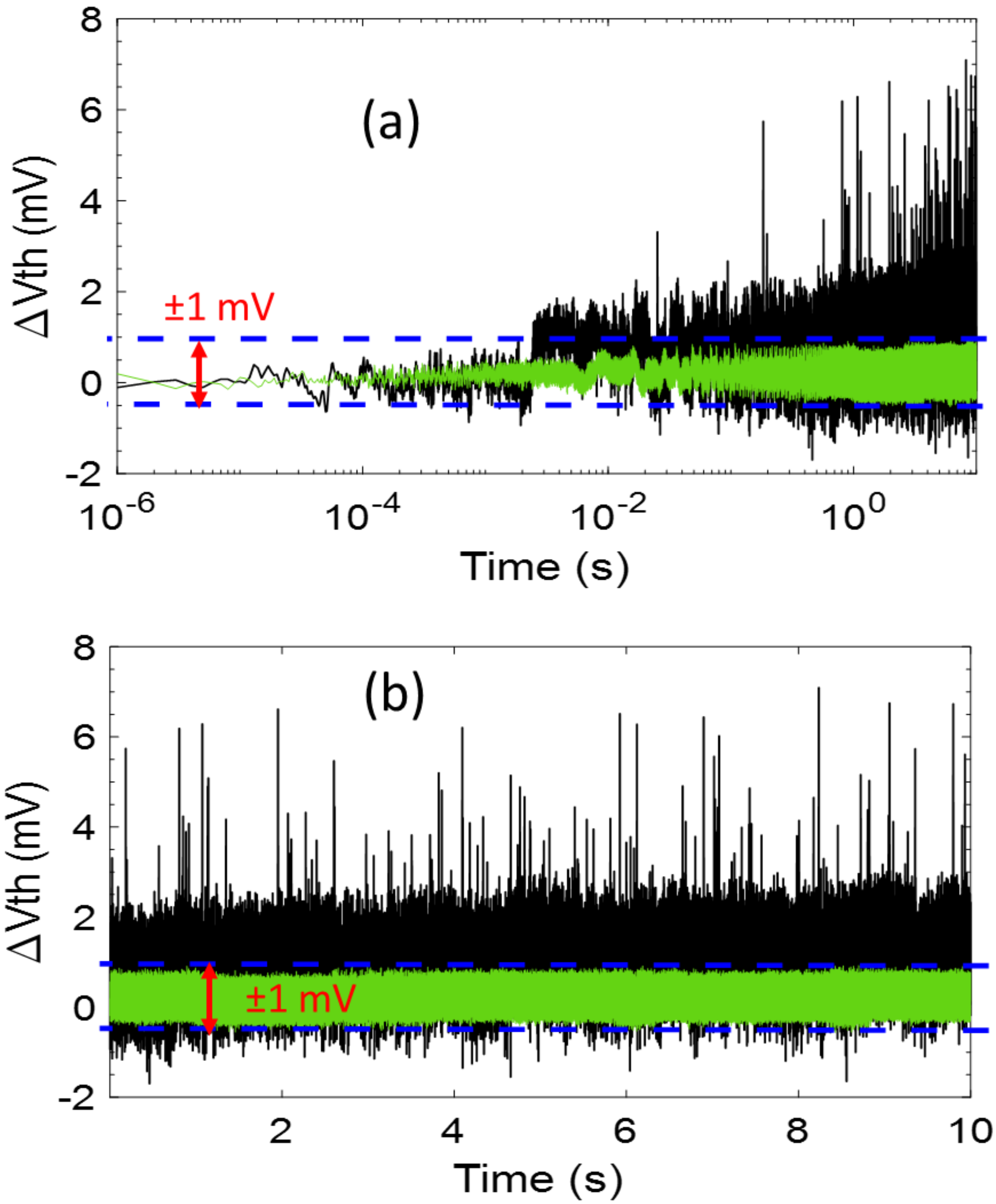


Figure 3.3.2 An example to represent the noise level of the system (a) in linear-log scale (b) linear-linear scale.

3.3.2 Sample Rate selection

One of the main challenges in characterising RTN signals is the difficulty of measuring small currents at a high sampling rate and how to perform this test for long time periods on many devices for statistical results. The continuous sampling of RTN fluctuation will need huge data storage capacity. The measurement time steps are related to the sampling frequency (Fs) in the signal by, Fs=1/Δt, where Δt is denoted as time intervals. As the time window increases, for example, with 1 MS/sec, where 'MS' is 'Mega-Sample points', the size of the data set for one measurement is 10 MS for a time window of 10 sec. The total measurement data points can be calculated by,

$$\text{Data Points} = \frac{\text{Time Window}}{\Delta \text{t}} \tag{15}$$

Therefore, the number of data points becomes larger with the longer time. For a time window of 10^5 sec (~ day), the total data size becomes intolerable as the data set size rises to 100 GS, which is beyond the memory depth of modern oscilloscopes. Hence, it requires developing new data acquisition techniques to control data size. This is achieved in this work by using different rates for data sampling and data recording (details given in section 3.5).

To choose a sampling rate for RTN measurement, a scoping test has been done. The RTN is measured under different sampling rates and then the extracted envelope (Env), as marked in red in Figure 3.3.3 (a)**,** is used to select the best sampling rate for this work. Figure 3.3.3 (a) is an example of an RTN signal sampled under 1 MS/sec. Although it shows that there are only a limited number of steps in the Env, it does not mean that a low sampling rate can be used to extract the Env. The plot in Figure 3.3.3 (b) is the Env at 10 sec time window that was obtained from the recorded test data at different sampling rates. It clearly shows that slower sampling rate leads to lower Env. This is due to it failing to capture the fast traps [90]. However, when the sampling rate reaches 1 MS/sec, the Env starts to saturate. Hence, a sampling rate of 1 MS/sec is used hereafter.

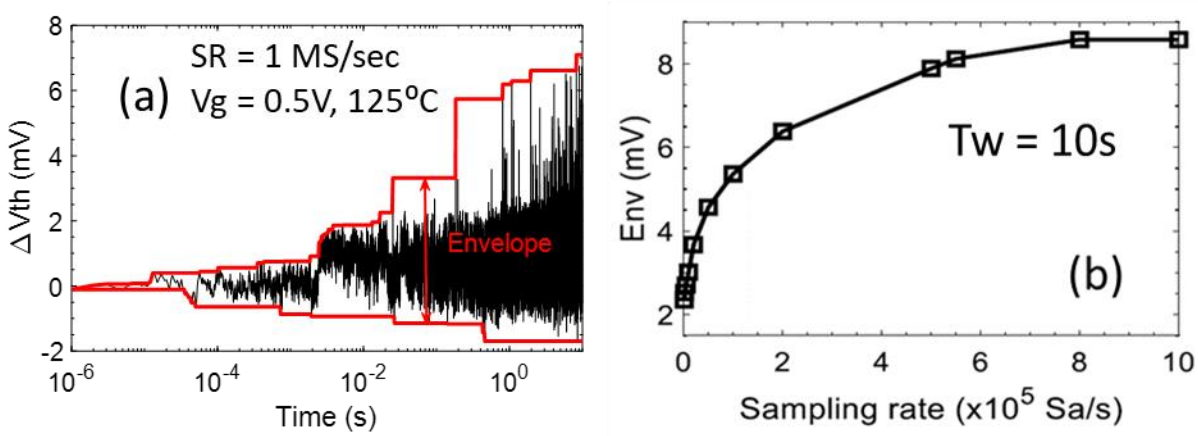


Figure 3.3.3 (a) An example showing extraction of RTN envelope from experimental data. (b) Evn at 10 second time window (Tw) extracted from (a) is plotted against sampling rate to depict the impact of sampling rate on the envelope.

3.4 Pulse IV measurement

The conventional Id-Vg measurement technique that is described in section 2.4.1 is not applicable for the fast transient instabilities in high-k materials and there is a need for a faster and more reliable measurement technique. Hence, the pulse IV technique is first introduced by Kerber *et. al* [91] and there are many works that have been performed by using this method [92]-[94]. This research also used this technique for the quick IV measurement to increase confidence in the basic connection and operation of the system. Also, the transfer characteristics (TC) for each device under test were recorded for further RTN analysis in the future.

The pulse generator is used to generate a pulse to be applied at the gate of the transistor. The constant drain to source voltage, V_D is applied at a low level of 100mV, with the help of a home-made circuit to eliminate any noise from bias circuitry. The amplified output, V_{OUT} is then monitored and recorded by a digital oscilloscope during the pulse edge which is further converted to the corresponding current value by simply dividing by the amplified gain, Id = V_{OUT}/Gain. The amplifier gain used for the devices of this work is 10k. This available pulse measurement system is a calibrated system used by previous researchers AT LJMU [57], [84]-[86], which ensures that the system has an acceptable noise level and performance. The waveform that is applied to the gate terminal is shown in Figure 3.4.1 (a) and (b) shows the typical results of a pulse IV measurement. Figure 3.4.1 (c) shows the typical expected TC behaviour from the applied pulse IV sweep on the CSR 27×90 nm nMOSFETs which depicts that the system is working as expected. The system is ready to carry out RTN measurement.

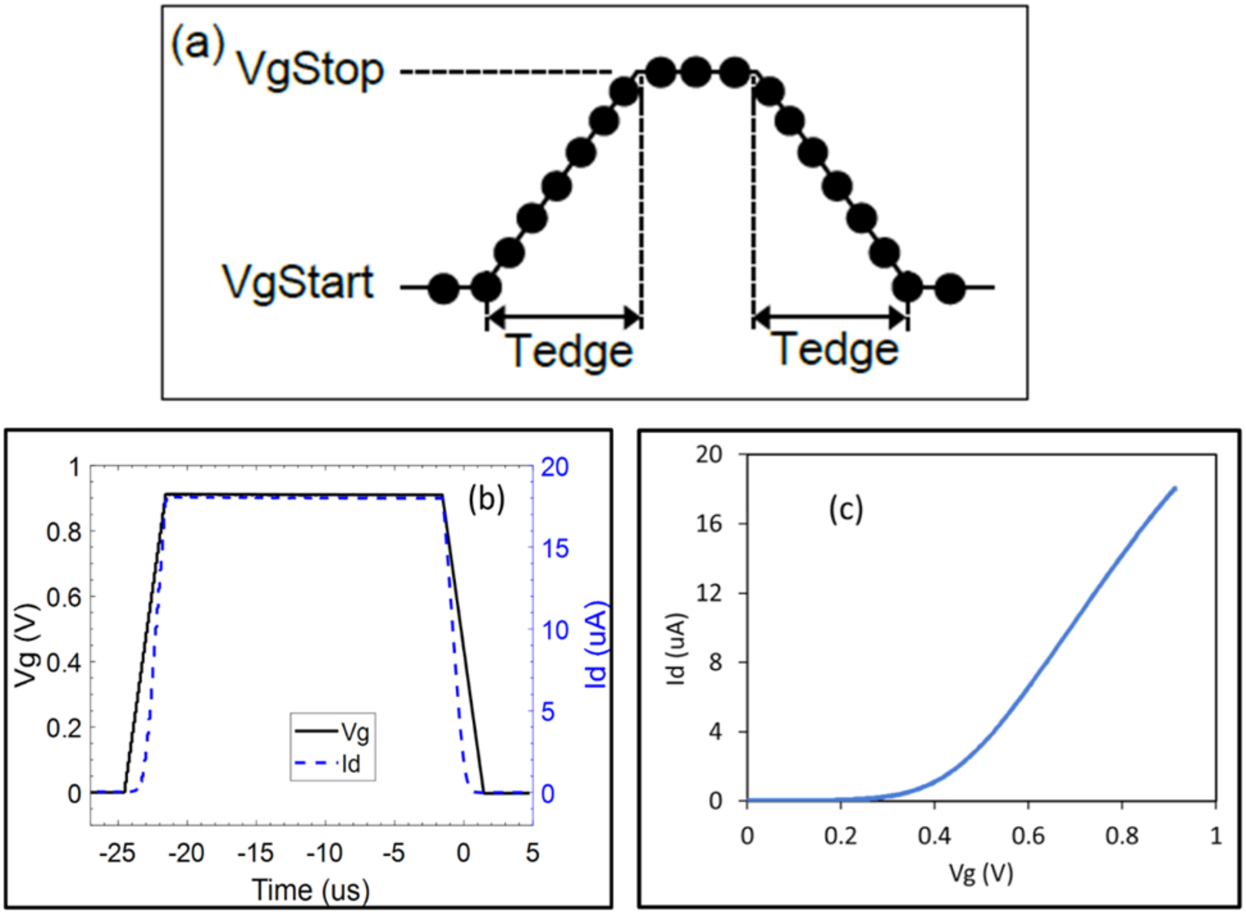


Figure 3.4.1 (a) Typical waveform of a Pulse IV measurement [57]. (b) Typical results of pulse IV. (c) The expected (Id~Vg) transfer characteristics of the device.

3.5 RTN overnight test

The conventional RTN is measured from the drain current (Id) of a device that occurs due to switching between two discrete levels under a constant gate bias, Vg and drain voltage Vd. In this work, the test starts by measuring a pulse Id~Vg with Vd as 0.1V and pulse edge time of 3us. The Vg is stepped from zero to 0.5 V and Id is monitored against time under Vd = 0.1 V. The waveform of Vg and Vd used in this work are showing in Figure 3.5.1. The average threshold voltage (Vth) of the nMOSFETs used here is 0.45 V and Vg is chosen to be Vth+0.05, as the requirement of low power is driving Vdd towards Vth and near threshold computing acutely suffers from RTN [95]. The temperatures used are 28 ºC and 125 ºC.

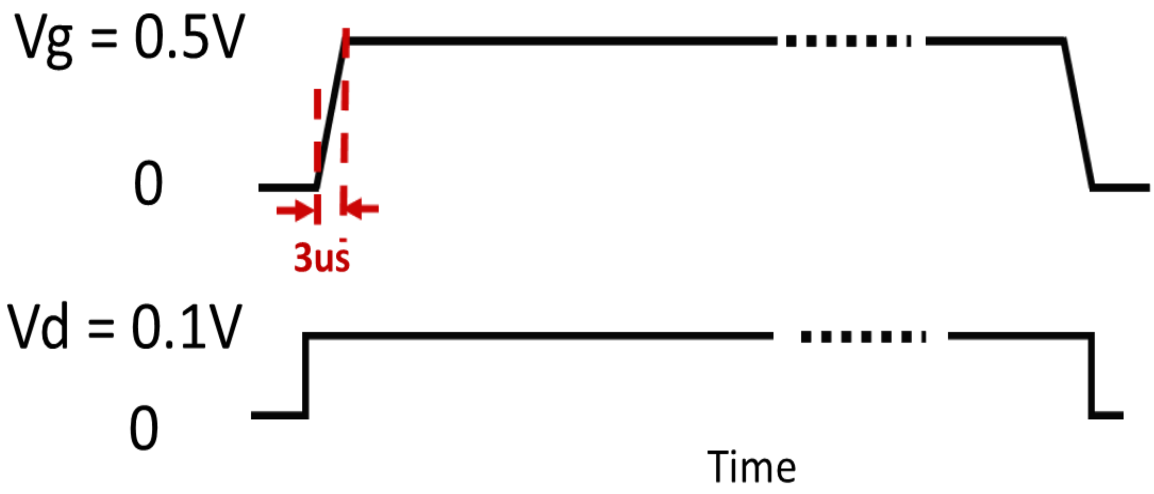


Figure 3.5.1 Applied voltage waveform for conventional RTN measurement.

In this research, the hardware limitations that have been taken into account are: handling too much data, limited storage space and time-consuming processes. To handle huge amounts of data, two oscilloscopes have been used to monitor current, Id. One of them has a time window of 10 sec and every data point is recorded. The other has a time window of 10^5 sec and monitors the maximum and minimum points of the signal at 1 MS/sec. The oscilloscopes automatically detect the max and min values of the Id current on the screen and that value is recorded every 20 seconds for the overnight test. This is because the purpose is only to depict the envelope and most of the time the envelope remains the same (refer to section 4.3.2). This is how the data storage rate is made different from data sampling rate. To overcome the oscilloscope reaching the memory limit, the data is recorded and directly saved in the PC with an external hard drive. As the entire test will be time-consuming, the best possible methods have been adopted to carry out the expected test while maintaining reasonable data collection time. Such as how much data can be saved in the PC in a reasonable saving time? A quick scoping test has been done to find the solution, as shown as an example in Table 1. The values in Table 1 are the data transferring and saving protocol from the oscilloscope to the personal computer. Figure 3.5.2 shows that the time for saving data directly in the PC increases as the data size increases. From the plot and the table, it could be calculated that the data saving time increases approximately 10 times with every decade increase of data size.

Table 1 Data acquisition protocol from oscilloscope to PC.

Time window (ms)	Number of data points	Time to save data ~ (ms)
1.00E-03	1	22
1.00E-02	10	22.6
1.00E-01	100	34
1.00E+00	1000	44
1.00E+01	10000	100
3.33E+01	33300	300
1.00E+02	100000	990

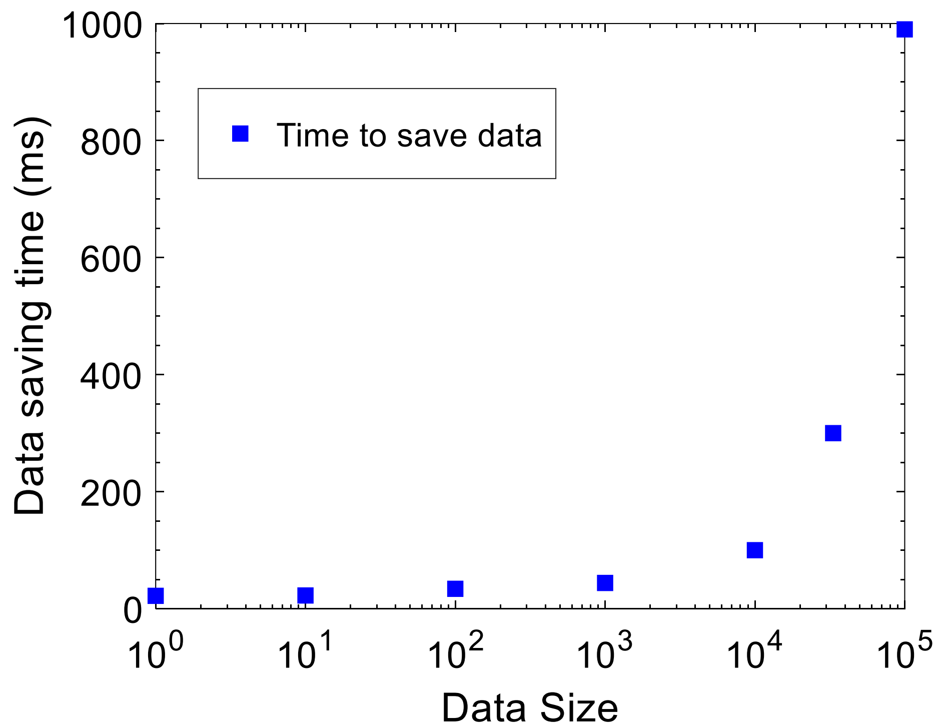


Figure 3.5.2 Data saving time against data size.

The systematic characterisation protocol of the work is summarised in the flow chart Figure 3.5.3. The whole test was automated to carry out the steps shown in the flow chart for each measurement. At first, a quick IV check is done and recorded. The system is then stopped to get ready for the RTN test by applying the bias shown in Figure 3.5.1. Initially in the RTN test, 10 sec of data is saved. The aim is to capture the fast traps under 1 MS/sec. Then, the measurement is restarted to collect continuous overnight RTN data at a constant applied gate bias. As stated, two oscilloscopes (OSC) are used, OSC 1 monitors RTN raw data and records it every 0.1 sec. The 0.1 sec is the minimum delay and saving time of 1 data point from the OSC needed before the OSC screen updates and allows the next data point. If delay time, less than 0.1 sec is used then there will be overlapped of saving same data point twice. Since, the purpose was to save continuous acquisition of RTN data as quickly as possible, the minimum delay time of 0.1 sec has been used rather than using a greater value. Here, the main intention was to capture the slow trap response in the presence of fast traps and find any relation between RTN time constant distribution with time window. On the other hand, OSC 2 is used to monitor the envelope of the signal. The target was to collect data to find the real time constant distribution without considering any parameters.

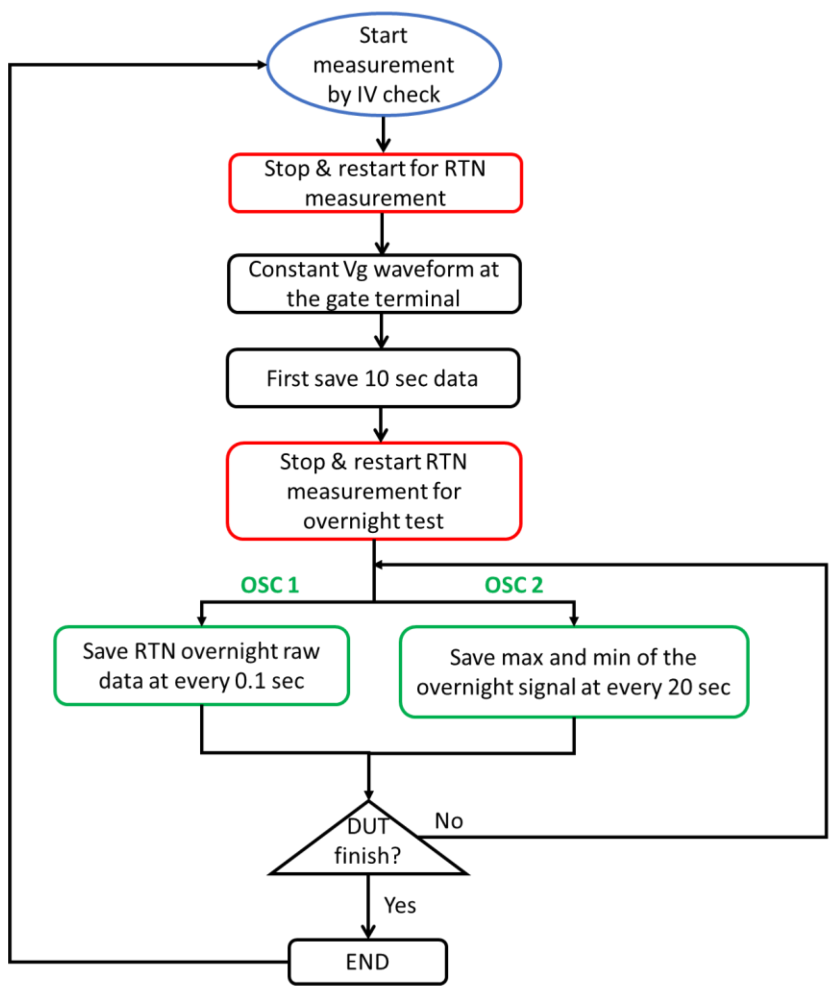


Figure 3.5.3 The flow chart for the overall overnight test.

An example of an RTN overnight measurement result is shown in Figure 3.5.4 where the black line is the 10 sec data with upper and lower envelope (red lines) and the blue line is the overnight raw data with upper and lower envelope (green lines) data recorded every 20 sec from the experimental measurement by using another oscilloscope (OSC 2). As already mentioned above, the oscilloscopes automatically detect the max and min values of the Id current on the screen. These max and min values can be saved directly from the OSC screen to the OSC hard disk by generating one file. This was the limitation that arose while saving these values from OSC as these data couldn't be append in the same generated file while recording, and each time one new file has to be created to save a single data point. For instance, 1M data point will be saved in 1M files which reaches the hardware limitation of the work for overnight measurement. Since the purpose is only to depict the envelope and most of the time the envelope remains the same (refer to section 4.3.2), here, with some trial and error process for the data analysis, handling software program and plot convenience, the concept of downscaling of data had been taken into account and each file is saved every 20 sec. However, as seen in the Figure 3.5.4 from the 10 sec raw data, there are 10 million data points. In the case of envelope extraction, there is no need for 10 million points to depict the shape of the RTN distribution as most of the time the upper and lower envelope remain steady, as seen from the Figure 3.5.4. Hence, the measurement is carried out under 1 MS/sec but the data storage rate is made different to avoid the difficulty of handling huge amounts of data, as already described earlier. However, both the upper and lower Env measured by the two oscilloscopes join together smoothly as shown.

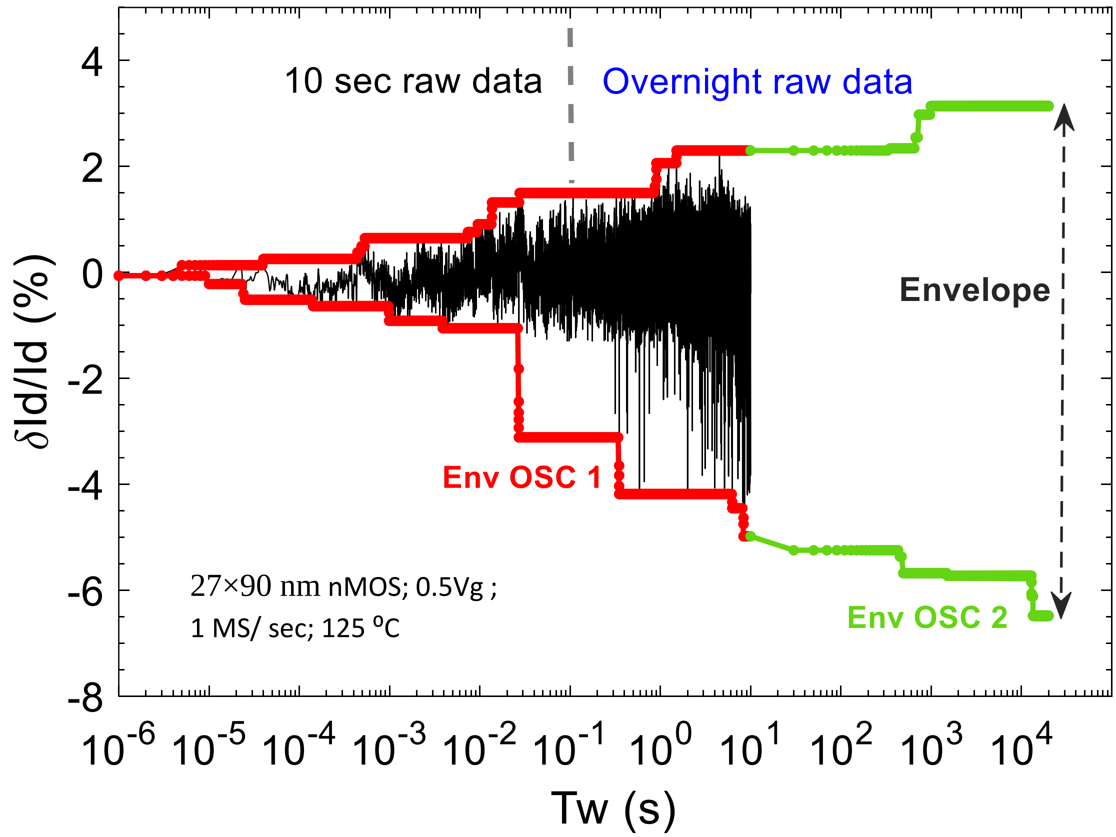


Figure 3.5.4 An example for RTN overnight measurement.

Although from the above figure, it can deduce that there are both acceptor and donor traps present, as the fluctuation can be seen in both positive and negative directions unfortunately this work does not have enough results needed to convince and support this as discussed in section 1.3.3.2.

Chapter 4 Statistical distributions of time constants

4.1 Introduction

As discussed in Chapter 1 , the downscaling of transistor size continues, and therefore the impact of random telegraph noise (RTN) is becoming an increasingly important consideration [4], [67], [96]-[98], for three main reasons. First, a single trapped charge has a larger impact on smaller devices. Second, RTN-induced malfunctions in a system are mainly caused by the devices in the tail of its amplitude statistical distribution. More transistors per chip increase the number of devices in the tail. Third, low power operation requires smaller overdrive voltage, (Vdd-Vth), so there is less room to tolerate the RTN-induced jitter of threshold voltage, ΔVth.

To take RTN into account, when optimising circuit design, substantial efforts have been made to model RTN [99]-[104]. For dynamic Monte Carlo modelling, one needs the statistical distributions of the number of traps per device, the amplitude of RTN per trap, and the capture/emission time (CET) of traps [67], [104], [105]. Earlier works [102], [70] have focused their attentions on the amplitude distributions but CET distributions have rarely been reported, based on test data [4], [20], [88], [106], [107] This is because it is difficult to obtain a sufficient amount of experimental CET data to establish a convincing statistical distribution.

The difficulties arise when CET is measured directly from the two discrete states of drain current, it requires a device having one trap only within the test time window [88]. This limits the number of CETs available. The Hidden Markov Model (HMM) [20], [108] has been used to extract trap properties. To analyse the RTN of multiple traps, Factorial HMM (FHMM) is proposed, where the measured signal is assumed to be a superposition of a number of independent two-level RTNs, with each of them from one trap and modelled by a Markov chain [77], [109]. Although this raises the number of traps analysable from one device, it becomes increasingly difficult to apply as the number of traps in a device increases with the time window. Although it is generally believed that there is no clear upper limit for CETs [4], [67], [105], [110] the time window used in earlier works is often limited, e.g. 10 sec or less [20], [88] (mentioned in section 3.2), partially to control the number of active traps in one device and partially for test convenience. RTN has been measured for longer time windows [111]-[114], but the statistical CET distributions were not established based on these test data.

4.2 Motivation for the investigation of statistical CET distribution

So far, most published RTN analysis is based on limited data from which two cumulative distribution functions (CDF) have been proposed for CET: Log-uniform [4], [67] and Log-normal [106], [107]. A Log-uniform distribution means that CET is statistically uniformly distributed against logarithmic time. With the help of a Monte Carlo simulation, CET is assumed to be either Log-normal or Log-uniform distributed as shown in Figure 4.2.1. 2,000 traps are then Poisson distributed into 400 devices. The CETs are randomly generated to create the simulated devices and verify the distributions. The following equations were used in MATLAB simulation to generate CETs:

$$CETs = \begin{cases} lognrnd(t_{start}, t_{end}, [1, N])\,; & for\ Log\text{-}normal; \\ 10^{(t_{start}+(t_{end}-t_{start})*rand(1,N))}; & for\ Log\text{-}uniform. \end{cases} \tag{16}$$

where t_{start} and t_{end} are the start and end points of the time window, respectively and N is the number of traps.

As demonstrated in Figure 4.2.1, the two distributions are very different, especially if they are used to predict long term RTN outside of the time window. Log-uniform CDF predicts that the number of active traps increases linearly against logarithmic time without saturation, while the Log-normal CDF predicts that there are fewer traps with long CETs and the CDF approaches saturation. As a result, the long-term RTN modelling cannot be trusted unless one has a trustable CET distribution.

The motivation of this work is to investigate the statistical distribution of the capture and emission times of the traps responsible for RTN. To do so, the objectives of this work are three-fold: to obtain the long-term RTN data experimentally and, based on them, to assess if any of these two and other distributions of CETs are correct; to propose a methodology for characterising the CET distribution; and to address the issue how accurately a distribution can make long term RTN predictions. As the practical time window for statistical tests is ∼1day, it is of importance to assess how accurately this data can be used to predict the RTN for years ahead.

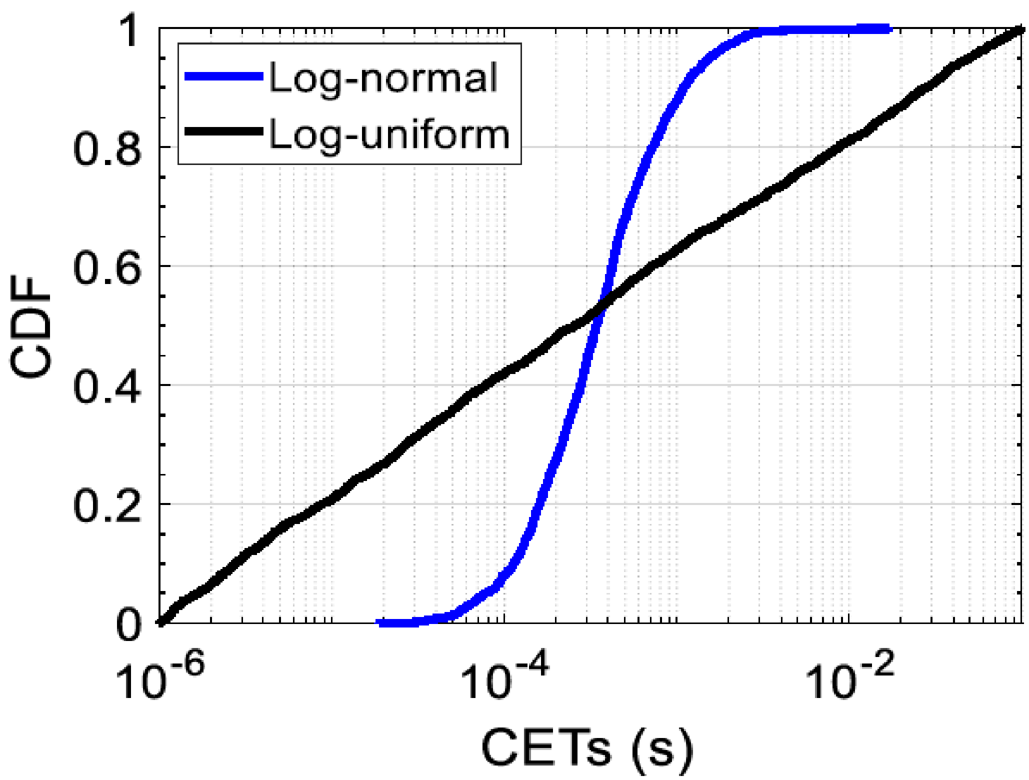


Figure 4.2.1 A comparison of the cumulative distribution functions (CDF) proposed for CETs: Log-normal versus Log-uniform.

4.3 Methodology and Measurement

Earlier works used two approaches to obtain the statistical capture/emission time distribution. These are: extracting CET directly [20], [88], [106], [107] or inferring the CET distribution from indirect measurements [4], [105]. As mentioned earlier, the difficulties in extracting CET directly often led to inadequate data to establish CET distribution unambiguously [4], [88]. Based on the measured CETs, some researchers proposed Log-normal CET distribution [106], [107].

The 1/f noise spectrum was used to infer the CET distribution [4]. It has been shown theoretically that a Log-uniform CET distribution will produce the commonly observed linear relation between power spectrum density and 1/f [4]. There are, however, two concerns with this inference. Kirton and Uren [4] showed that 1/f spectrum is insensitive to CET distribution and different CET distributions can produce a similar spectrum. Another concern is that the 1/f spectrum typically has a low frequency limit of ~1 Hz, corresponding to an up-limit in the time domain of ~1 sec. There is a lack of data for the long-term distribution, therefore.

A Log-uniform CET distribution is also inferred from the negative bias temperature instability (NBTI) tests [105]. It has been shown that the ΔVth grows linearly against logarithmic time within the first ~1 sec [105]. Unfortunately, the charging kinetics starts deviating from this linear relationship [105], as new traps are generated [115], [116].

The approaches adopted by earlier works is not followed in this research as those works did not give long term data for establishing CET distribution. Instead, an overnight RTN test had been carried out. Figure 4.3.1 shows the result of an overnight noise measurement. Although the noise amplitude may appear insensitive to time, when plotted in linear scale in Figure 4.3.1 (a), the plot against logarithmic time in Figure 4.3.1 (b) shows that noise amplitude clearly increases over a longer time. It is difficult to extract CETs from such data unambiguously. Instead, the increase of noise amplitude with time in Figure 4.3.1 (b) can be used to uncover the underlying CET distribution and the methodology for this is described in the next section.

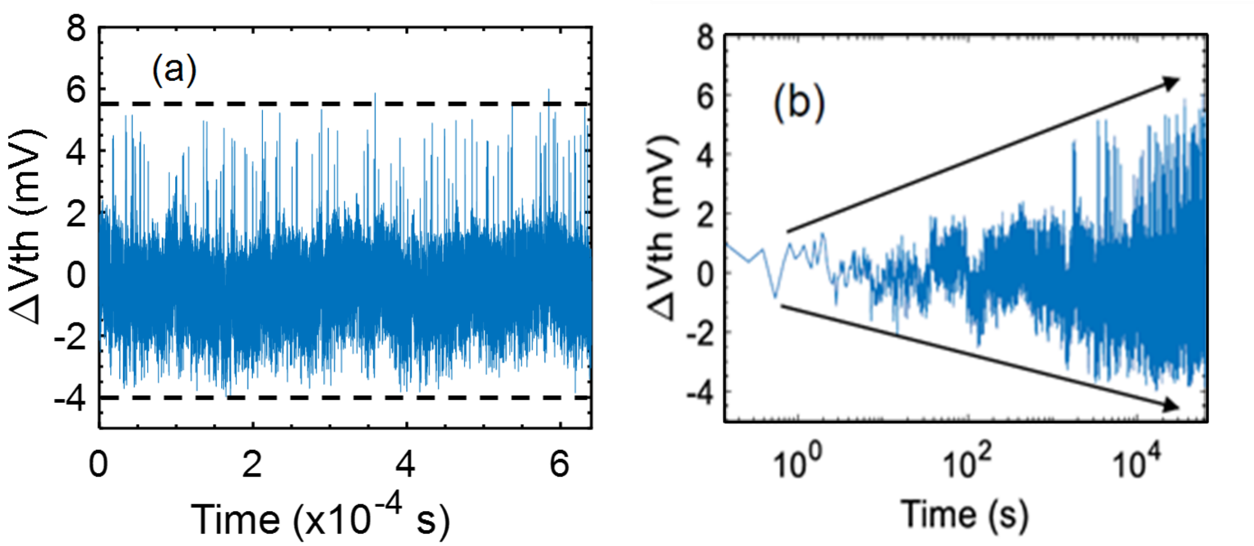


Figure 4.3.1 A typical overnight RTN measurement plotted against time linearly (a) and logarithmically (b). Although RTN amplitude appears constant against time in linear scale, it increases with time in log-scale in almost all measured data.

4.3.1 Methodology to extract statistical CET distribution

For a time window of tw, traps with CETs less than or close to tw are covered by the measurement. An increase in tw will bring slower traps into the measurements, leading to a higher cumulative RTN amplitude, as shown in Figure 4.3.1 (b). The build-up of RTN amplitude with time can be used to uncover the cumulative distribution function of CETs, therefore.

To illustrate this methodology, a case study is given in Figure 4.3.2. Figure 4.3.2 (a) shows the combined simulation results of 5 traps with their amplitude, capture and emission times listed in Table 2. The envelope of the complex multi-level RTN, Env, is extracted by,

$$Env(ti) = \begin{cases} \Delta Vth(ti), & if\ \Delta Vth(ti) > Env(ti-1), \\ Env(ti-1), & if\ \Delta Vth(ti) \leq Env(ti-1). \end{cases} \tag{17}$$

The RTN of each trap is given in Figure 4.3.2 (b-f). When the fastest trap makes a capture, it causes the first step-up of the envelope in Figure 4.3.2a, as marked out by '(1)' in Figure 4.3.2 (a) and (b). As the amplitude of this trap is fixed, the envelope remains the same when this trap goes through subsequent RTN events. When the second fastest trap becomes active, it causes the second step-up of the envelope, as marked out by '(2)' in Figure 4.3.2 (a) and (c). As time increases further, slower traps progressively become active, resulting in more step-ups in the envelope, as marked out by the corresponding numbers in Figure 4.3.2 (a) and (d-f). The evolution of the envelope with time in Figure 4.3.2 (a) originates from a distribution of time constants of the underlying traps, therefore.

For the simulation, the available model of R. Gao proposed in [158] was used by modifying. For each hypothetical devices, the number of traps, n, single trap impact, δV_{TH} , energy level, E_T, barrier height, ΔEB generated by using the statistical distributions in equation 18 :

$$\delta VTH = \eta^{-1} \exp\left(-\frac{\delta VTH}{\eta}\right) \tag{18}$$

Where, the average single trap impact, η is calculated out by using the proposed methodology in [98] which suggested that the corresponding η can be calculated from the mean and average sigma (σ) of the envelope at a specific time. For example, in this simulation work, mean and average sigma (standard deviation) is extracted from 400 test devices at 0.1 sec to get η by using the following equation 19 :

$$\eta = \frac{\sigma^2}{2\mu} \tag{19}$$

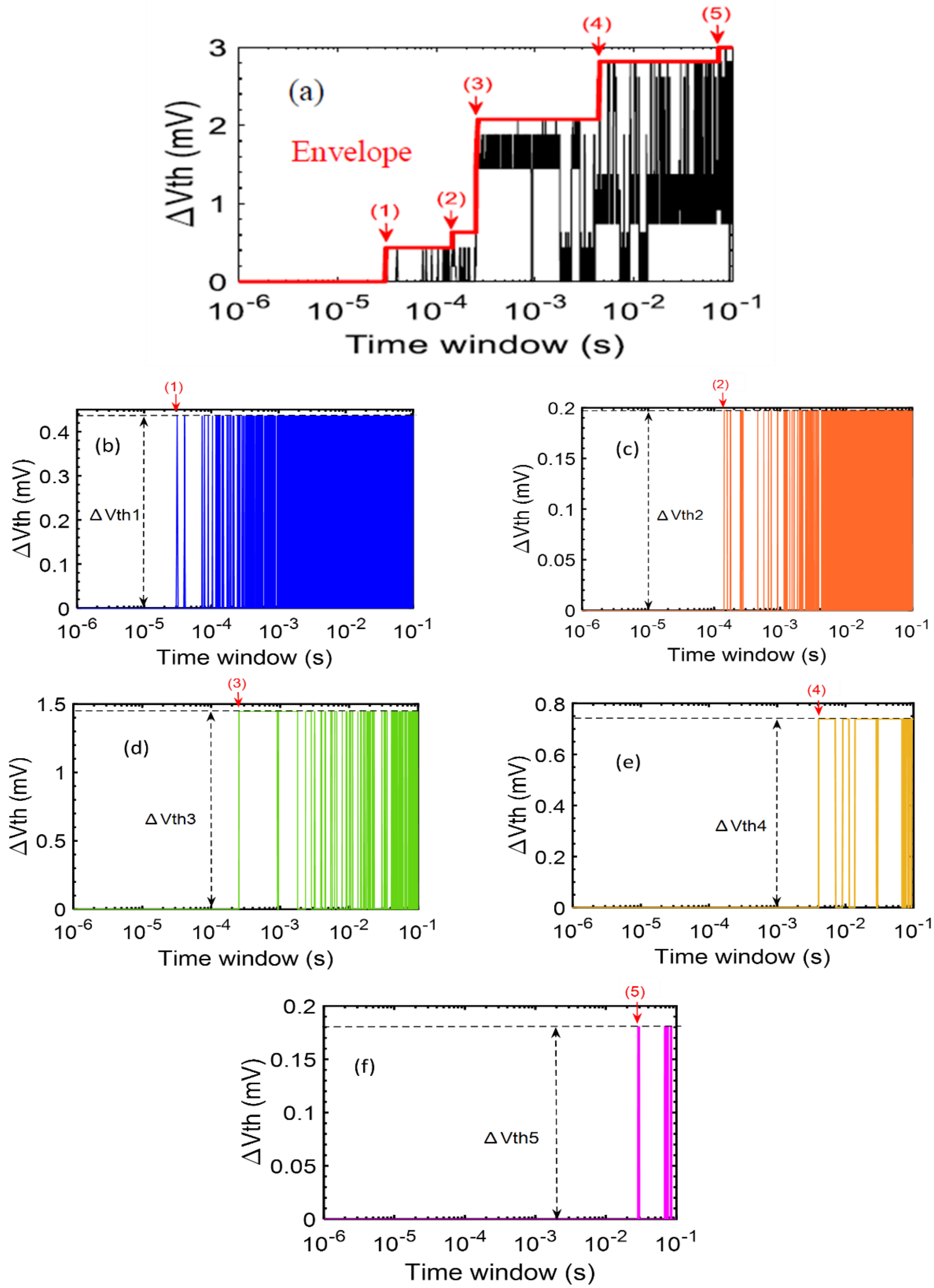


Figure 4.3.2 A simulation result of a device with 5 traps with their properties in Table 2. (a) shows the combined multi-level RTNs and the extraction of envelope. The RTN of each trap is shown in (b)-(f), respectively. The red arrows mark the first contribution of each trap to the envelope.

Table 2 Properties of traps used for the simulation in Figure 4.3.2.

Trapi	Capture Time (sec)	Emission Time (sec)	ΔVthi (mV)
Trap1	1.44×10^{-6}	1.94×10^{-7}	0.435
Trap2	5.84×10^{-5}	8.38×10^{-6}	0.198
Trap3	1.20×10^{-3}	2.68×10^{-4}	1.446
Trap4	2.60×10^{-3}	2.69×10^{-2}	0.739
Trap5	1.52×10^{-2}	3.90×10^{-3}	0.181

To support this methodology, dynamic Monte Carlo simulations were carried out. As mentioned earlier in the section, CET is assumed as either Log-normal or Log-uniform distributed, as shown in Figure 4.2.1. The 2,000 traps are Poisson distributed into 400 devices and the simulated results were recorded for analysis. Each grey line represents the envelope (Env) of one device in Figure 4.3.3 (a) for log-uniform and in Figure 4.3.3 (b) for Log-normal distributions. The black lines are the average results of 400 hypothetical devices. Although the envelope of individual device increases in steps, their average rises smoothly with time. A comparison with the CDF of CETs in Figure 4.2.1 clearly shows that the average Env correctly uncovers the underlying cumulative distribution of CETs. One can use the experimental Env of RTN to extract the CET distribution, therefore.

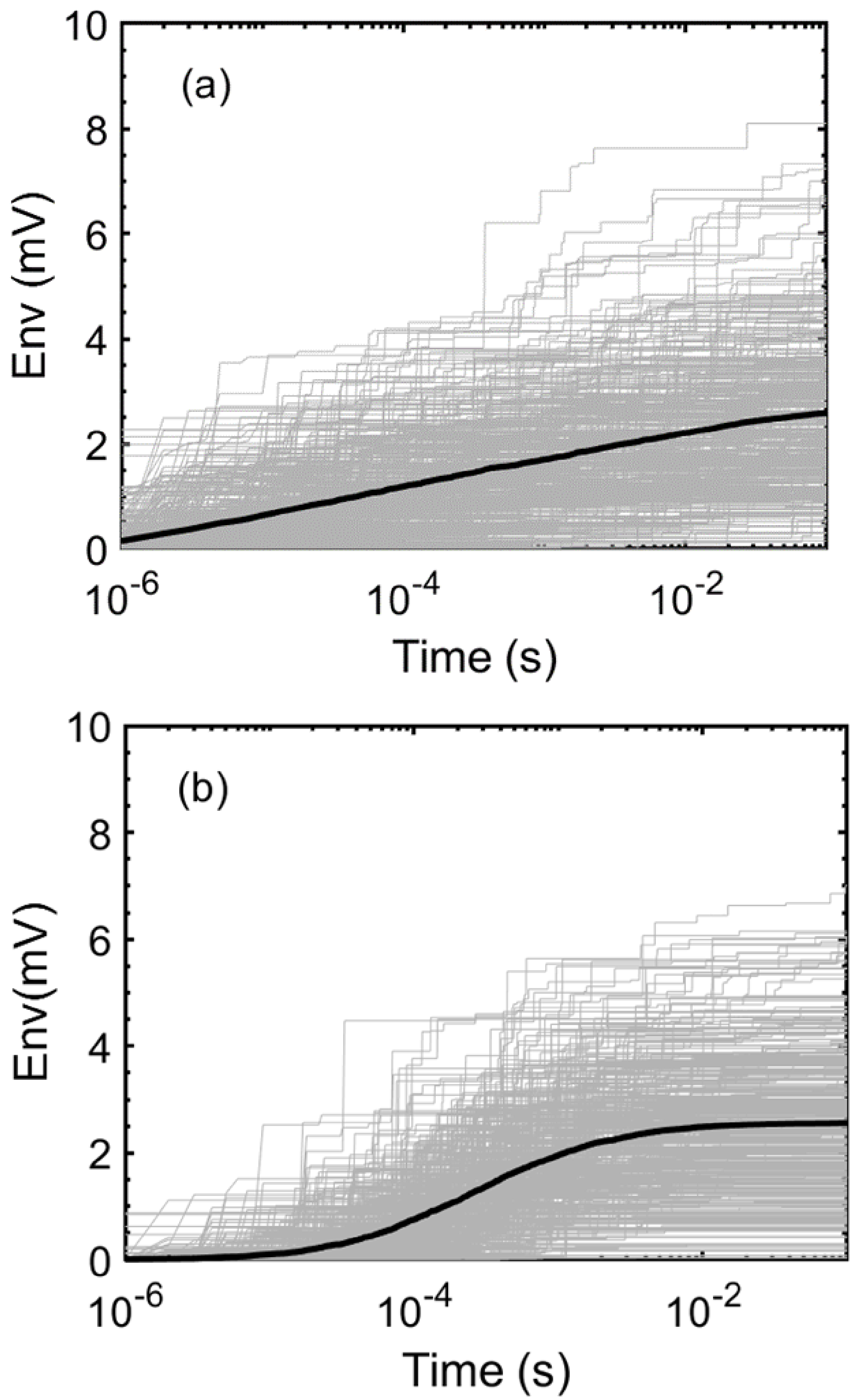


Figure 4.3.3 Simulation results for 400 devices with Log-uniform CDF (a) and Log-normal CDF (b). Each grey line represents one device. The black line is the average.

4.3.2 Devices and Measurement

nMOSFETs with a channel length and width of 27×90 nm were used in this work. The high-k/SiON stack has an equivalent oxide thickness of 1.2 nm and the gate is metal as discussed in section 2.3.

The test pattern used is as shown in Figure 3.5.1 from Chapter 3 . Tests start by measuring a pulse Id~Vg with a pulse edge time of 3 µs. The drain voltage bias used is Vd = 0.1V. The Vg waveform stepped from zero to 0.5 V and Id is monitored against time. The gate voltage Vg = 0.5V is used as the average threshold voltage of the nMOSFETs used in the work is 0.45 V and so the Vg is chosen to be Vth+0.05 V. As mentioned in section 3.5, the requirement of low power is driving Vdd towards Vth and the near threshold computing acutely suffers from RTN [95]. The temperatures used are 28 ºC and 125 ºC.

It has been reported that both as-grown traps and traps generated by stresses can induce RTN [55], [117], [118]. The generation process, however, follows power law [119], which is different from the Log-uniform [4], [67] or Log-normal distributions [106], [107] of time constants for charging-discharging as-grown traps. This work focuses on investigating the distribution of time constants for charging-discharging as-grown traps and a low Vg = 0.5 V is chosen for the tests to minimise the interference from the trap generation process [89]. Moreover, metastable and anomalous RTNs have been reported [67], [97] and their effects have been included in the experimental data.

The Id fluctuation, ∆Id, is calculated from Id-Iref, where Iref was evaluated from the average Id between 1 and 10 µs. As Vg is close to Vth, ∆Vth can be evaluated from: ∆Id/gm [98], where gm is the transconductance and is obtained from the pulse Id~Vg. The system noise is below ±1 mV as shown in Figure 4.3.4.

The extraction of the envelope from experimental data is illustrated in Figure 4.3.4. As discussed in section 3.3.2, the sampling rate for this work used is 1 MS/sec, where `MS' is `Mega-Sample points'. Figure 3.3.3 (b) shows the sampling rate impact on Env. Slower sampling rate leads to lower Env, as it fails to capture the fast traps [90].

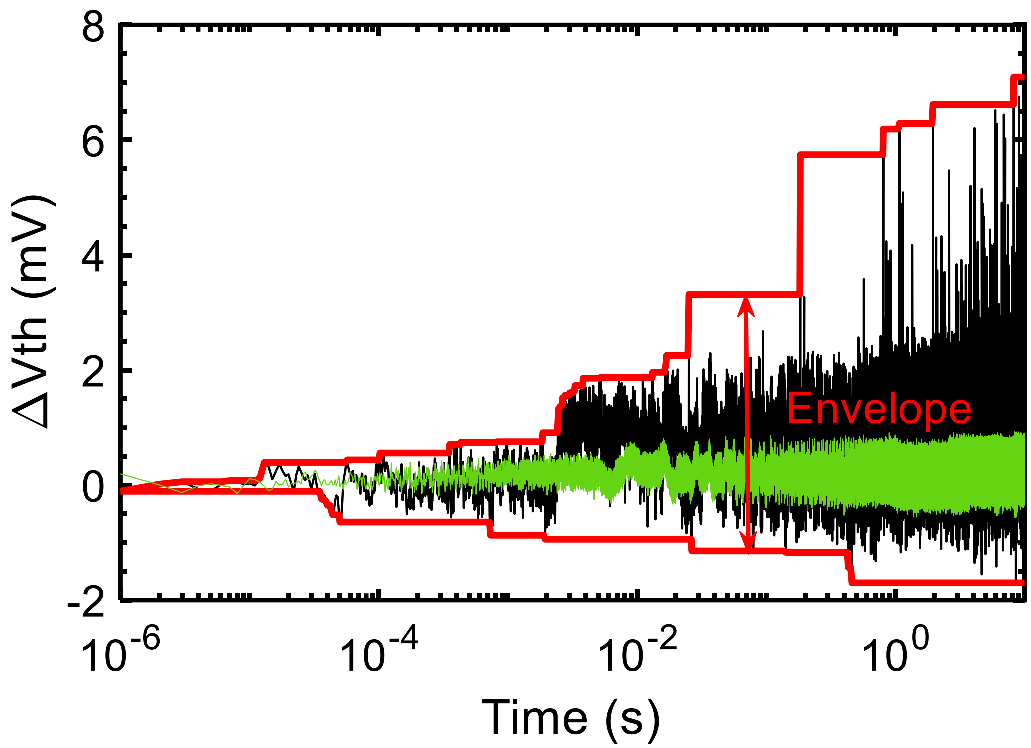


Figure 4.3.4 Extraction of RTN Envelope from experimental data (black lines). The green trace represents a device of limited step-like change in ΔVth.

With 1 MS/sec, the size of data set for one measurement is 10 MS for a time window of 10 sec. For a time window of 10^5 sec (~ day), the data size rises to 100 GS, which is beyond the memory depth of modern oscilloscopes. To overcome this difficulty, different rates for data sampling and data recording is used as details given in Chapter 3 . As shown in Figure 4.3.3, the number of steps in Env are limited and Env remains constant most of the time. Hence, Env can be recorded under a much slow rate, although it is measured at 1 MS/sec.

As mentioned, in this work two oscilloscopes were used to monitor Id. One of them has a time window of 10 sec and every data point is recorded. The other has a time window of 10^5 sec and monitors Env at 1 MS/sec, but the result is only recorded every 20 sec for the overnight test. The Env measured by this set-up is given in Figure 4.3.5 for 51 different devices. Each grey line represents one individual device and the red line is their average. The Env measured by the two oscilloscopes joins together smoothly.

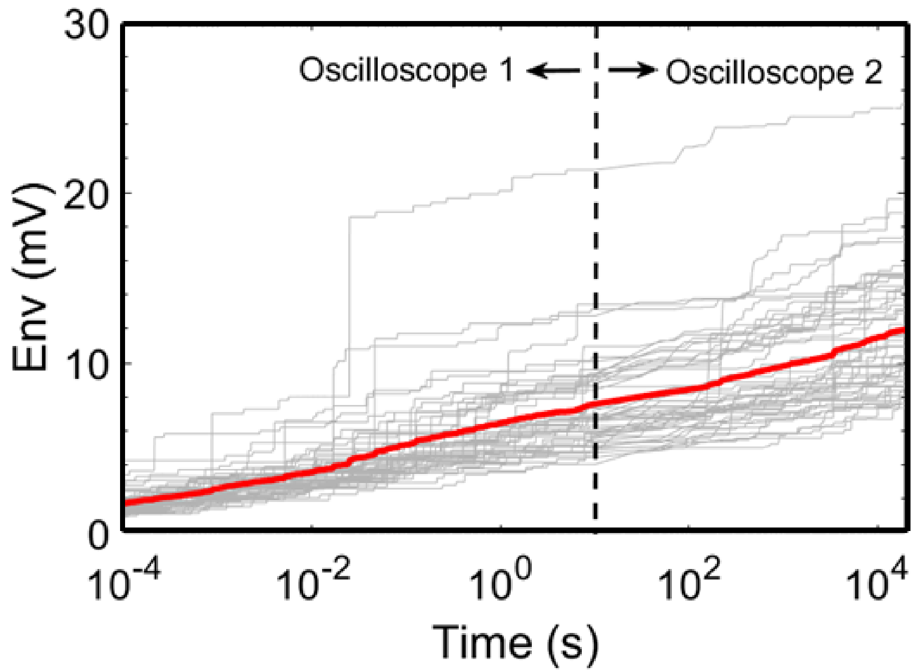


Figure 4.3.5 The overnight RTN envelopes measured by two oscilloscopes: The oscilloscope 1 covers up to 10 sec and the oscilloscope 2 covers longer time. Each grey line represents one device. The red line represents the average. The temperature is 125 ^{o}C.

The statistical tests require repeating the same test many times for different devices. For a time window of overnight, the test becomes costly in terms of test time and it is desirable to minimise the number of devices under test (DUTs). For a time window of 10 sec, DUTs up to 402 were used and the average Env at 10 sec is plotted against the number of DUTs in Figure 4.3.6. Initially, the average is sensitive to the number, but settles down within 2% when the number is over 50. Therefore, this work can use 50 DUTs to extract the average Env for the overnight tests.

It should be clarified that, in addition to RTN, the measurement can also include other sources contributing to the 1/f spectrum. By using the measured data to characterise RTN, this work effectively treated the other sources as additional RTN through a higher RTN amplitude. For nanoscale MOSFETs, RTN plays a dominant role. This can be seen from the step-like changes of the envelope in Figure 4.3.4 and Figure 4.3.5. Figure 4.3.4 also shows that, when the step-like changes are small, the total noise is much lower (the green trace).

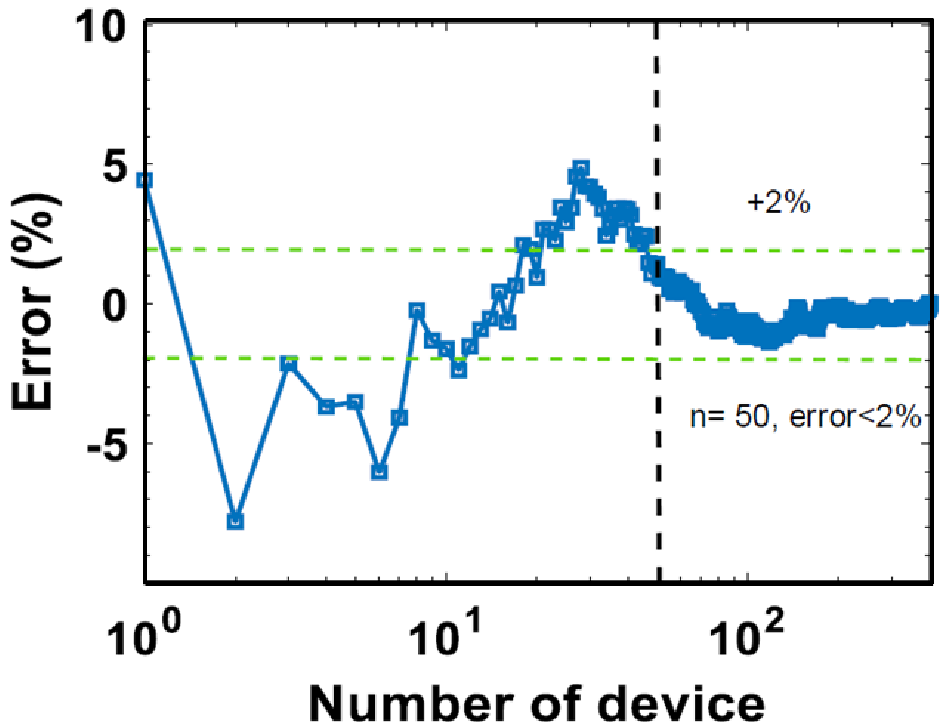


Figure 4.3.6 The Impact of the number of devices on the average envelope. When the number of devices is over 50, the error is within 2%.

4.4 Statistical Distribution of CETs

For the first time, the overnight RTN experimental results in Figure 4.3.5 are used to assess the statistical distribution of CETs. The non-saturation behaviour is widely observed for device ageing, which typically follows a power law [120]-[123]. To test if the Env also follows a power law, an attempted is taken to fit it with a power law. Figure 4.4.1 (a) shows that the agreement with power law is not good. Figure 4.4.1 (b) and (c) show that the experimental data fit reasonably well with Log-uniform and Log-normal distributions, respectively. This demonstrates that good fitting with experimental data is not a sufficient criterion for qualifying a model [89], [109], [120]. As the mission of modelling is to use the model to predict the device performance where experimental data are not available for model extraction, the predictive capability of these models will be tested next.

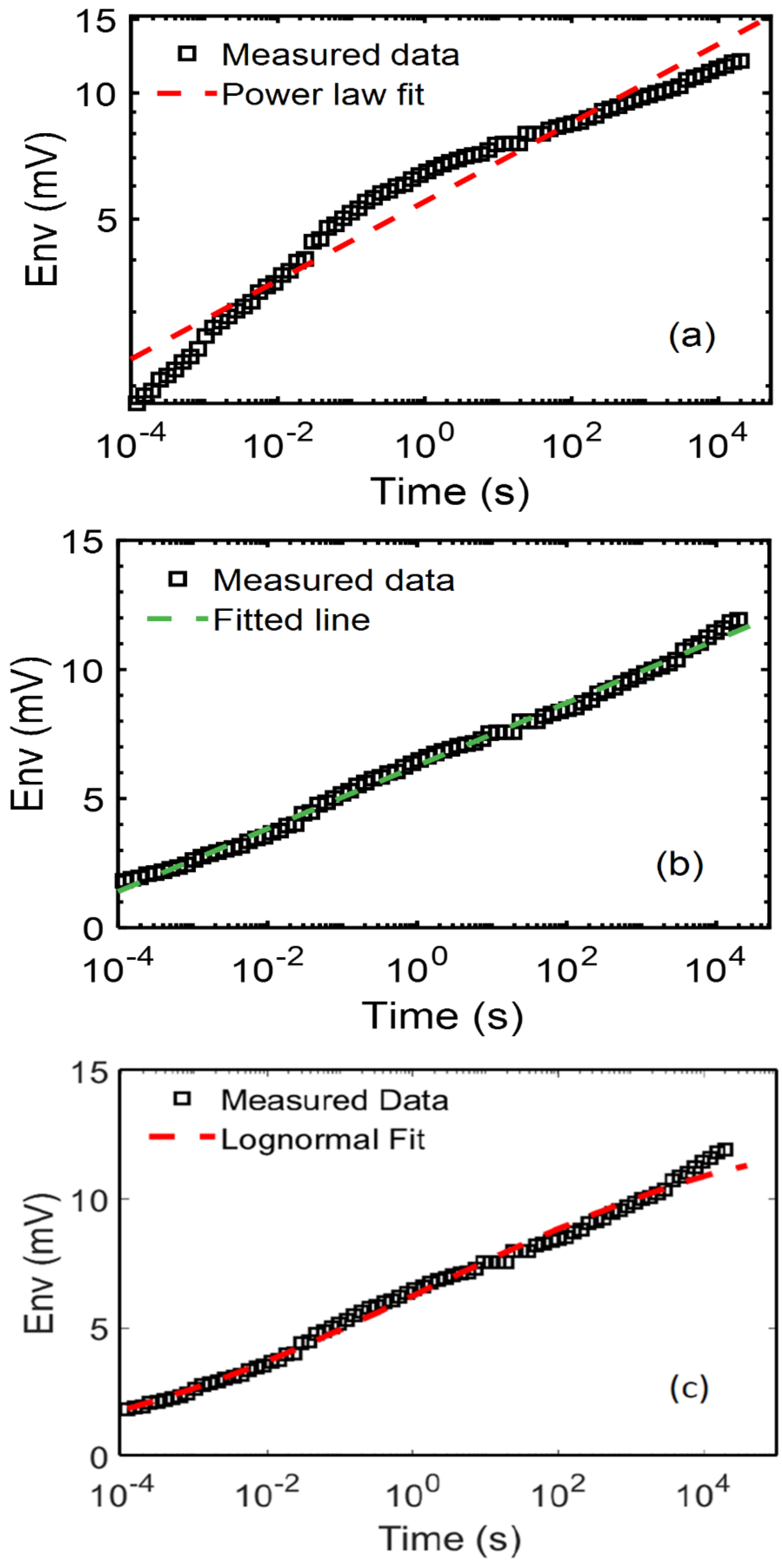


Figure 4.4.1 The evolution of envelopes with time. Symbols are experimental data and dashed lines are fitted with (a) power law, (b) Log-uniform, and (c) Log-normal.

4.5 Prediction of the long-term CETs

Although Figure 4.4.1 (b) and (c) show that the CETs within a time window of ~ day can be fitted reasonably by the Log-uniform and Log-normal distributions respectively, most electronic products require a lifetime of years, rather than days. To optimize a design, one needs to model the impact of RTN over the whole device lifetime. As it is impractical to carry out the repetitive statistical tests with a time window of years, one relies on the fact that the models extracted from the test of ~ day can be used to predict three orders of magnitude ahead to reach ~ years [89], [116], [120]. The question is how to verify this long-term prediction capability of a model.

As there is no test data of ~ years, it is impossible to have direct verification. What is available from this work is the test data up to 2×10^4 sec in Figure 4.3.5. Reducing it by three orders of magnitude gives a time window of ~ 10 sec. The model based on the data in a time window of 10 sec can be extracted and then it can be used to predict the RTN three orders of magnitude ahead to reach ~ 10^4 sec. As this work has the test data for ~ 10^4 sec, this prediction can be verified.

The solid black lines in Figure 4.5.1 (a-c) represent the model extracted from the data with a time window of 10 sec for power law, Log-uniform, and Log-normal distributions, respectively. The dashed lines are obtained by extrapolating the solid lines according to the extracted models. When compared with the experimental data that has not been used to fit the models (red symbols), the Log-uniform CDF in Figure 4.5.1 (b) is the clear winner. It predicts that Env reaches 18.5 mV at 10 years. The power law in Figure 4.5.1 (a) overestimates Env and gives a value of 47.5 mV at 10 years. On the other hand, the Log-normal CDF in Figure 4.5.1 (c) underestimates Env and gives a value of 12.8 mV at 10 years. The Log-normal CDF approaches saturation at a longer time, which was not observed in the test data. As a result, the experimental data supports the log-uniform distribution of CETs.

As the model extracted from the test data over five orders of magnitude of time between 10^{-4} and 10 sec can be used to predict three orders of magnitude ahead, it gives confidence that the model extracted over eight orders of magnitude from 10^{-4} to 2×10^4 sec can also be used to predict three orders of magnitude ahead, reaching ~ years.

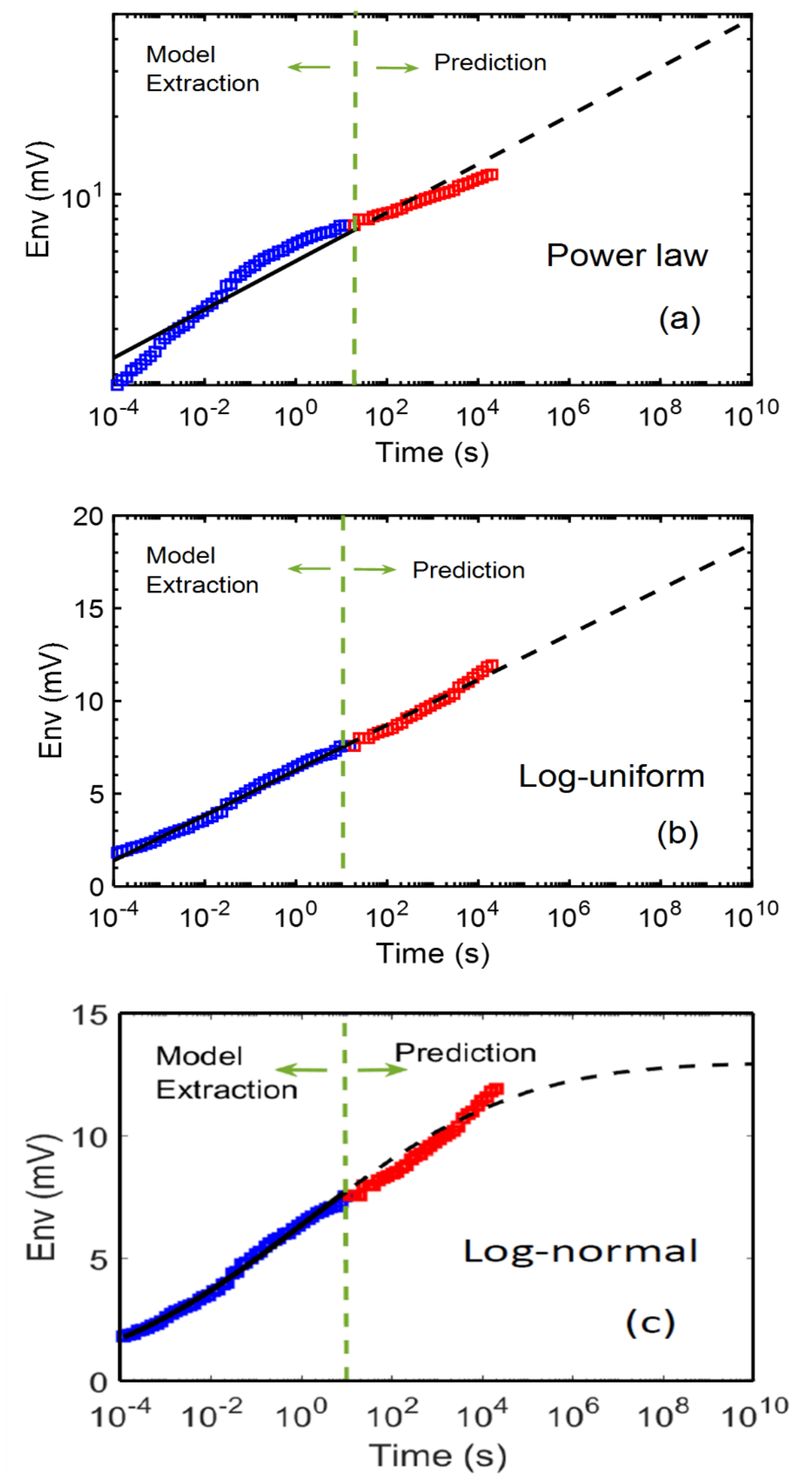


*Figure 4.5.1 Testing the predictive capability of the power law (a), Log-uniform (b), and Log-normal (c) CDFs. The experimental data up to 10 sec (**blue symbols**) were used to extract the CDFs (black solid lines). The obtained CDFs were then used to make prediction beyond 10 sec by extrapolation, as shown by the dashed black lines. The Log-uniform CDF has the best agreement between prediction and the experimental data (**red symbols**).*

4.6 Characterising Log-Uniform CDF

The Log-uniform CDF of CETs only have one parameter to be characterised: the number of traps per decade of time, Nt. This research proposes the following procedure to extract Nt:

- measure the RTN of multiple devices.
- extract the Env of each device, as shown in Figure 4.3.4.
- obtain the average envelope, as shown in Figure 4.3.5, and fit it with a straight line against logarithmic time, as shown in Figure 4.4.1 (b), and obtain the Slope.
- measure the amplitude of the RTN per trap and determine their average value, μ.
- Evaluate Nt by: Nt = Slope/μ.

For the process used in this work, the experimental results give Nt = 0.75/decade. Using this Nt and Log-uniform CDF for CETs and a Poisson distribution for traps per device, 400 hypothetical devices were generated for dynamic Monte Carlo simulation. Figure 4.6.1 shows that the simulated average Env agrees well with the measured one.

In principle, the Log-uniform distribution can be explained by two possible mechanisms: trapping-detrapping through elastic carrier tunnelling and inelastic multi-phonon trapping-detrapping.

It is well known that the carrier tunnelling probability decreases exponentially with the tunnelling distance [124], [125], resulting in an exponential increase of capture time with distance when moving from the dielectric/Si interface into dielectric. An assumption of a spatially uniform distribution of traps in gate dielectric can explain the Log-uniform CET distribution. Recent work [88], however, has reported that the CETs are not well correlated with the spatial position of traps. For the thin dielectric used in modern devices, carriers can readily tunnel through the whole dielectric in a short time [126], so that the depth into the dielectric typically does not control CETs.

For inelastic multi-phonon trapping-detrapping, carriers from the channel have to overcome an energy barrier, ΔE, to charge a trap. The capture time, τc, increases exponentially with ΔE [4], [88],

$$\tau_C = \tau_0 \exp\left(\frac{\Delta E}{kT}\right) \tag{20}$$

where τ_0 is a constant, k the Boltzmann constant, and T the temperature. A statistical uniform distribution of traps in ΔE will result in a Log-uniform distribution of CETs.

Furthermore, the advantages and disadvantages of the approach used in this work's, `envelope approach', are compared with the conventional method, for extracting the statistical distribution of CETs. Conventionally, a bottom-up method was used: the time constant of each trap is measured first and then used to establish statistical distributions [106], [107]. The advantage of this approach is that one knows the time constant of each trap and a lot can be learned about the property of individual traps from these earlier works [20], [88], [106], [107]. The disadvantage of this approach is that the number of traps and their time constants obtained through experiments is too limited to establish the statistical distribution convincingly, especially for slow traps. This research work could not do better than these earlier works if it followed the same bottom-up approach.

The envelope approach developed in this work can be considered as a top-down or integrated method: the results of multiple traps from multiple devices were combined and analysed together to extract the statistical distribution without knowing the precise time constant of each trap first. The advantage of this approach is that it allows for extraction of the statistical distribution of time constants, efficiently based on the long term RTN data, as shown in Figure 4.3.1 and Figure 4.4.1. The disadvantage of this method is that the precise time constant of each trap is not known and this precludes any quantitative comparison of simulation with test data for individual devices. As the precise time constant of each trap is not known, the time constant of each trap has to be statistically assigned according to the distribution for the simulation. Figure 4.6.1, however, shows that the simulation agrees well statistically with the test data.

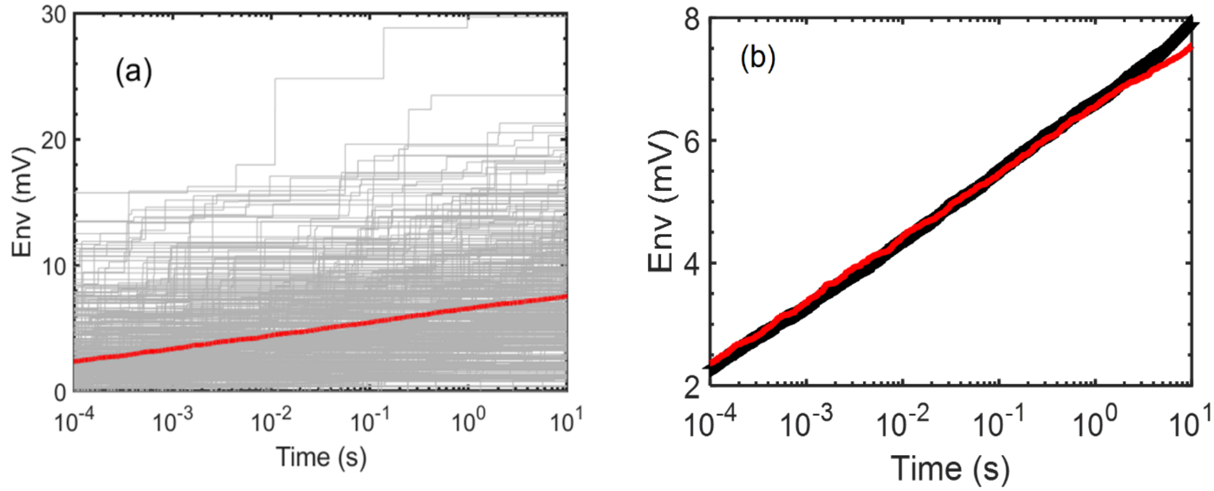


Figure 4.6.1 (a) Simulation results of 400 devices generated by the Log-uniform CDF extracted by the procedure given in section III.C. Each gray line represents one device and the red line is the average Env. (b) A comparison of the average envelope by simulation (red line) with the experimental average envelope (black symbols).

Finally, it is investigated if the Log-uniform CDF is applicable to RTN under different test conditions. As RTN is sensitive to temperature, overnight RTN was measured at 28ºC in Figure 4.6.2 (a), while the results in Figure 4.3.5 were measured at 125ºC. Figure 4.6.2 (b) shows that the Log-uniform CDF again fits the experimental data at 28ºC well.

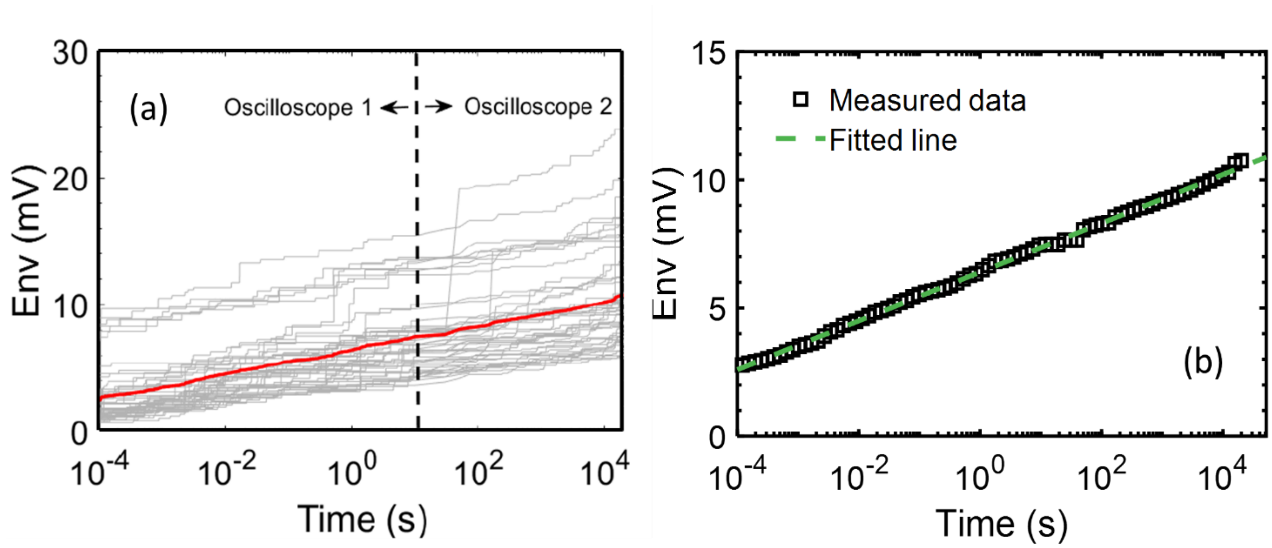


Figure 4.6.2 (a) The overnight RTN envelopes measured at 28ºC. Each grey line represents one device. The red line represents the average. (b) The symbols are the average experimental Env. The dashed line is fitted with the Log-uniform.

4.7 Discussion and Summary

In this chapter, the statistical distribution of the capture and emission times of traps responsible for RTN is investigated by developing a top-down methodology. The work started by using the dynamic Monte Carlo simulation to confirm that the average envelope of RTN, resultant from multiple devices and many traps, can uncover the underlying cumulative distribution of CETs. Overnight RTN tests were then carried out to extract the experimental envelopes for RTN. Based on this long-term RTN data, the CDFs proposed by earlier works for CETs were assessed. It is found that the power law, widely used for ageing, does not agree well with the test data and overestimates long term RTN. On the other hand, the Log-normal CDF underestimates long-term RTN. The overnight experimental data endorses the Log-uniform CDF for CETs. A methodology is proposed to extract the CDF of CETs efficiently. For the first time, the long-term prediction capability of the extracted Log-uniform CDF is verified, allowing assessing the RTN in years, based on the experimental data in days.

Chapter 5 Statistical distributions of RTN amplitude

5.1 Introduction

It is well known from the discussion so far that random telegraph noise (RTN) is a step-like fluctuation of drain current under constant gate and drain voltages. It has received much attention, as it adversely affects the operation of electronic circuits [4], [67], [70], [77], [96], [97], [106], [108]-[111], [127]-[130]. As MOSFETs become smaller, RTN becomes increasingly important, driven by an increased impact of a single charge on smaller devices and an increase in the number of devices in a system [4], [67], [70], [96], [97], [110], [127], [128]. A large number of devices in a system will contain more devices in the tail of statistical distributions as mentioned in section 4.1, which can cause errors. Moreover, low power is a key requirement for many Internet-of-Things edge units and this drives the operation voltage towards threshold voltage, Vth [95], [131], [132]. The minimisation of overdrive voltage, (Vg-Vth), in the future leaves little room to tolerate RTN induced jitter [67], [131], [132].

There have been many efforts to model RTN, both in the frequency domain [4], [65], [99], [133], and in the time domain [4], [67], [70], [96], [97], [110]. It is widely accepted that RTN originates from the trapping/detrapping of charge carriers from/to the conduction channel [4], [65], [67], [70], [77], [95]-[97], [99], [106], [108]-[111], [127]-[133]. The number of traps per device follows a Poisson distribution [67], [70], [97], [110]. To perform Monte Carlo simulation in the time domain, one needs the capture-emission times and RTN amplitude of traps [67], [104], [120], [132]. In this chapter, the statistical distribution of capture/emission time constants was studied in Chapter 4 [132] and this chapter will focus on RTN amplitude distribution.

RTN amplitude can be measured as a Vth shift, ΔVth, or a normalised drain current fluctuation, ΔId/Id. ΔVth is the accumulative effect of multiple traps on a device and here δVth is used to represent the RTN amplitude of one trap. δVth is stochastic and one feature of its cumulative distribution function (CDF) is a long tail, when compared with the Gaussian/Normal distribution, as shown in Figure 5.1.1 (a) [70], [96]. It has been proposed that this long tail originates from the uneven distribution of current [70], [96], [127] since the impact of a trapped charge in the oxide on the device depends on the local current density beneath it [98], [127]. As schematically illustrated in Figure 5.1.1 (b), the current near threshold voltage flows through a narrow percolation path. It is rare to have a trap located just above this percolation path and such a trap will cause a large δVth and result in the long distribution tail [70], [96], [98], [127].

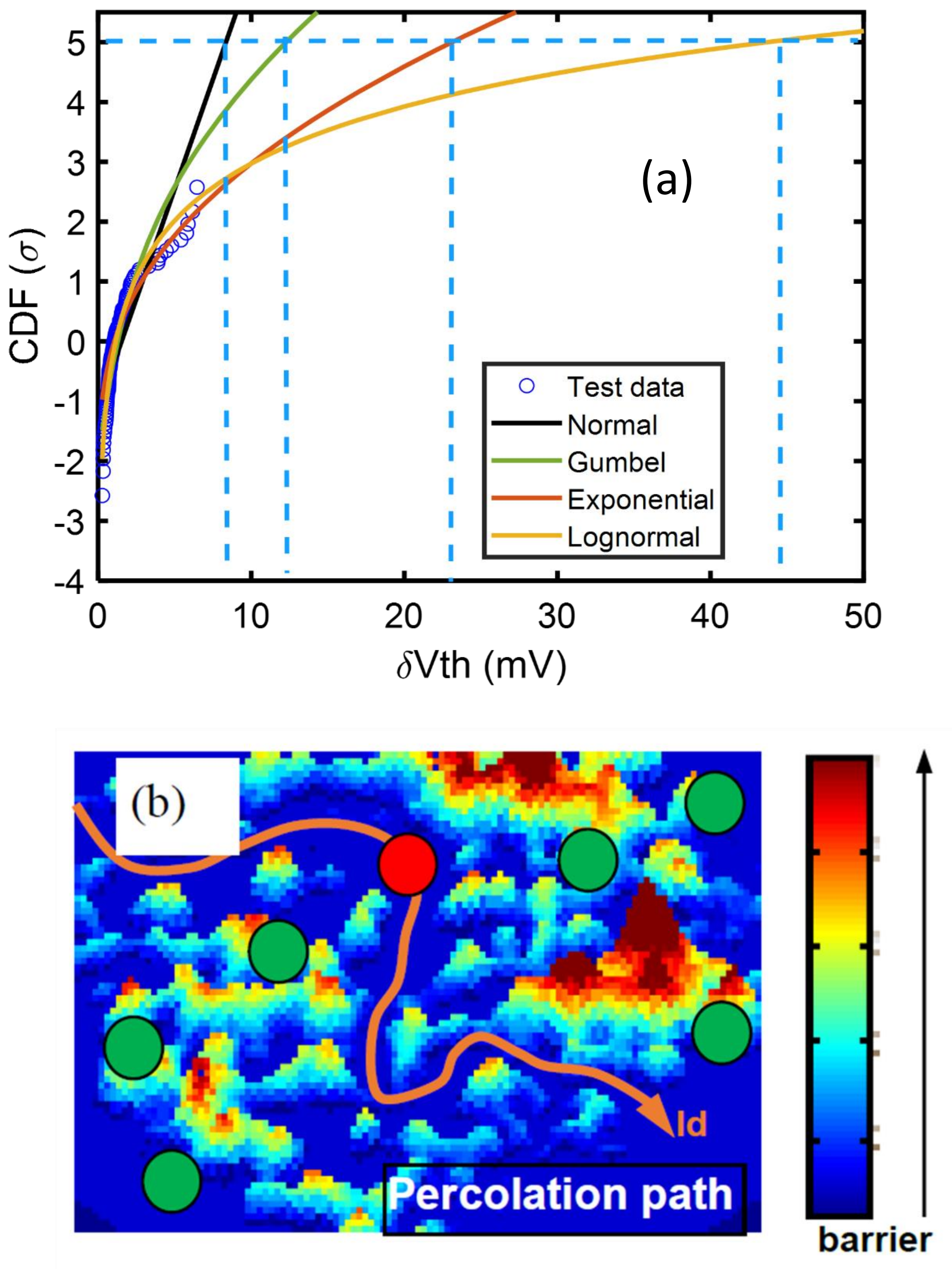


Figure 5.1.1 (a) A comparison of different cumulative distribution functions (CDF) of threshold voltage shift, δVth. Each 'o' represents δVth induced by one trap and there are 100 traps here. Although both the Exponential and Lognormal CDFs describe the test data well, they give very different results when their tails were used to make predictions, for example at 5σ, as shown by the dashed lines. (b) A schematic illustration of the impact of traps (circles) on current path near threshold condition. The red circle represents a trap just above the percolation path of current, which has a large δVth and is in the distribution tail.

Modelling the long tail in the CDF is a tall order and three statistical distributions have been proposed in the literature: Exponential [13], [21], [67], [68], [70], [97], [104], [134], [135], Lognormal [4], [6], [64], [67], [70], [127], [128], [136] and Gumbel [106], [129], [130]. The success of RTN modelling in terms of yield prediction for a system, such as SRAM, requires an accurate statistical distribution tail [6], [64], [67], [96], [127]. For a data set of 100 traps, Figure 5.1.1 (a) shows that both Exponential and Lognormal CDFs agree well with the test data, but they have substantially different tails. For example, at 5σ where σ is the standard deviation, the δVth predicted by Exponential and Lognormal CDFs is 23 mV and 44 mV respectively. This uncertainty calls the accuracy of RTN modelling into question.

Agreement has not been reached on which distribution should be used. Many earlier works [4], [13], [21], [68], [106], [110], [129], [134] only fitted their data with one statistical distribution. Different distributions were not compared and the reason for selecting a specific model is not given. For the works that compared the Exponential and Lognormal distributions [6], [64] it was reported that the Lognormal fitted the data better. There are, however, more fitting parameters in the Lognormal distribution than the Exponential distribution, so that it is not clear whether the improved fitting with the Lognormal originates from using extra fitting parameters.

5.2 Novelty and Motivation

The motivation of this work is to address the uncertainty in model selection for RTN amplitude in two ways. Firstly, this work attempts to find a statistical distribution that has lower error without using a higher number of fitting parameters. In addition to the three distributions mentioned above, thirteen other distributions are evaluated. Second, to propose a new criterion for selecting statistical models. It will be shown that if the data truly follows a specific CDF, the error per trap should decrease when increasing the trap number.

This work starts by examining the three distributions mentioned above in terms of their errors both over the whole distribution and in the distribution tail. The number of traps used in some early works [68], [106], [127] is ~100, leaving too few traps in the tail (e.g. >95%) to evaluate the error reliably. To enable the tail evaluation, 1,178 traps were used here.

The CDF parameters are extracted by the Maximum likelihood estimation (MLE). Earlier works suggest that the accuracy can be improved by using either bimodal Exponential [135] or bimodal

Lognormal distributions [70]. In this work, the impact of using bimodal distributions on the accuracy will also be examined.

An analysis of the distributions proposed by earlier works [4], [67], [70], [96], [97], [106], [110], [127]-[130] have not identified a clear winner. This leads to searching for new statistical distributions. Since there is little research on whether the RTN amplitude can be modelled better by other statistical distributions, apart from the three distributions mentioned above, a scoping study of different distributions was carried out. To emphasise the importance of the accuracy in the distribution tail for RTN modelling, the Z-score of corresponding CDF will be used to calculate errors, where $Z=(\delta Vth-\mu)/\sigma$ and μ is the average and σ the standard deviation. After comparing 16 distributions, it is found that Generalized Extreme Value (GEV) distribution [136] gives the lowest errors. GEV also meets the new criterion.

The last issue addressed in this work is the impact of trap number on the CDF accuracy in the distribution tail. The more traps used for extracting the distribution, the better the accuracy should be. In practice, however, the number of traps available is always limited. It is of importance to assess how reliably a CDF, extracted from a limited number of traps, can be used to predict the distribution tail at high sigma.

5.3 Methodology and Measurement

5.3.1 Devices and measurement

This work uses nMOSFETs fabricated by an industrial CMOS process, which have a metal gate and a high-k/SiON stack. The channel length and width are 27×90 nm, respectively. The equivalent oxide thickness is 1.2 nm.

The test conditions remain same as section 4.3.2 i.e. the test starts by applying a step voltage to the gate and drain. The Id is then monitored under a fixed Vg and Vd by an oscilloscope at a sampling rate of 1 Mpoint/sec [90], [132]. The gate and drain biases used are Vg = 0.5V and Vd = 0.1V for monitoring RTN as described in detail in Chapter 3 . Unless otherwise specified, tests were carried out under 125ºC.

Some typical results are represented in the next section in Figure 5.3.1 and Figure 5.3.2, where the current fluctuation is plotted as $\delta Id/Id=(Id-Iref)/Id$. The reference Id, Iref, was taken from the average of the first ten points of the measurement [132]. As Vg is close to Vth, δVth can be evaluated from -δId/gm, where gm is transconductance [98]. The gm is evaluated from a pulse (3 μs) Id-Vg, taken before the RTN test for each device [98]. There are more or less 2 to 3 traps per devices and in total the number of devices used in this work was more than 400 devices. An example of 1 and 2 traps distributions are showing in the following Figure 5.3.1 and Figure 5.3.2.

5.3.2 Method for extracting RTN amplitude

In the test of negative bias temperature instability (NBTI), the impact of one trap on Id is typically measured directly from its discharge induced step-change of Id [13], [137]. For RTN tests Hidden Markov Method [108] and Factorial HMM [77], [109] have been used, when both the RTN amplitude and time constants are needed. The time-lag plot has been often used to measure the RTN amplitude [138].

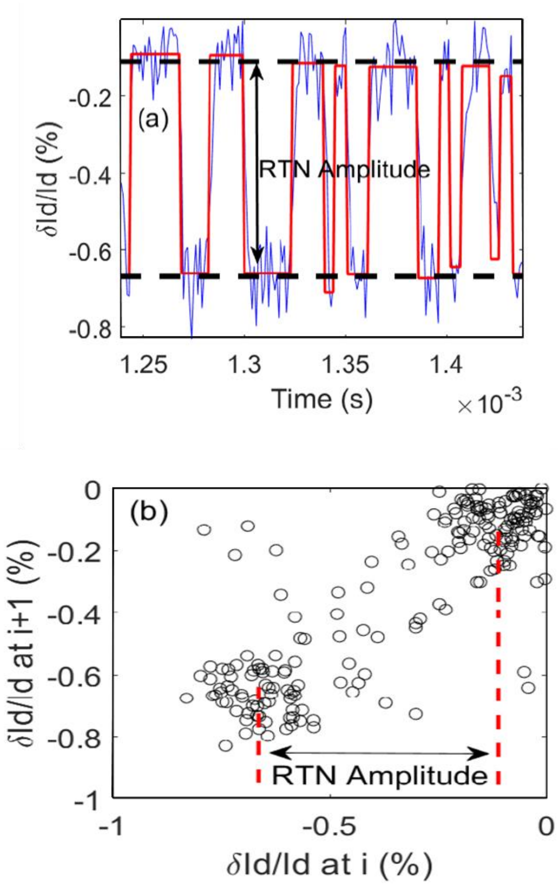


Figure 5.3.1 (a) Extraction of RTN amplitude directly from the two discrete levels of Id used in this work. (b) The same data in (a) was used to extract the RTN amplitude by the conventional time-lag method. The RTN amplitudes extracted by these two methods agree well when there is only one trap in a device.

Similar to the NBTI measurement [13], [137] the RTN amplitude directly from the step-changes in Id is measured in this work. As shown in Figure 5.3.1 (a), once a step-like change is observed, the Id for each discrete level is taken from the average of that level to minimize the effect of thermal noise. Moreover, unlike NBTI tests, where discharging one trap is often a one-off event [13], [137], here the advantage of the multiple charge-discharge events in RTN has been taken into account and the average of step-heights has been used to further improve measurement accuracy. The minimum detectable δId/Id is ~0.2%, corresponding to a δVth of ~0.2 mV.

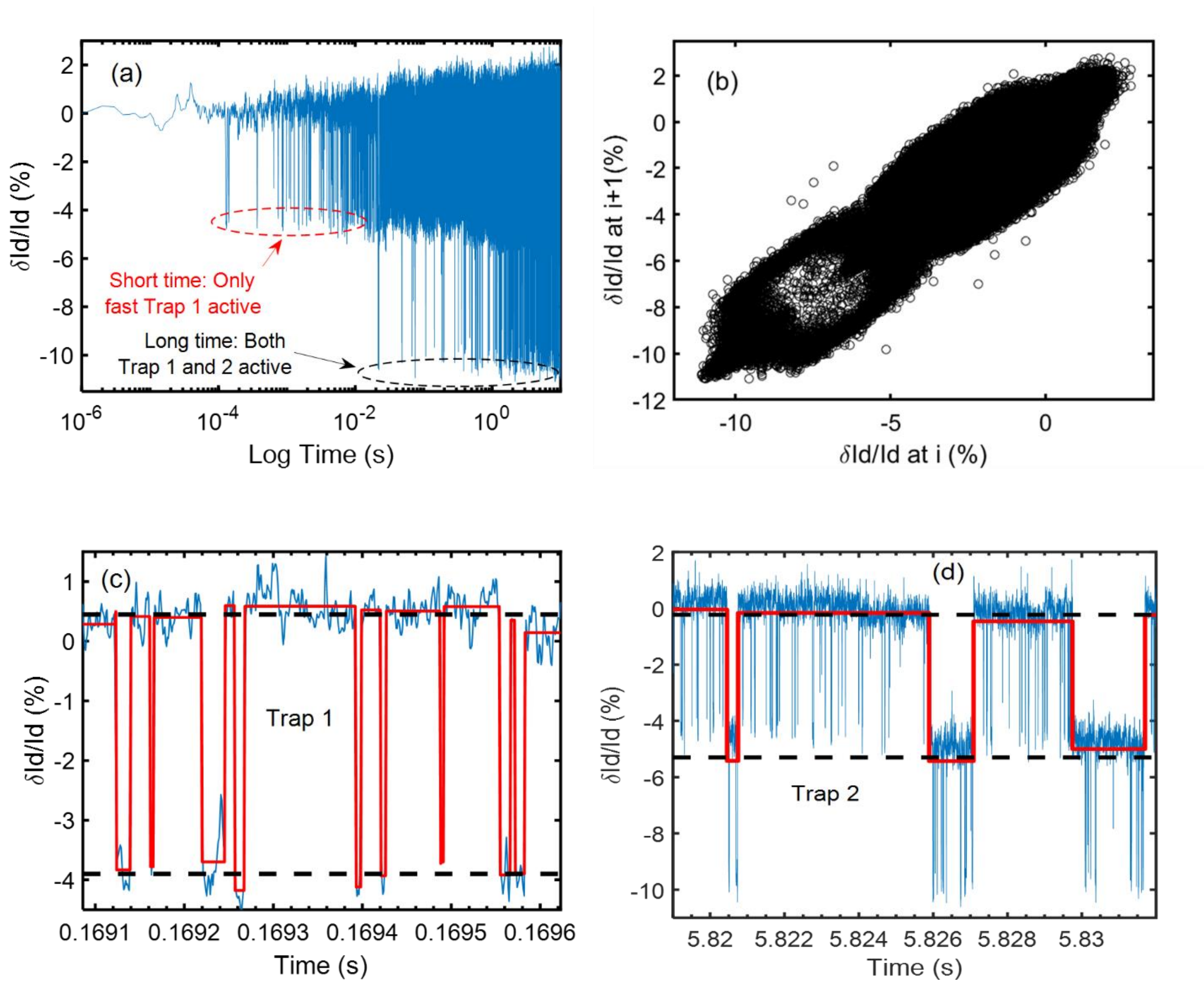


Figure 5.3.2 (a) An example of two active RTN traps in a device. (b) It is difficult to use the time-lag method to extract the RTN amplitudes for the data set in (a). (c) Extraction of the RTN amplitude of fast trap 1 by applying our method in the short time range. (d) Extraction of the RTN amplitude of slow trap 2 by applying our method in the long-time range.

Figure 5.3.1 (b) shows that the amplitude extracted by the new method agrees well with that of the time-lag method, when there is only one trap in a device. The time-lag method, however, uses data in the whole time window and is difficult to use when there are multiple traps, one example is given in Figure 5.3.2 (a) and Figure 5.3.2 (b). The advantage of the amplitude extraction method introduced in this work is that it can be applied to a selected time range where two-level RTN events are identified. For the same data set in Figure 5.3.2 (a), Figure 5.3.2 (c) shows that the amplitude of fast trap 1 can be measured in a short time window. For a longer time window in Figure 5.3.2 (d), the slow trap 2 becomes active and its amplitude can be extracted from the difference in the two discrete levels after averaging out the impact of the fast trap.

5.4 Problems with the proposed statistical distribution

For the RTN amplitude per trap, two popular statistical distributions used in earlier works are Lognormal [4], [67], [70], [127], [128], and Exponential [13], [67], [70], [97], [110], [135]. In addition, Gumbel distribution has been used to capture the long tail of RTN [106], [129], [130]. Their cumulative distribution functions (CDF) are summarised in Table 3. Table 3 also gives the formula for the Generalized Extreme Value (GEV) distribution, which will be investigated in Section 5.6. This section will focus on the three distributions used in earlier works: Lognormal, Exponential, and Gumbel.

Table 3 The cumulative distribution functions (CDFs).

CDF (Exponential)	CDF (Lognormal)
$1-e^{\left(-\frac{\delta Vth}{\eta}\right)}$	$\frac{1}{2}erfc\left(-\frac{\ln(\delta Vth)-\varepsilon}{\theta\sqrt{2}}\right)$
CDF (GEV)	**CDF (Gumbel)**
$e^{\left\{-\left[1+\xi\left(\frac{\delta Vth-\alpha}{\beta}\right)\right]^{-\frac{1}{\xi}}\right\}}$	$e^{\left(-e^{\frac{-(\delta Vth-\alpha)}{\beta}}\right)}$

Using the equation in Table 3, the parameters of different statistical distributions are extracted by the Maximum Likelihood Estimation (MLE) [135], [139]. MLE uses different weightings for different data to maximize the probability of test data set occurrence [139]. Based on the 1,178 measured traps, the estimated parameters are given in Table 4. The extracted CDFs are plotted together with test data in Figure 5.4.1.

Table 4 The estimated CDF parameters.

CDF (Lognormal)	CDF (Exponential)	CDF (Gumbel)	CDF (GEV)
$\epsilon = 0.143$	$\eta = 1.640$	$\alpha = 1.033$	$\xi = 0.540$
$\theta = 0.792$	-	$\beta = 0.843$	$\alpha = 0.831$
-	-	-	$\beta = 0.542$

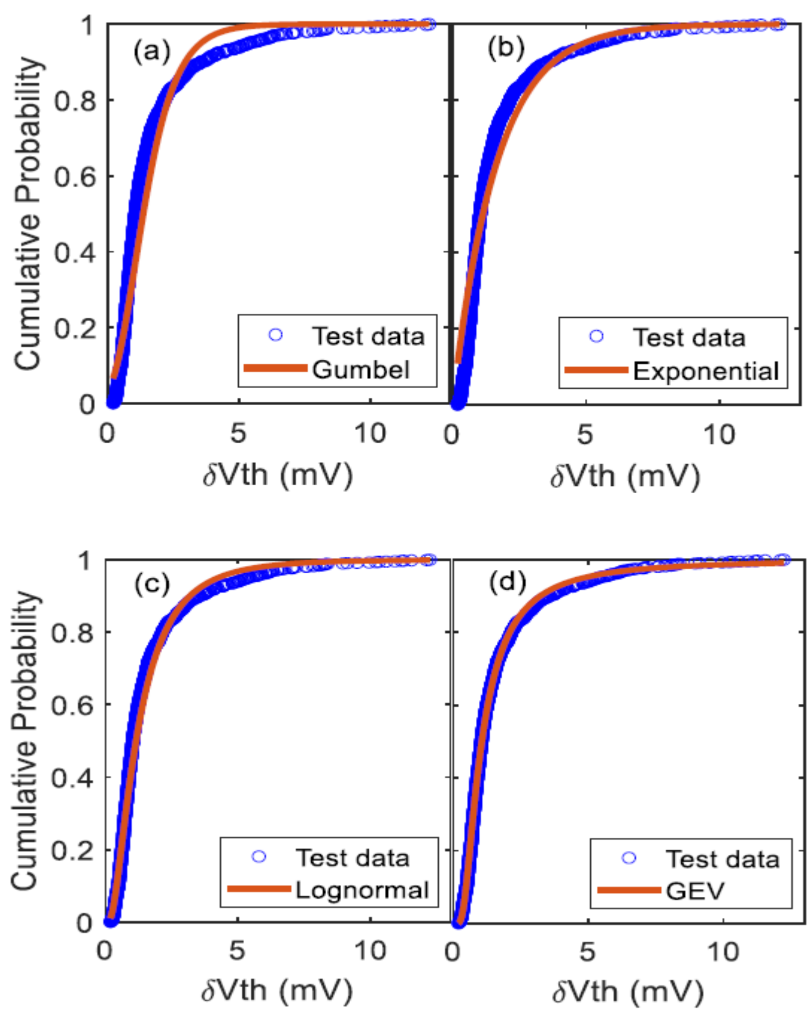


Figure 5.4.1 The CDFs (lines) extracted from 1,178 traps by the MLE method are compared with the test data (symbols) for: (a) Gumbel, (b) Exponential, (c) Lognormal, and (d) GEV, respectively.

Following the methods used in earlier works [6], [64], [67], [70], [140] the error between the extracted CDFs and the test data are used to compare different statistical distributions. Figure 5.4.2 (a) shows the sum-square-error (SSE) per trap. Gumbel and Exponential CDFs have similar errors, while Lognormal CDF has a lower error. For convenience, the result of GEV distribution is also given in Figure 5.4.2, which will be discussed in Section 5.6. If one uses the minimum error of the whole dataset as a criterion, the Lognormal distribution should be better than the Exponential and Gumbel distributions.

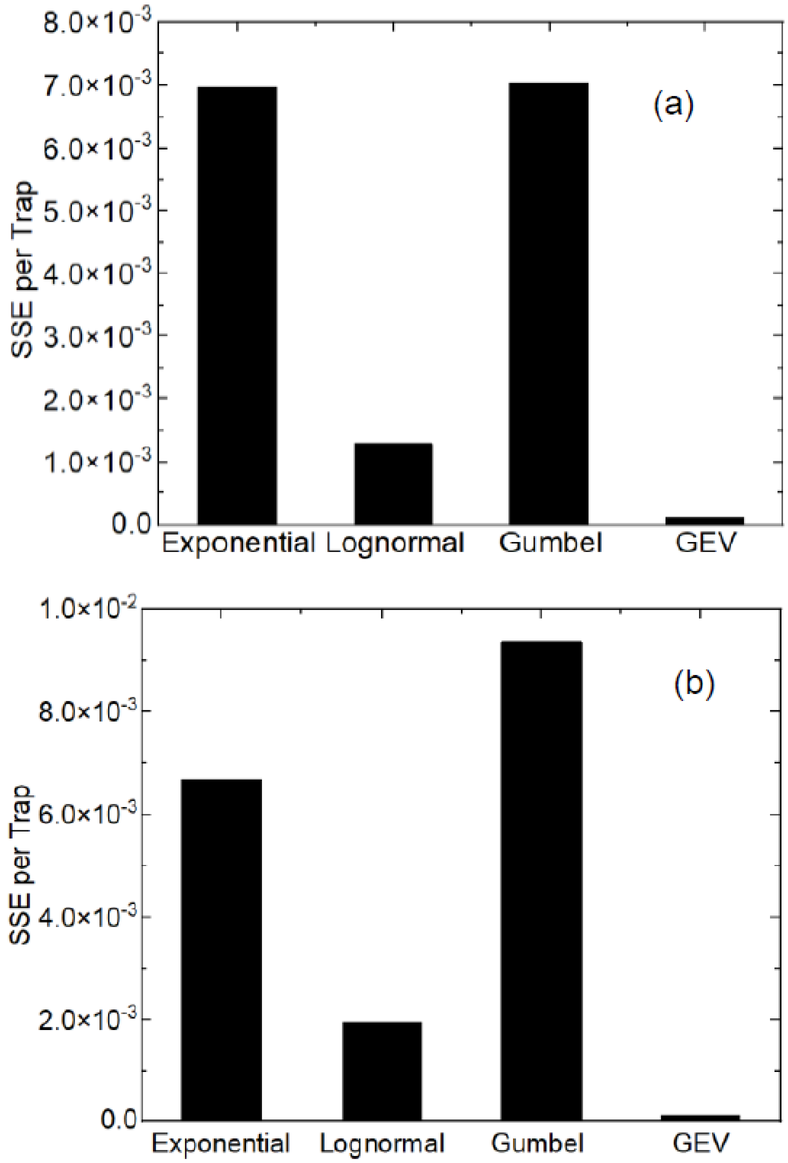


Figure 5.4.2 A comparison of the sum of square errors (SSE) per trap for CDFs extracted by Maximum likelihood estimation (MLE) at 125 °C (a) and 28 °C (b). The whole data set was used in the error evaluation. Lognormal has smaller error than Exponential for the whole data set.

The results in Figure 5.4.2 (a) were obtained at 125 ºC. To show that the observation is independent of test conditions, Figure 5.4.2 (b) gives the results at 28 ºC. RTN is generally sensitive to temperature and 814 traps were measured under 28 ºC. The error of Lognormal distribution again is lower than that of Exponential and Gumbel distributions.

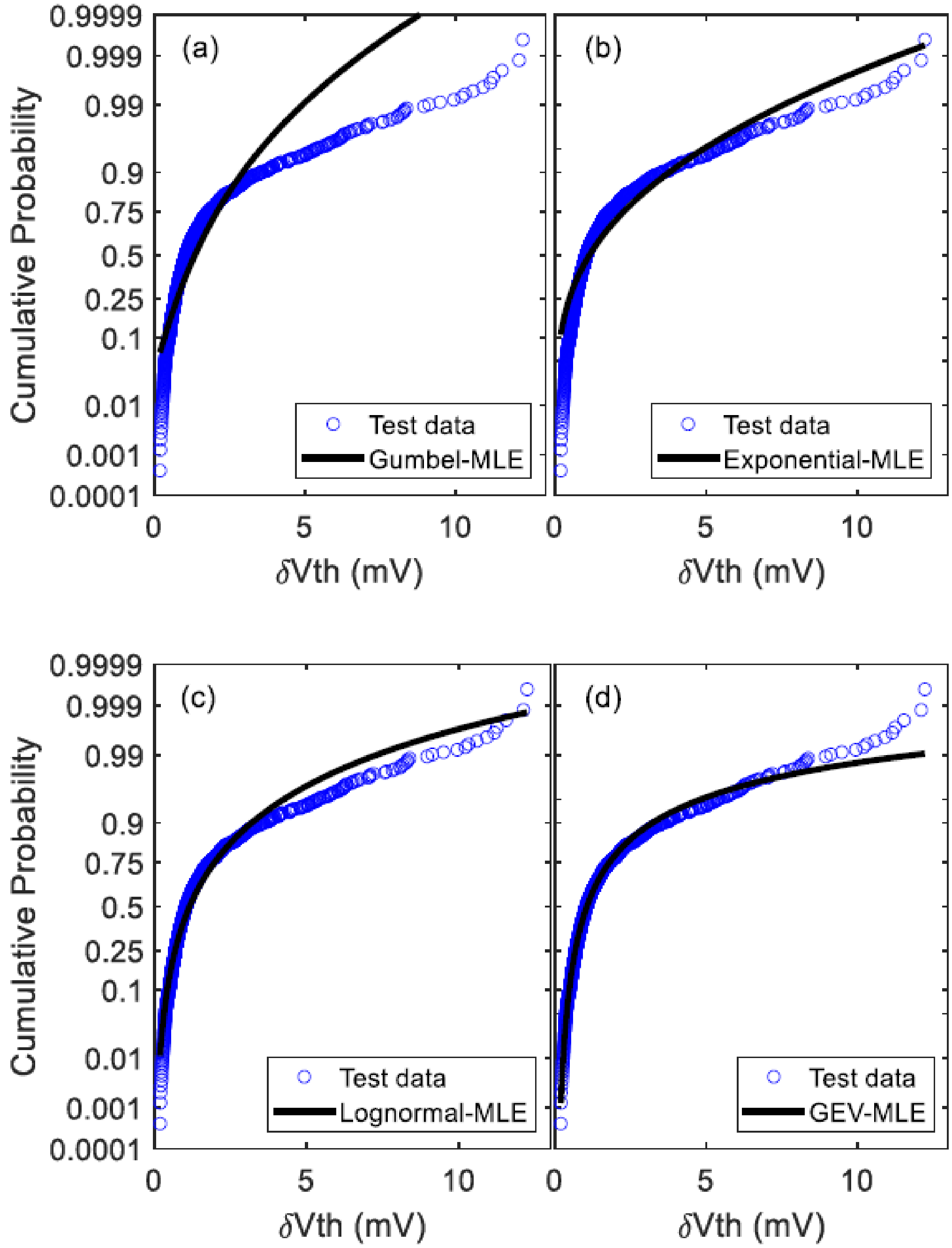


Figure 5.4.3 A comparison of the tail region between the test data (symbols) and the CDFs (lines) extracted by Maximum likelihood estimation (MLE) for (a) Gumbel, (b) Exponential, (c) Lognormal, and (d) GEV. The vertical axis is plotted linearly for the Z-score corresponding to the cumulative probability.

The minimum SSE per trap for the whole data set should not be the only criterion for selecting statistical distribution functions. As the loss of yield is mainly caused by traps in the distribution tail, the SSE in the tail region should also be examined. To see the tail clearly, the corresponding Z-score of CDF is plotted linearly in Figure 5.4.3. Some earlier works used ~100 traps [68], [106], [127] so that there are too few traps to evaluate the SSE reliably in the >95% tail. With 1,178 traps here, their SSE in the >95% tail is compared in Figure 5.4.4. Although the Lognormal CDF has lower SSE for the whole data set in Figure 5.4.2 (a&b), Figure 5.4.4 (a&b) shows that the Exponential CDF actually matches the test data better in this tail region. The choice between Lognormal and Exponential is not straightforward, therefore.

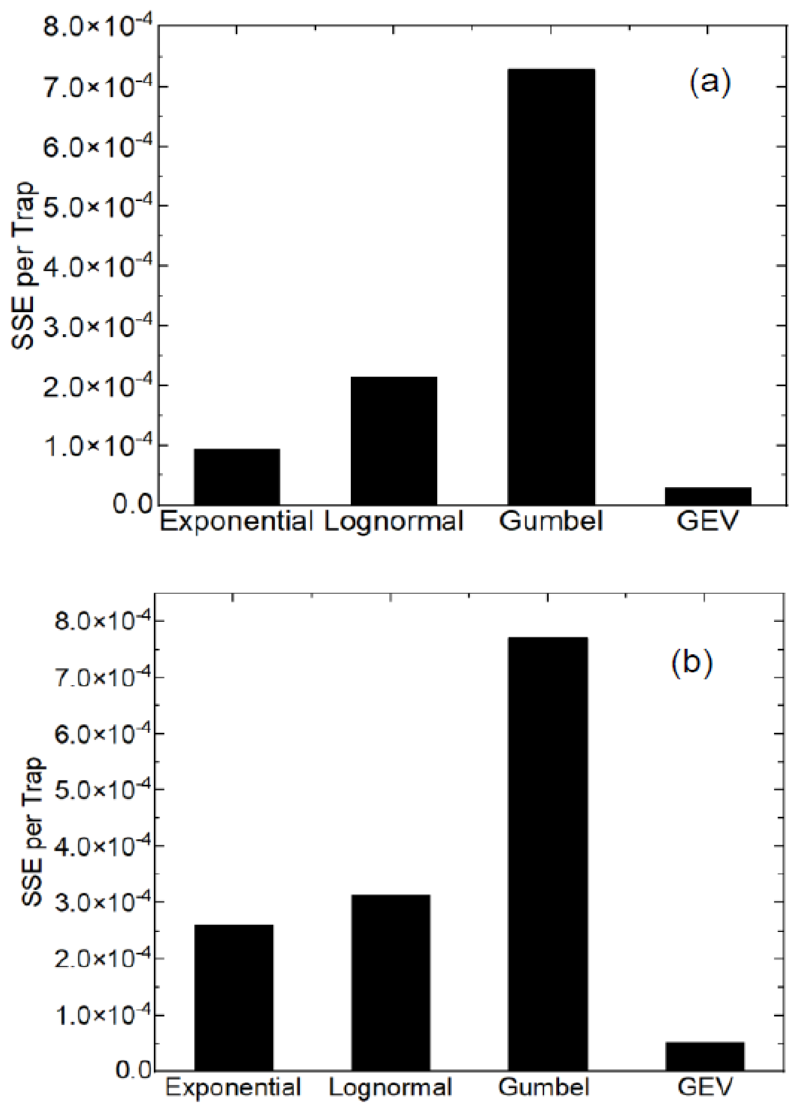


Figure 5.4.4 The error per trap of CDFs in the >95% tail extracted by MLE at 125 ^{o}C (a) and 28 ^{o}C (b). Exponential has smaller error than Lognormal in the tail.

5.5 Bimodal Statistical Distributions

To improve the accuracy of CDFs, bimodal CDF, BCDF, has been proposed:

$$\mathrm{BCDF} = \mathrm{p} * \mathrm{CDF1} + (1 - \mathrm{p}) * \mathrm{CDF2} \tag{21}$$

where $0 \leq p \leq 1$ is an adjustable parameter that can be fitted by using the MLE method [139]. CDF1 and CDF2 are two monomodal distributions. Both bimodal Exponential [135] and Lognormal [70] CDFs have been used. It has been suggested that CDF1 and CDF2 originate from traps in different layers of gate dielectric that have different statistical properties [135].

Figure 5.5.1 (a-f) shows the bimodal CDFs extracted by the MLE method for Lognormal, Exponential, and Gumbel, respectively. The bimodal Lognormal in Figure 5.5.1 (a&b) is dominated by the first Lognormal CDF and the contribution of the second Lognormal CDF is weak with a p value of only 0.033. For bimodal Exponential CDFs, the second CDF only counts for 9% in Figure 5.5.1 (c&d). This increases to 25% for the bimodal Gumbel in Figure 5.5.1 (e&f).

To compare the bimodal CDFs with their monomodal counterparts, the errors from their Z-score plot were calculated as shown in Figure 5.5.1 (b, d, and f). This places more weightings on the distribution tails where accuracy is important for RTN modelling. It is more appropriate than the error calculation directly from CDF values in Figure 5.4.2, therefore.

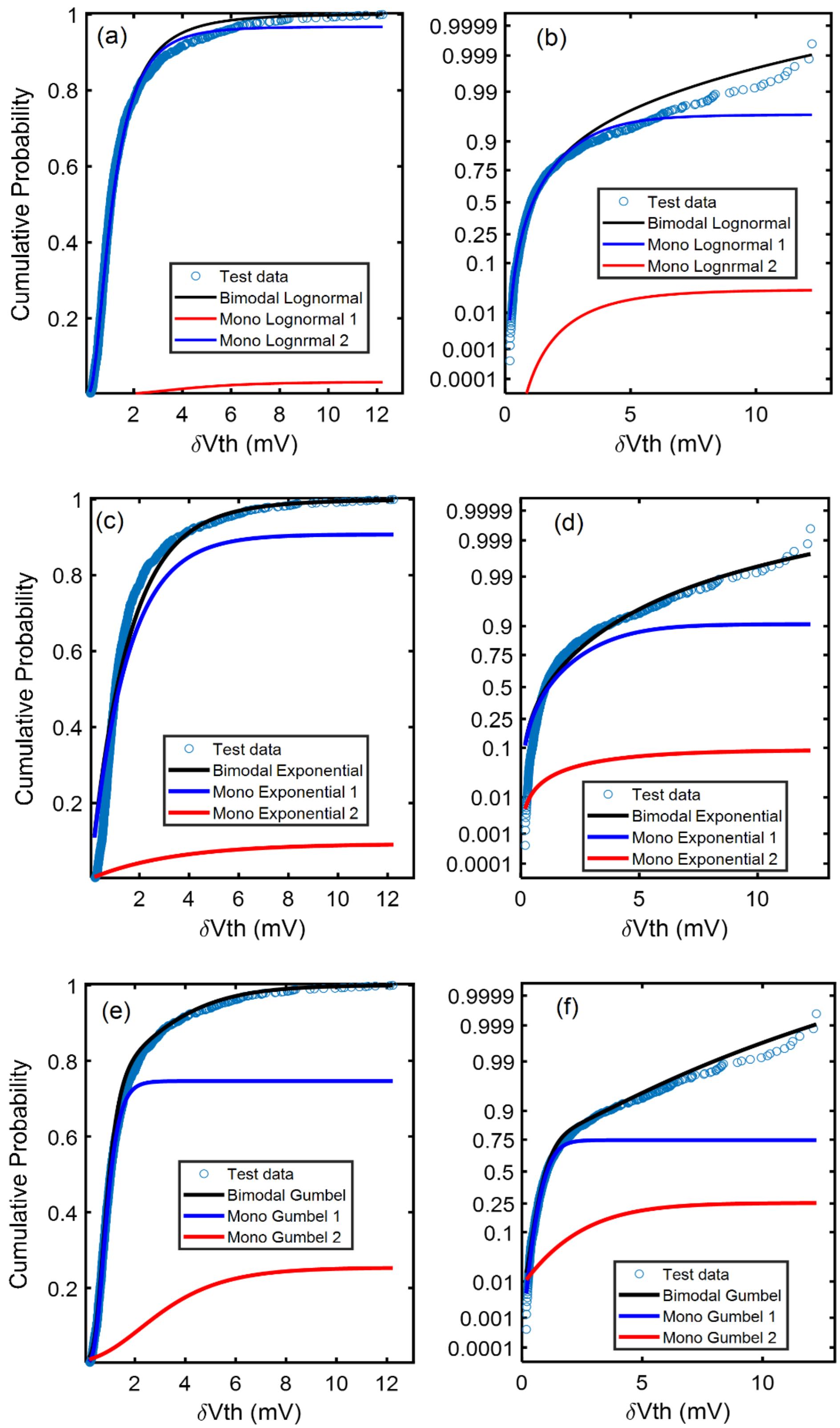


Figure 5.5.1 Bimodal CDFs for Lognormal (a) and (b), Exponential (c) and (d), and Gumbel (e) and (f). (b), (d) and (f) are the Z-score to enlarge the tail region. The symbols are test data. The black lines are the sum of two CDFs. The blue and red lines are the monomodal CDF1 and CDF2, respectively.

Figure 5.5.2 shows that, although bimodal Gumbel has a smaller error than monomodal Gumbel, it is still well above the error of monomodal Lognormal. The impact of using bimodal CDFs on the error is modest for both Lognormal and Exponential. When compared with monomodal CDFs, bimodal CDFs more than double the number of fitting parameters. According to the Bayesian Information Criterion [141], a penalty should be applied to models with more fitting parameters, so using bimodal CDFs is not strongly supported by the data in this work. The question is whether there is a monomodal CDF that can give a similar or even smaller error rate than the lowest error achieved by the bimodal Lognormal CDF in Figure 5.5.2. This will be investigated in the next section.

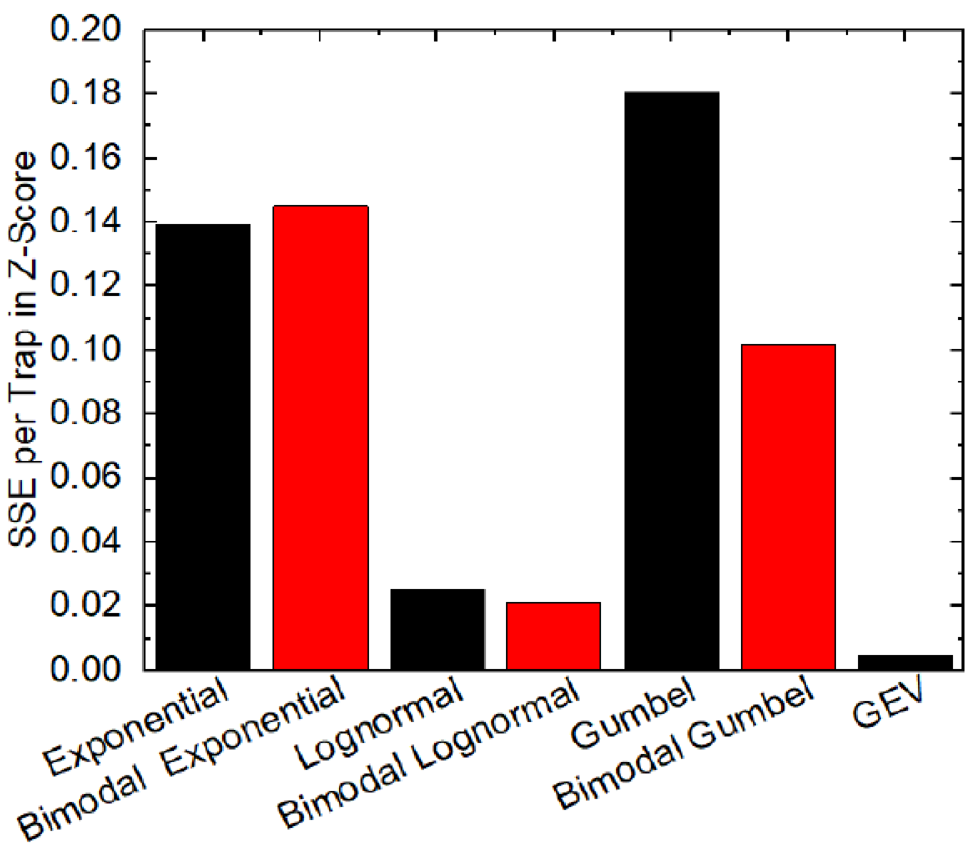


Figure 5.5.2 A comparison of errors in bimodal CDFs with their monomodal CDFs for Exponential, Lognormal, and Gumbel: The SSE per trap is calculated from the Z-score for the whole data set. The use of bimodal CDFs has not reduced errors below the level achieved with a monomodal GEV.

5.6 Generalized Extreme Value (GEV) distribution

Driven by the desire to find a statistical distribution that has the lowest SSE per trap without using its bimodal CDF, 13 other distributions [142] have been evaluated and their SSE per trap is compared in Figure 5.6.1, together with the three distributions used in earlier works. Among the 16, the Generalized Extreme Value (GEV) distribution has the lowest error. It is worth exploring this distribution further, therefore.

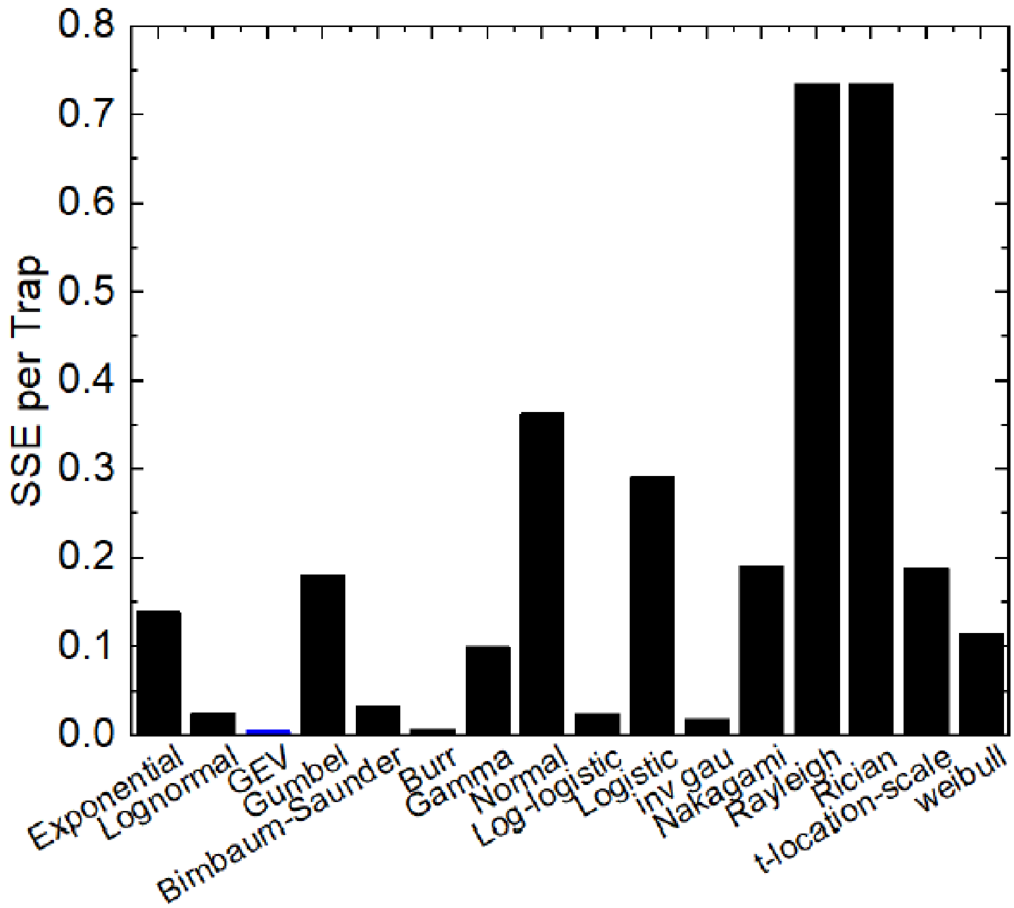


Figure 5.6.1 A comparison of the SSE per trap for 16 CDFs [39]. The error is calculated from the Z-score for the whole data set. The Generalized extreme value (GEV) distribution has the lowest error.

The equation for GEV is included in Table 3 and the extracted parameter values are given in Table 4. Figure 5.4.1 (d) shows that the CDF of GEV agrees well with the test data overall. Although Figure 5.4.3 (d) shows that the difference between GEV and the highest few data points appears to be increasing, this is an artefact, as the last few points of any test data are always lifted upwardly by the limitation in the size of the data set. The Z-score approaches infinity when CDF approaches

1. As the last data point has CDF=1, its Z-score would be infinity. To avoid this, it is a common practice to calculate the CDF of test data by equation (22) [143],

$$\mathrm{CDF}(\delta \mathrm{Vth}, \mathrm{i}) = \frac{\mathrm{i} - 0.5}{\mathrm{N}} \tag{22}$$

where i=1 has the lowest δVth and i=N=1,178 has the highest δVth in our test data set. This brings the last CDF point from 1 to 0.999576 and their corresponding Z-score from infinity to 3.34. It, however, cannot completely eliminate the artificial up-swing of the last few data points.

Figure 5.4.2 and Figure 5.4.4 show that the GEV has the lowest error for both the whole data set and the tail region when compared with other CDFs. Figure 5.5.2 shows that the error of monomodal GEV CDF is also lower than that of the bimodal Lognormal, Exponential, and Gumbel CDFs. The number of fitting parameters is 5, 3, and 5 for the bimodal Lognormal, Exponential, and Gumbel CDFs, respectively. It is 3 for the GEV in Figure 5.5.2, so the better accuracy of GEV was not gained from using larger number of fitting parameters.

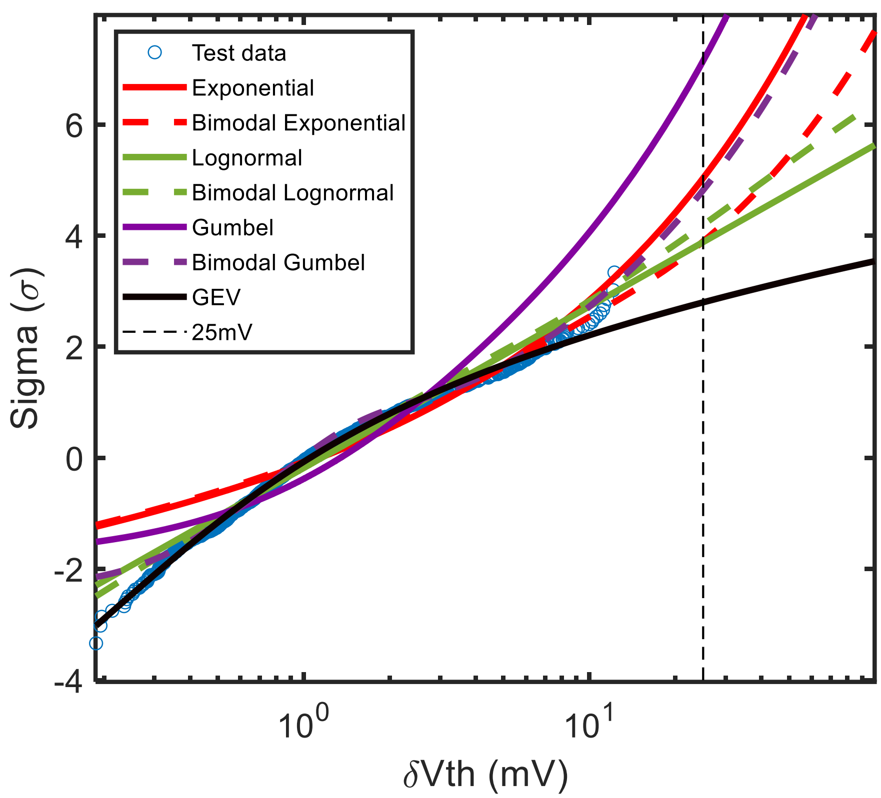


Figure 5.6.2 A comparison of different CDFs extracted from the same data set (symbols). The solid lines are the monomodal CDFs and the dashed lines are their bimodal counterparts for the same color.

Figure 5.6.2 compares the CDFs of different distributions extracted from the same data set. The predicted distribution tail is sensitive to model selection. The δVth at high σ increases in the order of Gumbel, Exponential, Lognormal, and GEV. In other words, Gumbel has the shortest tail and gives an optimistic prediction, while the GEV has the longest tail and gives a pessimistic prediction. Figure 5.6.3 compares the probability for δVth ≥ 25 mV predicted by different CDFs. Quantitatively, it is 4.5×10^{-7}, 0.24, 52, and 2553 parts-per-million (ppm) for Gumbel, Exponential, Lognormal, and GEV, respectively. This highlights the uncertainty of RTN prediction when using different CDFs.

On the applicability of the conclusions drawn here to other fabrication processes, ideally this work should compare the results of samples fabricated by different processes. However, only one wafer from one company was available. The test samples used here were fabricated by the 28 nm CMOS process, which has been widely used commercially. The results reported here should be a typical representation of industrial processes, but further work will be needed to confirm this.

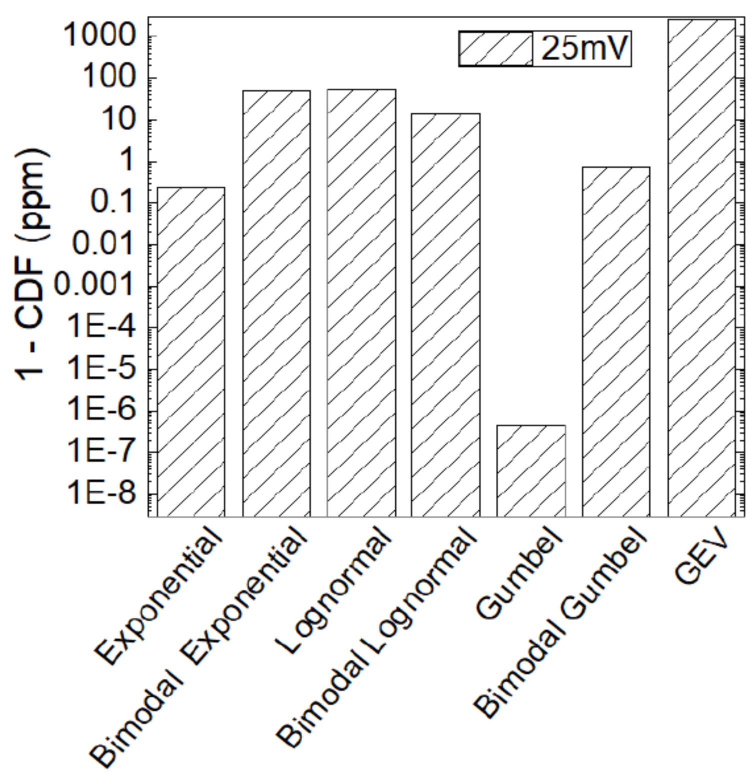


Figure 5.6.3 A comparison of the probability of occurrence for δVth≥25 mV predicted by different CDFs. The CDF values at δVth=25 mV were taken from Fig. 11, as marked out by the vertical dashed line.

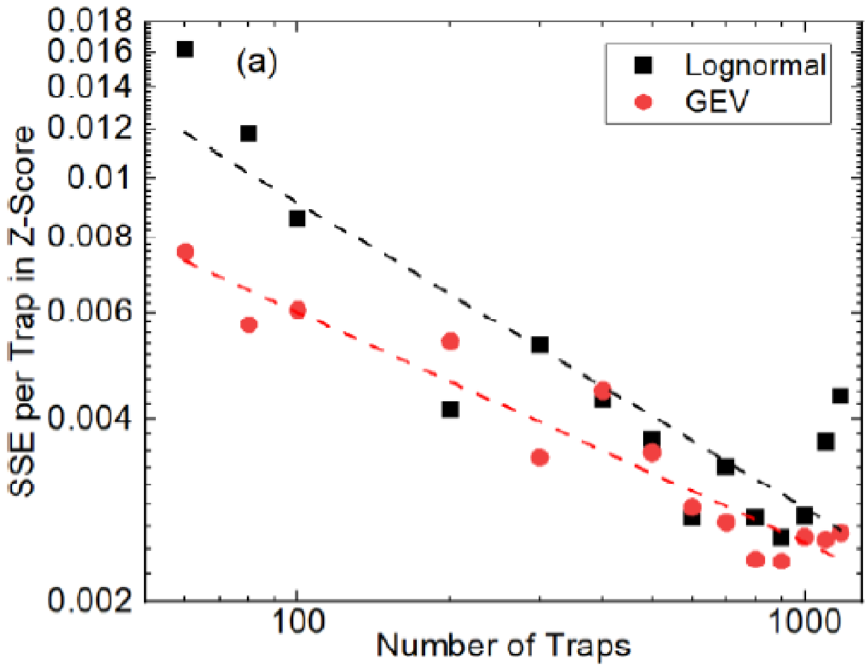
5.7 New Model Selection Criterion

Given the large uncertainties in RTN predicted by different CDF models, further work is needed to justify their selection, in addition to their errors. Ideally, the selected model should be justified by device physics. Unfortunately, this work was unable to link the Exponential, Gumbel, and GEV with a physical process, as these models are empirical [4]. GEV is developed from the extreme value theory to capture the long distribution tails, with Gumbel, Fréchet, and Weibull distributions as its special cases [136]. The number of traps in a device is minimised in a modern commercial CMOS process through quality control and one may consider that having a trap right above the narrow percolation current path in Figure 5.1.1 (b) is extremely rare.

The Lognormal CDF has been interpreted physically [70], [127]. As the number of charge carriers in the channel depends on (Vg-Vth) exponentially in the subthreshold region, a local Vth fluctuation spatially leads to an exponential fluctuation of local density of charge carrier, n. If Vth varies spatially by following a Normal distribution, Log(n) will vary by following a Normal distribution. The impact of a trapped charge on the channel is proportional to n, so that Log(δVth) will also follow a normal distribution, i.e. δVth follows Lognormal distribution [127].

There are, however, two difficulties with this interpretation. One is that Id was monitored above threshold in typical RTN tests, where n no longer depends on (Vg-Vth) exponentially. The other is that the impact of trapping on carrier mobility is neglected here [127]. It has been reported that the contribution of charge-induced mobility degradation is similar to that of carrier number reduction [5].

In searching for further criteria for model selection, the dependence of error per trap on the number of traps was examined. If the test data truly follows a specific CDF, then it is expected that the error per trap decreases with an increasing number of traps, because an infinite number of data points should produce this specific CDF perfectly. To support this statement, a theoretical Lognormal CDF is used to randomly generate a number of data points and then treated as ‘test data’. This ‘test data’ was used to extract the Lognormal CDF and the errors were evaluated in the same way as for the real test data. Figure 5.7.1 (a) shows that the SSE per trap indeed reduces for a higher number of traps, despite the statistical scattering. The same also applies to the GEV CDF.



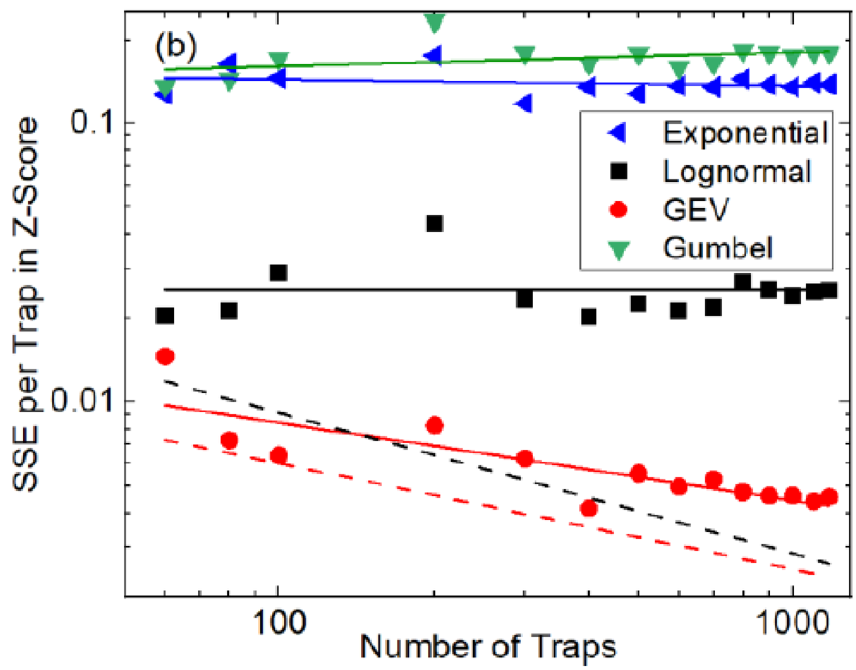


Figure 5.7.1 Dependence of SSE per trap on the number of traps used to extract the CDFs. (a) The data are generated randomly from the theoretical Lognormal (■) and GEV (●). They were treated as the 'test data' and used to extract Lognormal and GEV CDFs, respectively. Their SSE per trap decreases with increasing trap number, as shown by the fitted dashed lines. (b) The real test data were used to extract CDFs and calculate SSE per trap. The solid lines are fitted. For comparison, the two dashed lines in (a) were replotted in (b). Only GEV clearly shows the expected decrease of errors with increasing trap number.

Figure 5.7.1 (b) shows the dependence of SSE per trap on the trap number for the real test data. Only the error of GEV exhibits a clear decrease for a higher number of traps. To quantitatively compare the error of theoretical and real test data, the two fitted lines in Figure 5.7.1 (a) were reproduced as the two dashed lines in Figure 5.7.1 (b). The difference in the error between the theoretical and real test data is substantially larger for Lognormal, when compared with a GEV distribution. This supports the use of the GEV model.

5.8 Impact of trap numbers on Prediction accuracy

Figure 5.7.1 (a) shows that the error per trap reduces for a higher number of traps. In practice, the number of available traps is always limited. When the CDFs extracted from a limited number of traps are used to predict the RTN in the long tail, an important question is how accurate it is.

To assess the impact of trap number on this accuracy, one needs a reference distribution as the benchmark. Here, the GEV distribution extracted in Figure 5.6.2 is used as the reference and its parameters are given in Table 4. One set of ‘N’ data is randomly generated according to this distribution, as shown in Figure 5.8.1. This ‘N’ data set is then used to extract the statistical distribution, which gives the orange curve in Figure 5.8.1. The difference between the fitted and the reference distributions (the blue curve in Figure 5.8.1) at a given σ can then be determined, as illustrated by the dashed lines in Figure 5.8.1. By repeating this process 1,000 times and each time with a different and randomly generated set of ‘N’ data, the confidence for the accuracy of statistical distributions extracted from a set of ‘N’ data can be obtained [122].

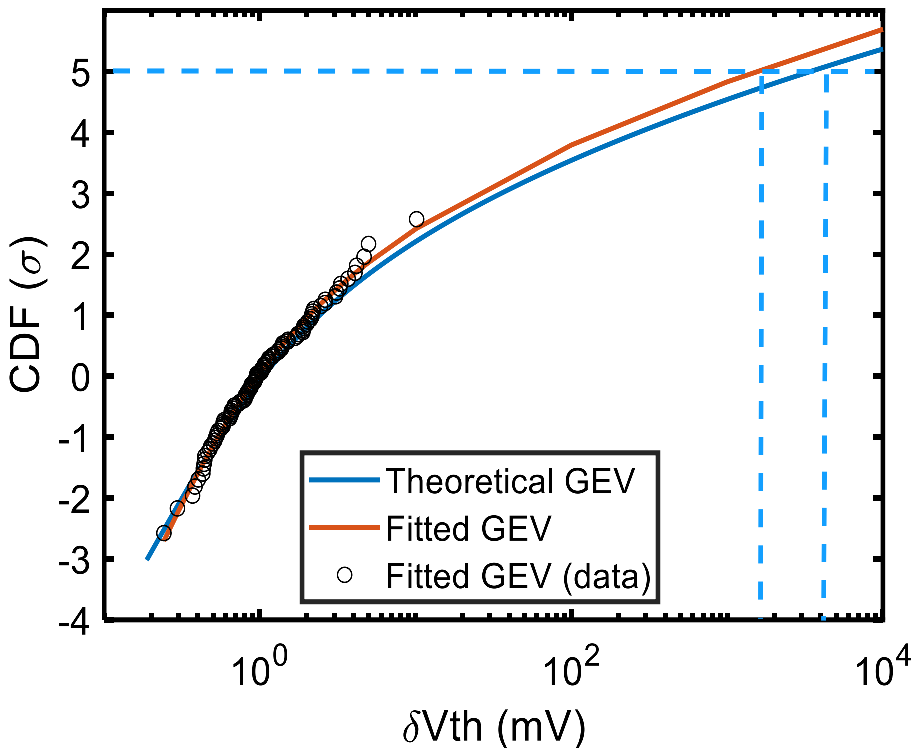


Figure 5.8.1 A comparison between theoretical GEV by using parameters in Table II and the GEV fitted by using 100 hypothetical traps, which were randomly generated by the theoretical GEV. The difference between theoretical and fitted GEV at 5σ is shown by the dashed lines.

Figure 5.8.2 (a) shows the error at 3σ for different N. For N=100, the error at 90% confidence is -58.35% and 57.42%, respectively. For N=1,000, these two errors are reduced to -13.68% and 13.18%. If one targets an accuracy of 15% at 3σ with 90% confidence, 1,000 traps can be used.

Figure 5.8.2 (b) shows the errors for N=1000 at different sigma. The error increases from -13.68% and 13.18% at 3σ to - 37.15% and 26.5% at 5σ (a probability of 0.57 parts per million) for 90% confidence. To be conservative, the guide-band for RTN induced δVth at 5σ should be increased by 26.5% from the value predicted by the statistical distribution extracted from 1,000 traps, therefore.

With 1.000 traps, the probability of occurrence for δVth ≥ 25 mV is between 1584.2 and 3537.6 with 90% confidence. This uncertainty is substantially smaller than that from using different CDF models shown in Figure 5.6.3. Here, it is concluded that the uncertainty in RTN amplitude prediction is dominated by model selection.

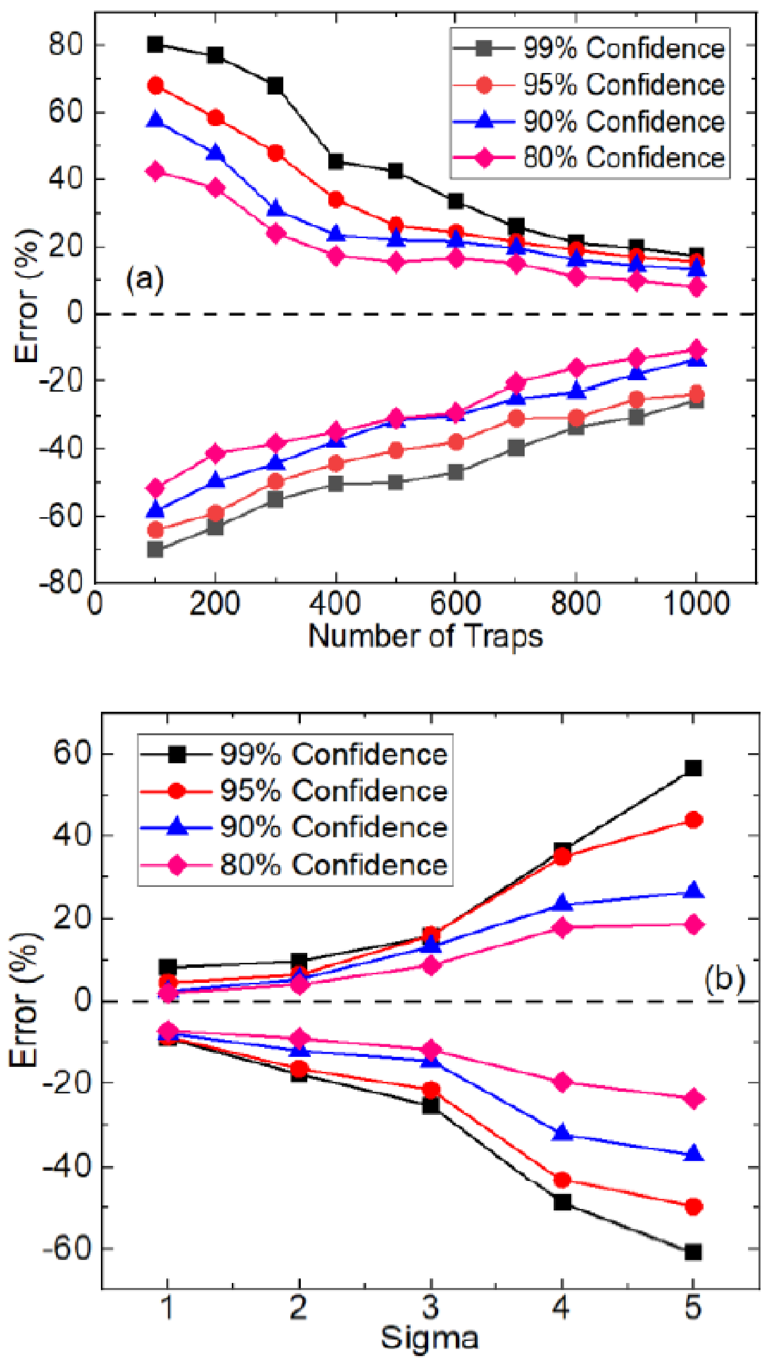


Figure 5.8.2 (a) shows the errors of prediction at 3σ by the CDFs extracted from different number of traps with 80%, 90%, 95% and 99% confidence. (b) shows the errors at different σ for 1,000 traps.

5.9 Summary

This work assesses the accuracy of the statistical distributions for the RTN amplitude per trap. Its novelty includes proposing a new model selection criterion based on the relation between error and trap number, exploring the applicability of a wide range of statistical distributions to RTN amplitudes, and finding that the Generalized Extreme Value (GEV) distribution has the least Z-score based error. The new model selection criterion requires a monotonic error decrease for a higher number of traps. The GEV meets this criterion, while the Exponential, Lognormal, and Gumbel distributions do not. Based on the data used in this work, using bimodal Exponential and Lognormal CDFs only has a modest impact on the error, despite the increased fitting parameters.

The accuracy of CDF extracted from a limited number of traps is also assessed. For 90% confidence, the guide-band for RTN induced δVth at 5σ should be increased by 26.5% from the value predicted by the statistical distribution extracted from 1,000 traps. The uncertainty caused by using a limited number of traps is relatively small and the selection of the CDF model dominates the uncertainty in RTN amplitude prediction and modelling.

Chapter 6 Impact of Gate voltage

6.1 Introduction

Decreasing the size of MOSFETs down to nanometre scales induces an increased defect influence on carrier mobility, subthreshold swing, and threshold voltage. Smaller devices have a larger statistical spread, due to there being fewer traps per device and the larger impact of a single trapped charge on them as mentioned in Chapter 5 [13], [144]. If the number of devices per chip increases, this will lead to a larger statistical spread [13], [144]. This, combined with a high data transmission rate, will require tight control of fluctuation [145]. Hence, fluctuation has become a major concern for circuit design and has attracted much attention [20], [85], [88], [112], [146]-[149][148]. Fluctuation is commonly monitored as random telegraph noise in the drain current under constant gate bias [20], [85], [88], [112], [147]-[149]. It is known that RTN fluctuation defined in terms of ΔId/Id or as a corresponding ΔVth does not follow a normal distribution but shows a long distribution tail. Most transistors typically have low ΔVth fluctuation while there are a few devices that have extremely large ΔVth, which affects circuit and SRAM stability. Thus, the analysis of the statistical distribution of RTN is important.

Chapters 4 and 5 cover an assessment of the accuracy of the statistical distribution of capture/emission time constant and amplitude distribution of RTN, respectively. In both cases a low bias condition close to threshold voltage was selected for low power applications and also to minimise trap generation. However, it is known that RTN is Vg dependent [67], so that this Chapter is conducted to assess the impact of gate voltage on RTN amplitude and time constant distribution. The objective is to study the impact of gate voltage (Vg) on RTN distribution and discuss the unique features of different statistics that make them suitable for different gate biases.

6.2 Device and Measurement condition

As in section 4.3.2 and 5.3.1, the same 27x90 nm nMOSFETs were used with all the test conditions kept the same but a varied gate voltage. This time the gate voltage applied is Vg = 0.9V and the test was carried out under 28ºC.

In previous chapters, the RTN amplitude analysis was expressed in terms of ΔVth, evaluated from -ΔId/gm because the gate voltage used was Vg = 0.5 V, which is close to the threshold voltage

(Vth). However, as demonstrated in Figure 6.2.1, RTN is dominated by a different group of traps when gate bias is increased [150]. RTN is dominated by traps near Fermi-level at dielectric/substrate interface, Ef, as shown in Figure 6.2.1 (a). As Vg increases, the traps near to Ef change. There are more traps at higher |Vg| and the RTN is dominated by a different group of traps under different Vg. The charging and discharging of traps responsible for RTN is highly dynamic in nature so at any specific time, the traps can be neutral and could be missed if the conventional measure-stress-measure (MSM) technique is used. Due to these difficulties, early work [20], [149], focuses on measuring the fluctuation in drain current ΔId under fixed Vg and then converting to ΔVth by dividing gm. The accuracy of this ΔVth conversion was not clear as it was not compared with the real ΔVth measured at Vg = Vth.

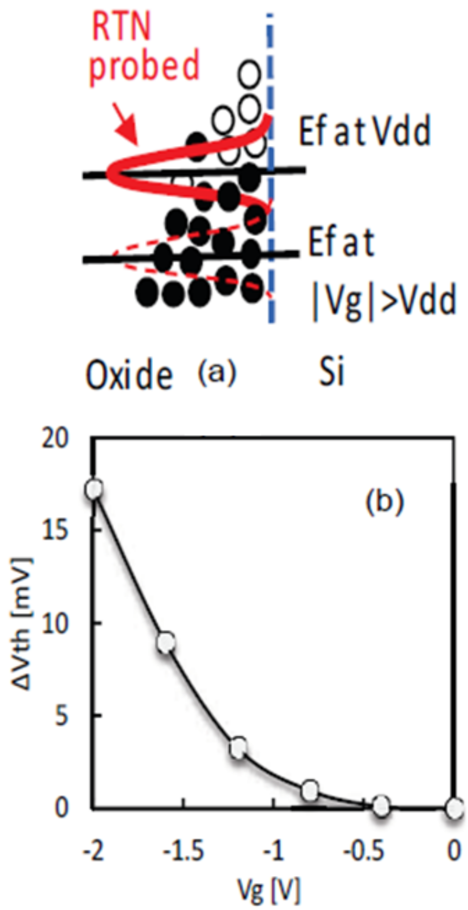


Figure 6.2.1 (a) RTN is dominated by traps near to Ef at the interface. Vg-acceleration changes the traps near Ef probed by RTN. (b) More traps under higher |Vg| results in higher ΔVth [150].

In the recent work of A. Manut *et al*, the Trigger-When-Charged (TWC) method [98] assesses the inaccuracy of measuring ΔVth from ΔId/gm (Vdd). The TWC technique exploits the oscilloscope-based fast measurement system in which the edge of RTN is used to trigger the entire Id-Vg measurement. Understanding the entire Id-Vg (the transfer characteristic curve) and its shift induced by the RTN is important as it can provide valuable information in understanding its underlying physical mechanism [151]-[153]. A. Manut *et al.* [98] measured ΔVth under Vdd and compared it with the real ΔVth (Vg=Vth) and reported that the former is twice the latter on average. One reason given is the mobility degradation at higher Vg and the current distribution also contributes to this discrepancy. Therefore, the ΔVth from ΔId/gm method is not used in this work. The magnitude of RTN is measured in terms of ΔId/Id instead.

6.3 Gate Voltage impact on RTN amplitude distribution

In this research, the measurements have been done under two different gate biases. Firstly, Vg close to threshold voltage and then Vg at operating condition, that is Vg = 0.5V and Vg = 0.9V respectively. In Chapter 5 , the accuracy of the RTN amplitude distribution fitting has been studied and it has been concluded that the Generalized Extreme Value distribution (GEV) fits the data best giving the lowest error when applied gate voltage is 0.5V. A new model selection criterion has been introduced and GEV distribution also met the criterion [154]. From the early works [154] it is known that the amplitude of different RTN varies with gate bias and on average the distribution of RTN induced ΔVth significantly increases with gate bias. In contrast ΔId/Id decreases under higher bias [63]. To find the impact on the statistical distribution of RTN amplitude with different biases, Vg = 0.9V is applied at the gate in this Chapter. The method introduced in Chapter 5 is repeated under the new Vg condition to extract the RTN amplitudes. The same distribution fitting criterion as Chapter 5 is then applied to find out whether the RTN distribution model fits with GEV distribution or changes when different gate bias is applied. Further investigation and discussion are in the following sections.

6.3.1 Effect of Sensing Vg on ΔVth

Power consumption is a key issue, especially for the edge devices or units in an IoT system. One of the effective ways to reduce power is by lowering the operating voltage. As the overdrive voltage, (Vg-Vth), becomes smaller, the device gets more vulnerable to threshold voltage jitters.

RTN is considered one of the sources for the jitter that causes fluctuation in both drain current, ΔId, and threshold voltage, ΔVth. Early works on RTN were focused on measuring ΔId and then evaluating ΔVth from ΔId/gm as mentioned above. However, the accuracy of ΔVth obtained in this way was not known and the work in [98] assessed its accuracy and reported that ΔVth must not be evaluated from ΔId/gm. One of the reasons reported about this discrepancy is the device-specific localized current distribution near the threshold [98]. Figure 6.3.1 (a-c) are three different devices' results taken from [150], showing device-to-device variation is substantial and the dependency of ΔVth on the sensing Vg.

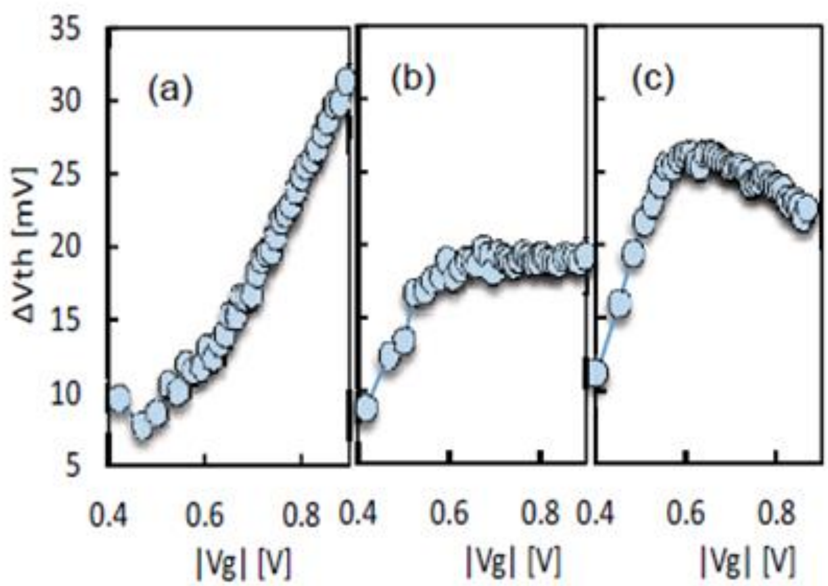


Figure 6.3.1 The dependence of ΔVth on sensing Vg for three nano-scale devices. (a) ΔVth increases with |Vg| (b) ΔVth saturates (c) ΔVth turns around [150].

In Figure 6.3.1 (a) the ΔVth increases with |Vg| and in (b) the ΔVth saturates as |Vg| increases and in (c) ΔVth turns around i.e. it increases initially and then decreases for higher |Vg|. It has been reported that the current can have a narrow percolation path near Vth but becomes more evenly distributed as |Vg| increases. The impact of a charged trap on Vth depends on the current density beneath it. That is, the higher the current density, the larger the impact [98],[155]. As illustrated in Figure 6.3.2, the trap for Figure 6.3.1 (a) is away from the localized current path at Vth so there is small impact and ΔVth is low. As |Vg| increases, a more evenly distributed current increases the relative local current density, so that the ΔVth increases in Figure 6.3.1 (a). For Figure 6.3.1 (c), the trap can be above the localized current path at Vth. As |Vg| increases, the initial rise is caused by increased mobility degradation and the subsequent decrease is caused by the relative reduction of current density under this trap, as the distribution spreads. Under Vth, localisation of current path leads to many traps being away from it, so that their impact on ΔVth is low. As |Vg| increases, the effect of these traps increases, contributing to the higher ΔVth for higher Vg [156]. Nevertheless, to avoid such discrepancy in this work, the RTN amplitude is expressed in terms of ΔId/Id. Two different voltage conditions: Vg near Vth and Vg at operating condition are used for comparison to observe the dependency of RTN amplitude distribution on Vg.

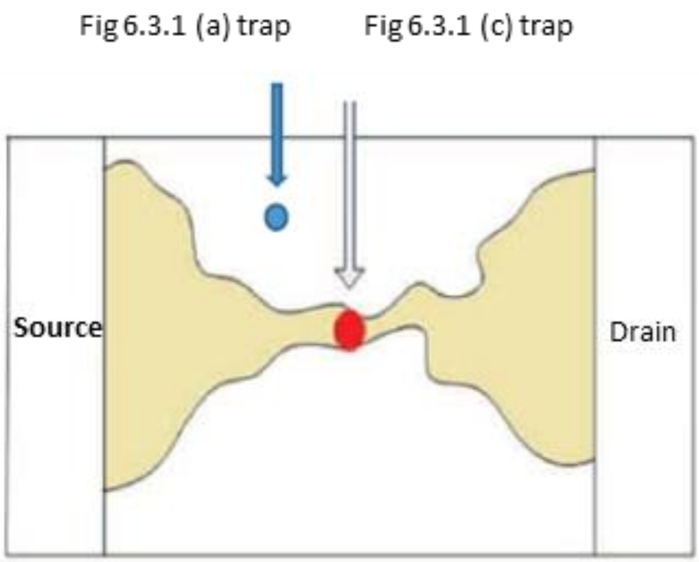


Figure 6.3.2 The localised current path at Vg=Vth. The trap in Fig 6.3.1(a) is away from the path, while the trap in Fig 6.3.1(b) is above it [150].

6.3.2 Different statistics at different Gate bias

Following the previous work from Chapter 5 [154], 1,020 traps were extracted from more than 400 devices at 28ºC and the cumulative distribution function (CDF) of the extracted traps is plotted together with the test data. In this work, further analysis of RTN amplitude distribution measured at Vg = 0.9V is carried out. The RTN is measured in terms of δId/Id.

To observe the impact of gate voltage on RTN amplitude distribution, the extracted traps are fitted using Gumbel, Exponential, Lognormal and GEV distributions to find out which distribution fits best to the test data. The CDF parameters shown in Table 5 for different statistics are estimated by using the equations in Table 3 and their corresponding extracted CDFs along with the test data are plotted in Figure 6.3.3.

Table 5 The estimated CDF parameters at Vg = 0.9V.

CDF (Lognormal)	CDF (Exponential)	CDF (Gumbel)	CDF (GEV)
$\epsilon = -1.0398$	$\eta = 0.468$	$\alpha = 0.3121$	$\xi = 0.6056$
$\theta = 0.7365$	-	$\beta = 0.2331$	$\alpha = 0.1604$
-	-	-	$\beta = 0.2497$

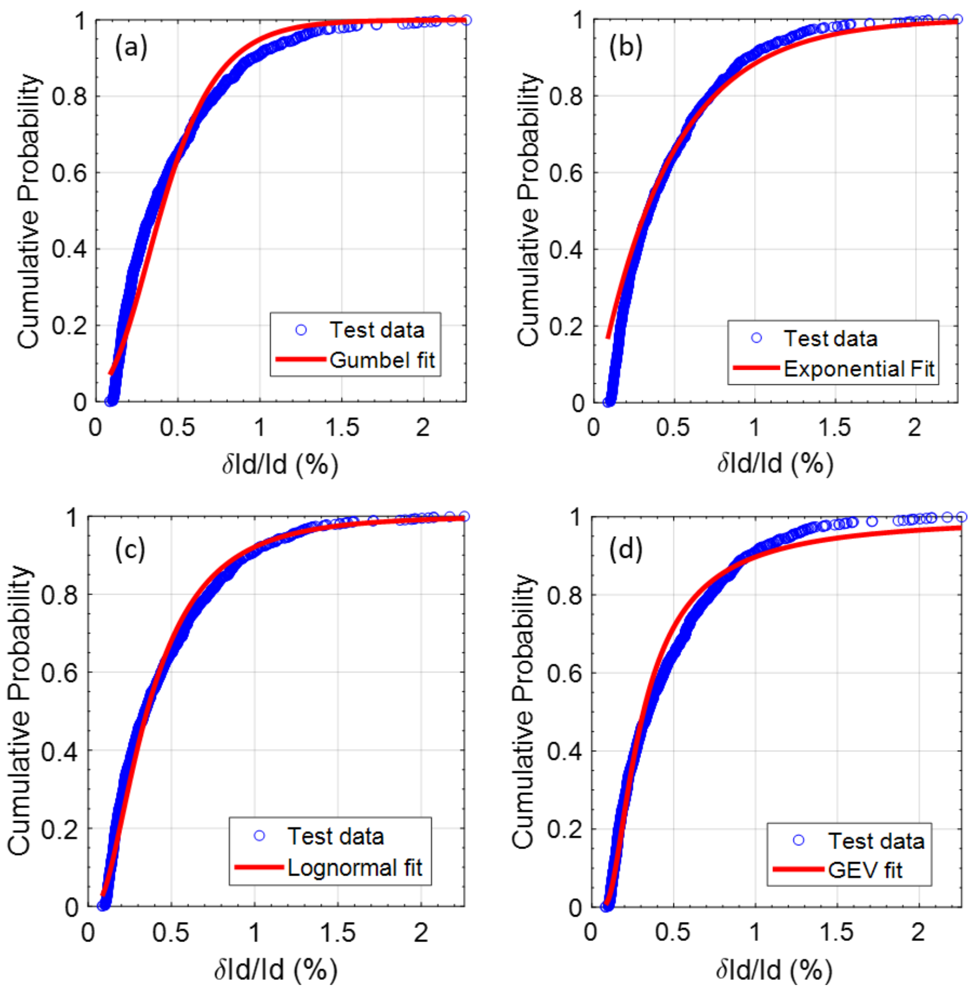


Figure 6.3.3 The CDFs (lines) extracted from 1020 traps by the MLE method are compared with the test data (symbols) for: (a) Gumbel, (b) Exponential, (c) Lognormal, and (d) GEV, respectively.

From Figure 6.3.3, Lognormal seems a better fit to the whole set of the data when biased at operating voltage. However, by considering only Figure 6.3.3, it is difficult to select which distribution best fits the test data. Hence, Figure 6.3.4 is examined to observe the tail by plotting the corresponding Z-score of CDF linearly.

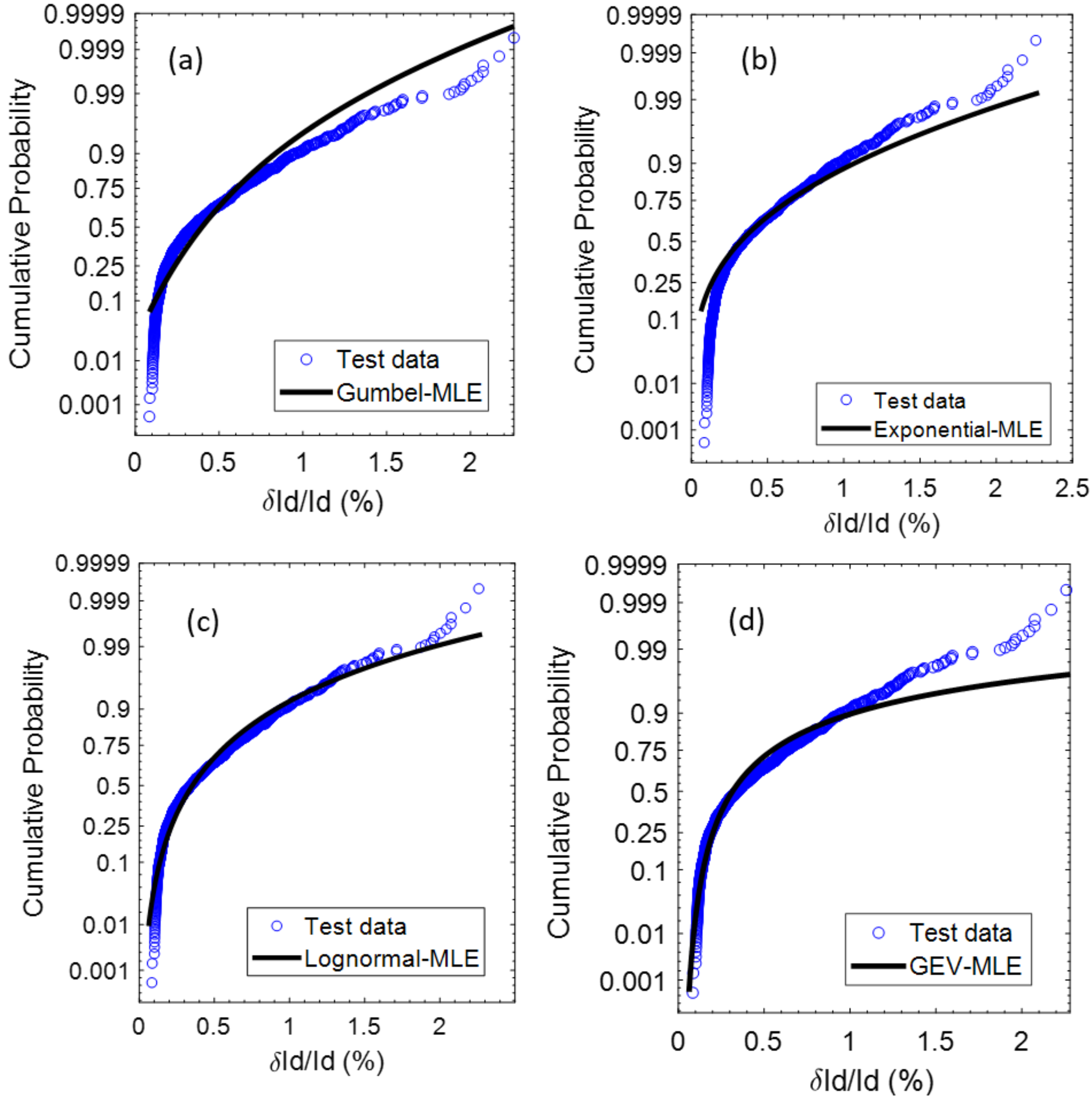


Figure 6.3.4 A comparison of the tail region between the test data (symbols) and the CDFs (lines) extracted by MLE for (a) Gumbel, (b) Exponential, (c) Lognormal and (d) GEV. The vertical axis is plotted linearly for the z-score corresponding to the cumulative probability.

From Figure 6.3.4 (c), it is observed that Lognormal also fits the tail distribution better than the other three distributions in (a), (b) and (d). For a quantitative discussion, from Figure 6.3.3 the corresponding SSE per trap for the whole data set and the data in the > 95% tail are also estimated as shown in Figure 6.3.5 (a) and (b), respectively.

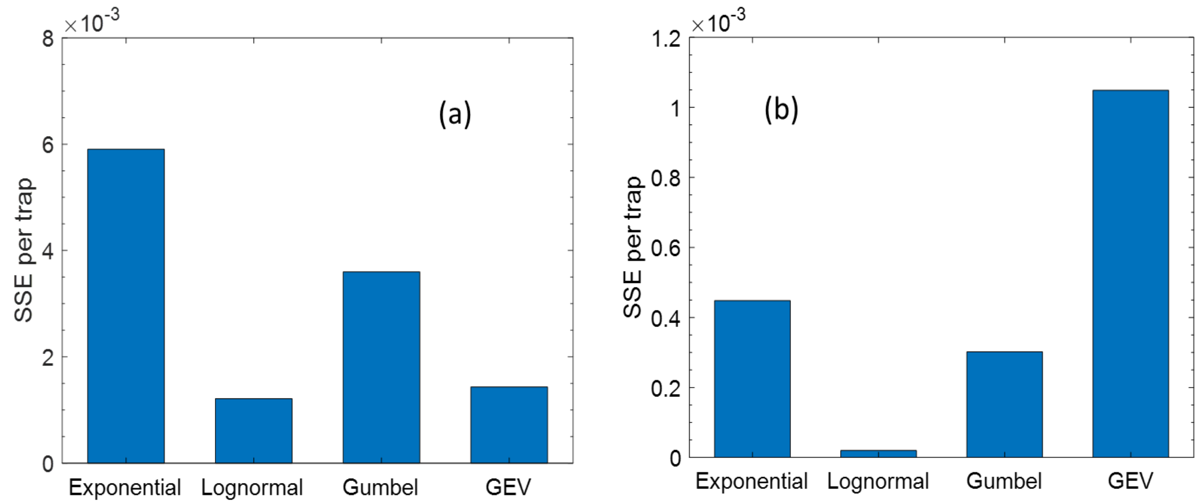


Figure 6.3.5 The error per trap of CDFs extracted by MLE at 28 ^{o}C for the (a) whole dataset used in the error evaluation and (b) in the >95% tail data. In both cases Lognormal has smaller error than other distributions.

The results in Figure 6.3.5 show the error for Exponential, Lognormal, and Gumbel. All three distributions decrease when considering the data in (b) tail region in comparison with the whole data set used in (a), except GEV distribution. The Lognormal distribution shows the lowest error of Exponential, Gumbel and GEV distributions when considering the whole data set for the error evaluation (a) and also when considering the data in the tail region. The physical interpretation of Lognormal CDF has already reported in [70], [127].

For further investigation, the newly introduced criterion in [154] is used. That is, the error per trap will decrease with the number of traps. If this is true, then the SSE per trap in Z-Score for Lognormal distribution should have decreased with increasing traps. Figure 6.3.6 (b) shows that the error per trap for all the distributions remains almost unchanged with a higher number of traps. Theoretically, based on the simulation result presented in Figure 6.3.6 (a), the error for Lognormal (black dots) in (b) should have decreased.

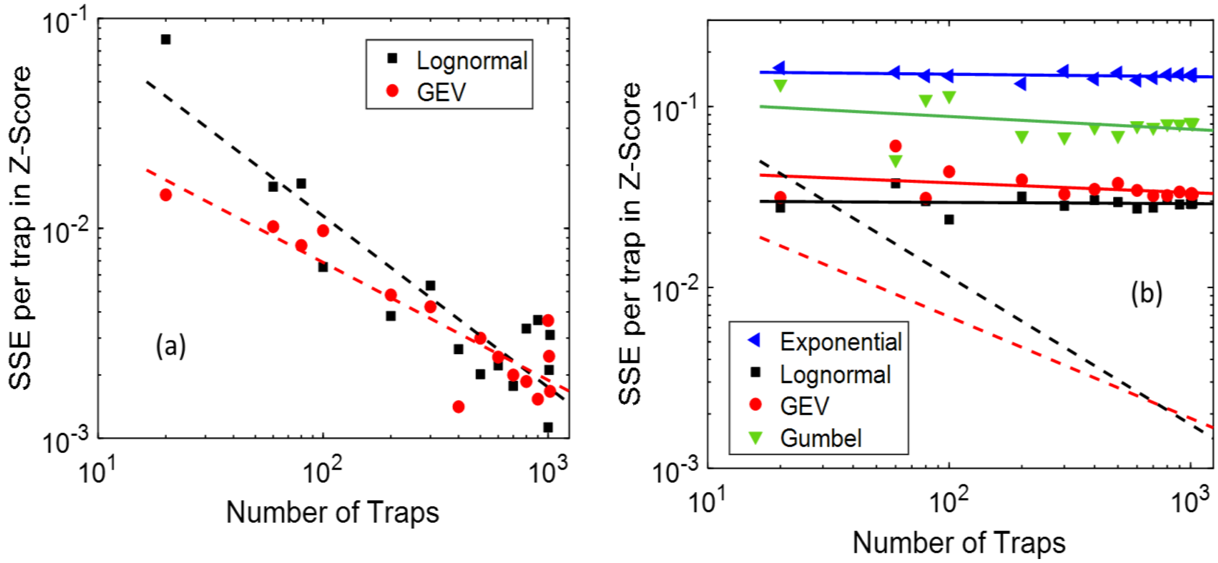


Figure 6.3.6 Dependence of SSE per trap on the number of traps used to extract the CDFs. (a) The data are generated randomly from the theoretical Lognormal (■) and GEV (●). They were treated as the 'test data' and used to extract Lognormal and GEV CDFs, respectively. Their SSE per trap decreases with increasing trap number, as shown by the fitted dashed lines. (b) The real test data were used to extract CDFs and calculate SSE per trap. The solid lines are fitted. For comparison, the two dashed lines in (a) were replotted in (b). None of the distributions show the expected decrease of errors with increasing trap number.

Therefore, based on the current data, it can be concluded that Lognormal fits the statistical distribution of RTN amplitude best when biased at Vg = 0.9V. However, this does not agree with the conclusion in chapter 5 for Vg=0.5V and needs further investigation.

6.3.3 Comparison between RTN amplitude distribution at different Vg

There can be different statistical distributions at different gate voltages. It is likely that current distribution is different at different gate bias. At 0.9 Vg there is a more uniform current distribution, whereas at 0.5 Vg, the conduction channel is weak and highly localized.

The best way to understand RTN distribution, is using the concept of a channel percolation path (PP) [67], [157]. The current will percolate through 'valleys' from source to drain as shown in Figure 6.3.7. When a trap captures a charge carrier from the conduction channel, it will generate a local potential barrier to disturb the current percolation path. Hence, the current fluctuation results from a trapped charge at various positions. If the trapped charge is located on the critical current path, as shown by red region in Figure 6.3.7, then RTN amplitude fluctuation is relatively large. On the contrary, if the trapped charge position is away from the percolation path, it will induce small current fluctuation.

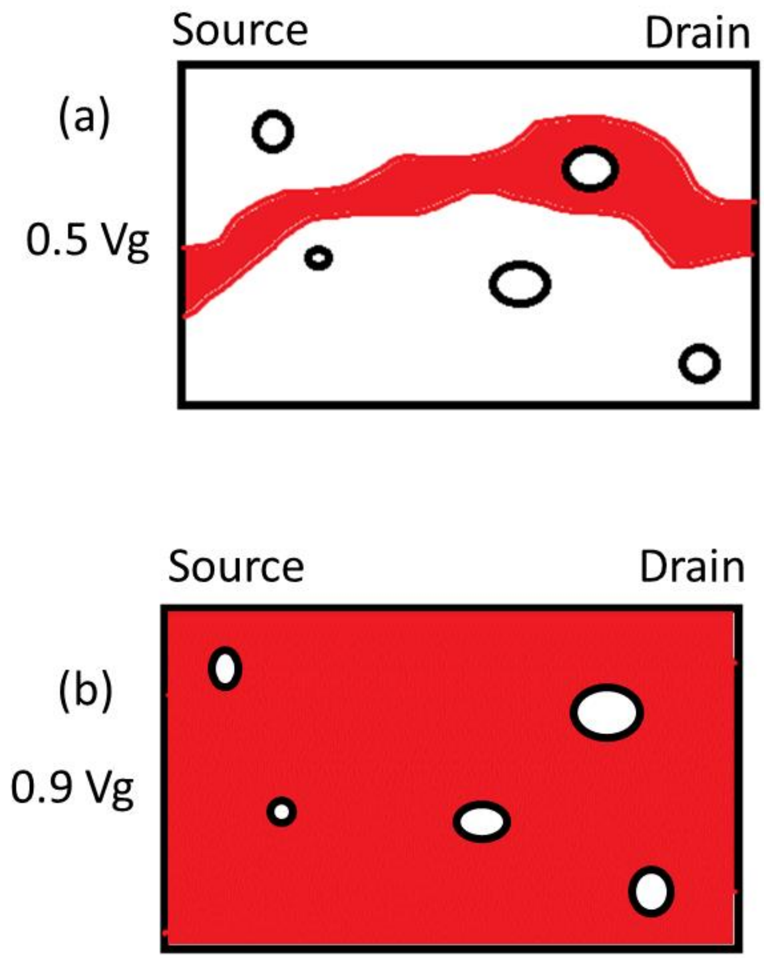


Figure 6.3.7 Illustration of percolation effect for current conduction width (a) narrower and (b) wider

As shown in Figure 6.3.7, (a) if Vg is close to Vth, the current conduction has a narrow percolation path (red region). A trapped charge on this narrow path will impact a greater portion of the percolation path and have a larger RTN amplitude. On the other hand, Vg=0.9V in (b) gives a more spread current path (red region), making an individual trap's impact proportionally smaller, since 0.5 Vg is a highly localised situation which is more likely to have extreme values. Therefore, amplitude distribution at 0.5Vg should have a longer tail in comparison to RTN distribution at 0.9 Vg. GEV captures the long tail better than the other distributions. This explains why GEV distribution fits these extreme cases better under Vg=0.5 V [154].

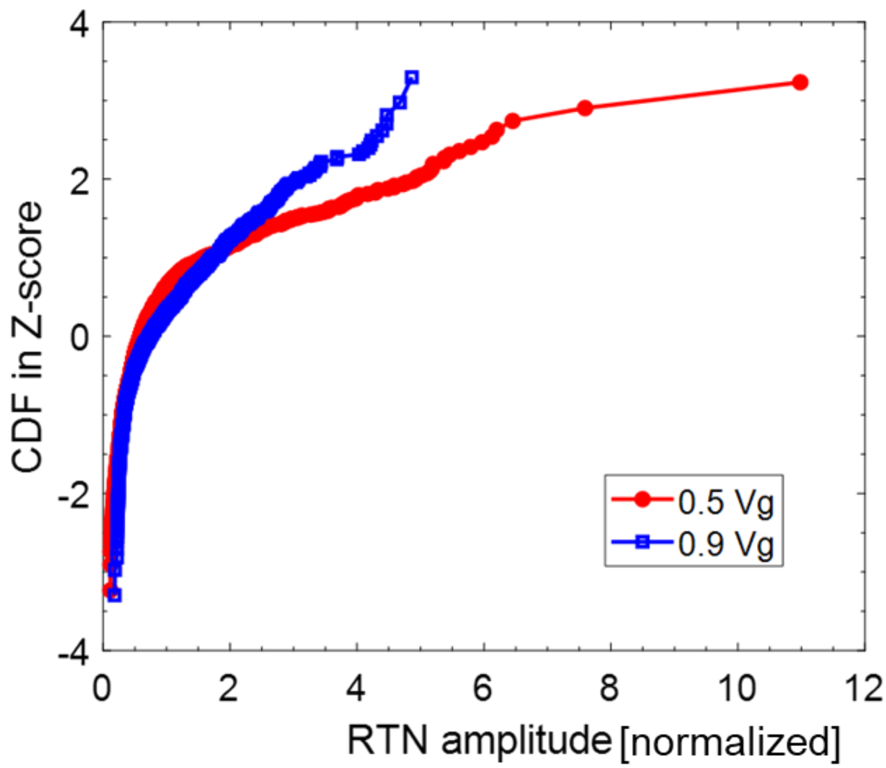


Figure 6.3.8 Comparison of RTN amplitude distribution at different gate voltage.

The amplitude distributions extracted from two different gate voltages are compared by taking their normalised values in figure 6.3.8. The amplitudes are normalised against the mean value of each data series. The result shows that the magnitude of RTN around threshold voltage has a longer tail than that under Vg=0.9 V. This agrees well with the 3D atomic simulation results reported by R. Wang [67].

6.4 Gate Voltage impact on RTN time constant distribution

Both amplitude and time constants are important parameters of RTN characterisation. An accurate evaluation of these parameters is essential for a proper modelling. These RTN parameters have wide distributions, because of the variation of trap positions. From the previous section, it was concluded that the statistical distribution of RTN amplitude is Vg dependent. This section will investigate the impact of Vg on RTN time constant distribution. As discussed in Chapter 4 , the capture emission time (CET) distribution of RTN is found to be uniformly distributed in Log-scale at Vg = 0.5V. It is known that the time constants of RTN are bias dependent [4] therefore the Vg impact on the statistical distribution of CETs of RTN will be studied in this section.

Following the same methodology used in Chapter 4 [132], the envelope (Env) of the RTN signal from 405 devices is extracted at Vg = 0.9V and temperature at 28ºC. The extracted Env data is then fitted with Log-uniform distribution to find the Vg impact on Env when 0.9Vg bias is applied. In Figure 6.4.1, it is depicted that the time constant distribution also follows the Log-uniform distribution under 0.9 V. Ideally further investigation should be done for different samples fabricated by different processes. This is out of the scope of this project.

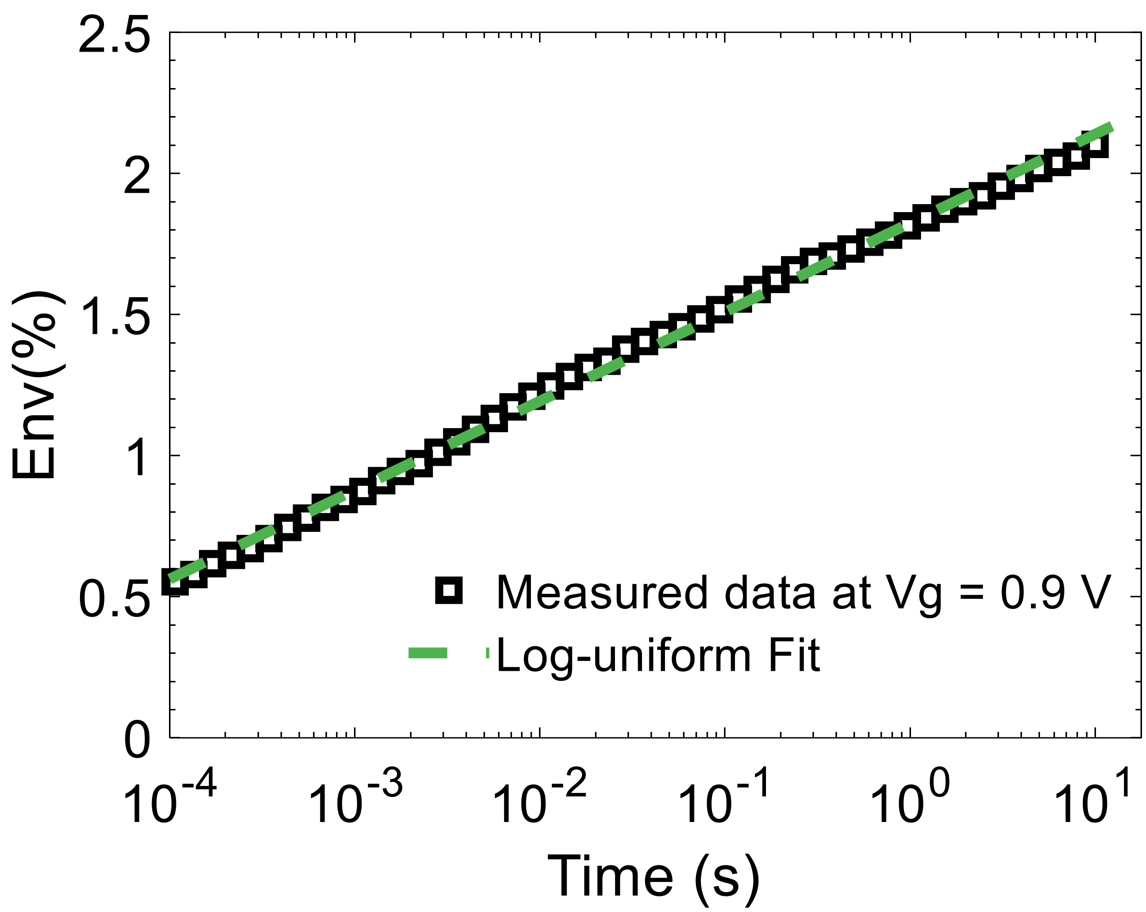


Figure 6.4.1 The evolution of envelopes with time. Symbols are experimental data and dashed lines are fitted with log-uniform distribution.

6.5 Summary

In this work, the impact of gate voltage on RTN amplitude and time constant distributions were assessed. The extraction of distribution function is important for the accurate prediction of RTN impact on circuits towards high-σ tail. Therefore, an operating voltage of 0.9V was applied at the gate and the corresponding RTN amplitude and CET distributions were monitored and compared with the distributions extracted from when a near threshold voltage of 0.5V was applied at the gate. It is concluded that RTN amplitude distribution is Vg dependent when Vg is increased from 0.5V to 0.9V. Different statistics should be applied under different gate biases, therefore. At 0.5Vg the RTN amplitude distribution follows GEV distribution, while the Lognormal distribution fits the data better under 0.9Vg. On the other hand, the time constant follows the Log-uniform distribution for both series of data extracted at 0.5Vg and 0.9Vg. However, the conclusion is drawn based on the current data and further investigation is required for its general applicability.

Chapter 7 Conclusions and Future perspective

The project is targeted at characterisation and modelling reliability and variability on modern nano-scale devices, as currently there is no complete and practically acceptable model or simulation tool available for nanometre transistors' RTN prediction. The project focuses on the instability induced by Random Telegraph Noise (RTN) in nano-scale CMOS devices and assesses its statistical properties to develop a model that can be used to predict long term RTN impact on real circuit designs in the future. This chapter will highlight the main objectives that have been delivered in this project to achieve the targeted aim. The potential work that could be undertaken in the future will also be discussed.

7.1 Conclusions on new RTN characterisation technique

Commercial devices are expected to perform at the rated operation parameters for at least 10 years, statistically. Based on the published results [140], [156], the experimental measurement time used for analysis is far away from 10 years. The results reported are mainly for less than 1 sec time windows with only a few works reporting between 1 sec and 20 sec. In this situation, the accuracy of the analysed results or statistical distribution of RTN within the provided time range are questionable. Due to this restraint, there is a desire for longer measurement beyond 20 sec or more. The main difficulty is the huge data size that is difficult to handle for the simulator and laboratory equipment, such as oscilloscopes which have a finite memory depth. For instance, 1us sampling rate in 1s will generate a data size of ~1M data. Another issue in the case of RTN measurements is the selection of devices. In previous works, they mainly used devices which have observable RTN. This does not represent the real devices well, as many devices do not have a clear RTN signal. Thus, the accuracy of the results again comes into question.

Therefore, the first major issue that has been addressed in this project is measuring long-term RTN from a large number of devices for statistical analysis. To do so, all devices were used without any device selection. This allows obtaining the trustworthy true statistical properties of the RTN instabilities. Secondly, a new technique was developed to record a longer time window by making the storage rate different from the sampling rate of the measured data. The longer measurement time used in this project is 10^5 sec which is nearly one day. With such a long time window and using 1 MS/sec sampling rate, the data size becomes hundreds of GBs for each device. Such a big data size is beyond the memory depth of modern oscilloscopes, so that the new method of recording data at different storage rates has been introduced in this project.

Two oscilloscopes were used to establish this technique: one records raw data of time window of 10 sec at every 0.1 sec as this is the minimum delay and saving time of the oscilloscope before its screen refreshes and allows the second lot of the data. Another oscilloscope records maximum and minimum points (referred to as envelopes) of the same signal at every 20 sec until the full desired time window is complete. The sampling rate of both the oscilloscopes is fixed at 1 MS/sec. From the obtained results, it has been observed that most of the time the envelope of the measurement signal remains constant hence other than recording every point the data were recorded every 20 seconds. The obtained results from the two oscilloscopes have good agreement and the method has been successfully introduced.

7.2 Conclusion on RTN distributions

The second major issue tackled in this project is the statistical distributions of both RTN amplitude and time constants. Time constants of RTN are the capture/emission times (CET) of traps within a device and the fluctuation in the drain current is its corresponding amplitude. As device size continues to decrease, proper modelling and characterising of RTN statistical properties becomes important due to a single trap having a larger impact. Several statistical distributions for amplitude and time constants of RTN have been proposed but the main issue of these results is the number of data points used for statistical distribution prediction. There are typically less than 200 data points [106] and that reduces the accuracy of the statistical distribution as the larger the number of the data points, the more accurate the statistical distribution will be.

At first, the statistical distribution of the capture and emission times of traps that are responsible for RTN is investigated. This work proposes an integrated methodology for extracting the statistical distribution of capture and emission time constants. Unlike the traditional modelling method, a top-down methodology is approached. The dynamic Monte Carlo simulation is used to confirm that the average envelope of RTN from multiple devices can uncover the underlying cumulative distribution of the CETs. The RTN envelopes were extracted from the overnight experimental data. The overnight RTN test was conducted by following the new RTN characterisation technique. By measuring the cumulative impact of traps on multiple devices against time window, the experimental results are used to extract the CDF of CETs. Afterwards, the average long-term RTN data is used to assess the CDFs proposed by early works. The power law distribution, which is most commonly used for ageing, was first considered. It is found that the results do not agree well with Power law and, moreover, it overestimates the long-term RTN. Next, the Log-normal distribution is used and it is found that it underestimates the long-term RTN data. Finally, Log-uniform distribution is used and the overnight experimental data supports the use of Log-uniform CDF for CETs. Although the test data within a short time window can be fitted well with both Log-uniform and Log-normal distributions, only the Log-uniform model correctly predicts the long-term RTN behaviour. The novelty of the work is that, for the first time, the long-term prediction capability of the extracted Log-uniform CDF is verified based on the experimental data. This will allow the assessment of RTN time constant distributions in a time scale of years.

Next, a comprehensive analysis on the accuracy in modelling the statistical distribution of RTN amplitude is undertaken with the new contribution of a distribution model and model selection criteria. The novelty of this work is proposing the new model selection criteria based on the relation between error and trap number. A new method for extracting trap amplitude from the measured raw data is also proposed. The cumulative distribution of trap amplitudes from multiple devices fitted with Maximum Likelihood Estimation (MLE) method. The fitted distribution should minimise the error not only for the whole data set but also in the distribution tail. It was discovered that Generalized Extreme Value (GEV) distribution has the least Z-score based sum-square-error (SSE). The new model selection criteria were then used to make sure the obtained results follow GEV distribution. The new method requires a monotonic error reduction with the increase in trap number. When this method is applied to the measured amplitude data, it is found that GEV meets this criterion well whereas other distributions like Exponential, Lognormal, and Gumbel distributions do not. To enhance the justification on the GEV model selection, bimodal Exponential and Lognormal CDFs are also used and it is found that they have a modest impact on the error even with more fitting parameters. A reference benchmark is also provided to access the impact of a limited number of traps on the CDF accuracy. For 90% confidence, the guide-band for RTN induced δV_{th} at 5σ should be increased roughly by 26.5% from the value estimated by the statistical distribution based on 1,000 traps. It has been shown that the uncertainty caused by using a limited number of traps is substantially smaller and concluded that the uncertainty in RTN amplitude prediction and modelling are dominated by CDF model selection.

RTN is a stochastic noise occurring when traps capture and emit electrons or holes. The voltage and temperature applied to MOSFETs are directly affecting the behaviour of RTN. The final work of the project is based on the impact of gate voltage on RTN amplitude and CET distribution. From the start of the project, all the experiments have been conducted under gate voltage close to threshold voltage of 0.5 Vg and at the end of the project an operating voltage of 0.9 Vg is applied to observe the impact of Vg on the distribution function at 28ºC. The same method and procedure for RTN amplitude and CET extraction are used as before. Based on the obtained results, it is concluded that RTN amplitude distribution is Vg dependent. In this case, the RTN amplitude distribution at 0.9 Vg fits better with Lognormal instead of GEV in terms of the least Z-score based error. However, the accuracy of the Lognormal CDF does not confirm the new model selection criteria and further work needs to be done. The reason behind the difference between the different statistics at different gate voltage is due to the narrower percolation path under 0.5 Vg in comparison with 0.9 Vg. As a narrower percolation path increases the single trap impact, the RTN amplitude distribution for 0.5 Vg has a longer tail than that at 0.9 Vg. In contrast, the RTN time constant distribution is shown to have no dependence on Vg, as in both bias conditions the distribution follows Log-uniform behaviour. The conclusion was drawn based on the collected data and due to the time limit of the project and its general applicability awaits further work.

7.3 Future perspectives

Before listing the suggestions for future work, it is important to address some of the challenging tasks in the research. The continuous RTN data acquisition has instrumental and experimental limitations, long-term RTN measurement becomes too time consuming, the limited available number of devices for statistical analysis and minimising system noise were the most challenging parts of the work. The experimental environment and the results can be improved if all the above can be optimised. While this thesis has highlighted a successful approach of systematic RTN overnight test, new RTN distribution models and long-term RTN prediction, there are several parts that need to be improved or addressed.

Firstly, the proper continuous RTN data acquisition is needed so that no information is lost. In this work, due to oscilloscope slow processing and memory limitation the test pattern was designed with two oscilloscopes and signal envelopes were used for overnight data. This limitation can potentially be further improved by looking into some potential solutions such as continuous Data

Acquisition Cards (DAC) platform. The quickest and the easiest way would be using a DAC for recording continuous data directly to an external hard drive. Once the DAC is connected with the PC or laptop, then the integrated hardware could be either controlled by existing code or by developing new LABVIEW code. This would improve the continuous RTN data measurement. The next issue would be addressing methods to simplify the huge data analysing procedure.

The experimental data show that the drain current fluctuates both below and above the initial value of Id, potentially containing both acceptor and donor traps in the device. The collected data at a certain time is mixed with negative and positive values. For accurate RTN modelling, both the negative and positive data should be covered. One potential approach could be assigning acceptor and donor Id according to their measured ratio and then the experimental data will be separated into two parts: Id above 0% and Id below 0% as a starting point. That means separate modelling for acceptor and donor traps of RTN.

This work was done under RTN DC conditions. Early works also typically measure the RTN under DC conditions for convenience, but the real circuits typically operate in AC conditions. Under DC conditions, the RTN is dominated by the traps located close to the Fermi-level, Ef, at the oxide/Silicon interface, as the traps below Ef are steadily filled. Under AC conditions, however, they can be empty and contribute to the instabilities. The differences in RTN between DC and AC operation have not been characterised in this project and are not well understood, thus opening another area for investigation.

Due to the limited number of available devices, the conclusion drawn in this project is only based on nMOS devices, hence further extension of the work could be done with pMOS devices and the comparison between them would be interesting.

RTN parameters are voltage and temperature dependent. However, in this work gate voltage has been varied only at two values: 0.5 V which is near to threshold voltage and 0.9 V considered as operating voltage. The extracted RTN amplitude distribution is found to be different at different biases, but the model selected for higher Vg does not agree with all model selection criteria introduced for the work. Therefore, further work needs to be done to gain confidence in the accuracy of the distribution model when bias is varied.

Moreover, in this paper several statistical distributions of single trap RTN amplitude are compared with experimental data on a single type (as per technology and size) of transistor, and it is concluded that GEV distribution fits the best when biased at 0.5 Vg. Since this distribution was chosen without any link to the underlying RTN physical mechanism, there is no guarantee that it is always good for other MOSFETs fabricated by other processes. The general applicability of the conclusions drawn in this project awaits further verification.

The most promising extension of this work would be developing a model that could provide the probability of RTN occurrence in different states or levels that could impact the performance of the designed circuit. Based on the experimental results, an industry-standard model would be developed to predict long term RTN probability in circuits.

References

[1]. T. Grasser, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, M. T. Luque, "The paradigm shift in understanding the bias temperature instability: from reaction–diffusion to switching oxide traps", *IEEE Electronic Device*, Transactions on, 58(11), 3652-3666, 2011.

[2]. B. Kaczer, *et al.*, " Atomistic approach to variability of bias temperature instability in circuit simulations*", IEEE international*, 2011.

[3]. K. S. Ralls *et al.*, "Discrete resistance switching in sub-micrometer inversion layers: individual interface traps and low-frequency (1/f ?) noise", *Phys. Rev. Lett*. 52 228–31, 1984.

[4]. M. Kirton and M. Uren, "Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/*f*) noise," *Advances in Physics*, vol. 38, no. 4, pp. 367-468, 1989.

[5]. K. Hung, P. Ko, C. Hu, and Y. C. Cheng, "Random telegraph noise of deep-submicrometer MOSFETs," *IEEE Electron Device Lett.,* vol. 11, no. 2, pp. 90-92, 1990.

[6]. S. Realov, and K. L. Shepard, "Analysis of random telegraph noise in 45-nm CMOS using on-chip characterization system," *IEEE Trans. Electron Dev.*, vol. 60, no. 5, pp. 1716 – 1722, 2013.

[7]. A. Ghetti, C. M. Compagnoni, F. Biancardi, A. L. Lacaita, S. Beltrami, L. Chiavarone, A. S. Spinelli, and A. Visconti, "Scaling trends for random telegraph noise in deca-nanometer flash memories," *IEDM*, pp. 835-838, 2008.

[8]. K. Zhang, Embedded Memories for Nano-Scale VLSIs, Integrated Circuits and Systems Series, Springer, 2009.

[9]. ITRS 2009 web site: http://www.itrs.net/Links/2009ITRS/Home2009.htm

[10]. C. Auth et al., "A 22 nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors," *in VLSI Symp. Tech. Dig*., pp. 12–14, 2012.

[11]. N. Tega, "Study on Impact of Random Telegraph Noise on Scaled MOSFETs", *Degree of Doctor of Philosophy*, University of Tsukuba, 2014.

[12]. J. F. Zhang, "Traps", *Wiley Encyclopedia of Electrical and Electronics Engineering* (1999): 1-10.

[13]. B. Kaczer, T. Grasser, P. J. Roussel, J. Franco, R. Degraeve, L. A. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger, "Origin of NBTI variability in deeply scaled pFETs," *Proc. Int. Rel. Phys. Symp.,* pp. 26-32, 2010.

[14]. Terman, L. M., "An investigation of surface states at a silicon/silicon oxide interface employing metal-oxide-silicon diodes", *Solid-State Electronics*, 5(5), 285-299, 1962.

[15]. C. Q. Wei, Y. Z. Xiong, X. Zhou, and L. Chan, "A technique for constructing RTS noise model based on statistical analysis", *NSTI-Nanotech,* vol. 3, p. 885, 2008.

[16]. J. M Woo *et al.*, "Statistical analysis of random telegraph noise in CMOS image sensors. In Simulation of Semiconductor Processes and Devices", *SISPAD,* International Conference on, pp. 77 –80, 9-11, 2008.

[17]. H. Kurata, "Random telegraph signal in flash memory: Its impact on scaling of multilevel flash memory beyond the 90-nm node. Solid-State Circuits*, IEEE Journal of,* vol. 42, No. 6, pp. 1362–1369, June, 2007.

[18]. G. I. Wirth, R. D. Silva, and R. Brederlow, "Statistical model for the circuit bandwidth dependence of low-frequency noise in deep-submicrometer MOSFETS," *IEEE Trans. Electron Dev.*, vol. 54, no. 2, p. 340, 2007.

[19]. E. Simoen, J. W. Lee, and C. Claeys, "Assessment of the impact of inelastic tunnelling on the frequency-depth conversion from low-frequency noise spectra*," IEEE Trans. Electron Dev.*, vol. 61, no. 2, p. 634, 2014.

[20]. H. Miki *et al*., "Understanding short-term BTI behavior through comprehensive observation of gate-voltage dependence of RTN in highly scaled high-k/metal-gate pFETs," *in Symp. VLSI Technol.-Dig. Technol.,* Jun. 2011, pp. 148–149.

[21]. T. Nagumo, K. Takeuchi, S. Yokogawa, K. Imai, and Y. Hayashi, "New analysis methods for comprehensive understanding of random telegraph noise," *IEDM,* pp. 32.1.1-32.1.4, 2009.

[22]. Y. Ye, C. C. Wang, and Y. Cao, "Simulation of random telegraph noise with 2-stage equivalent circuit," *ICCAD*, pp. 709 – 713, 2010.

[23]. R. Landauer, "The noise is the signal," Nature, vol. 392, pp. 658-659, 1998.

[24]. J. A. Connelly, "Low-Noise Electronic System Design", *John Wiley & Sons, Inc.*, 1993.

[25]. J. B. Johnson, “Thermal Agitation of Electricity in Conductors,” *Phys. Rev.*, vol. 32, pp. 97-109, 1928.
[26]. H. Nyquist, “Thermal Agitation of Electric Charge in Conductors,” *Phys. Rev.*, vol. 32, pp. 110-113, 1928.
[27]. B. L. Kish, "End of Moore's law: thermal (noise) death of integration in micro and nano electronics." *Physics Letters A 305*, no. 3-4 (2002): 144-149.
[28]. W. M. Leach, “Fundamentals of low-noise analog circuit design,” Proc. IEEE , vol. 82, no. 10, p. 1515, 1994.
[29]. M. V. Haartman *et al.*, “Low-frequency noise in advance MOS devices”, *Netherland, Springer*, 2007.
[30]. J. Lee *et al.*, “Noise model of gate-leakage current in ultrathin oxide MOSFETs,” *IEEE Trans. Electron Dev.*, vol. 50, pp. 2499 - 2506, 2003.
[31]. C. Fiegna, “Analysis of gate shot noise in MOSFETs with ultrathin gate oxides,” *IEEE Trans. Electron Dev.*, vol. 24, pp. 108-110, 2003.
[32]. D. K. Schroder, “Semiconductor Material and Devices Characterization”, *John Wiley & Sons Inc.*, 3rd edition, 2006.
[33]. A. van der Ziel, “Noise in solid state devices and circuits”, *John Wiley & Sons*, 1986.
[34]. Christensson, S., I. Lundström, and C. Svensson. "Low frequency noise in MOS transistors—I theory." Solid-State Electronics 11, no. 9 (1968): 797-812.
[35]. A.L McWhorter, "1/f noise and germanium surface properties,” Semiconductor Surface Physics, Philadelphia: University of Pennsylvania Press 1057, pp.207, 1957.
[36]. Vandamme, L. K. J. "Model for 1/f; noise in MOS transistors biased in the linear region." Solid-State Electronics 23, no. 4 (1980): 317-323.
[37]. F. N. Hooge, “1/f noise is no surface effect”, *Physics Letters*, 29A (3), 1969, 139–140
[38]. S. Machlup, “Noise in semiconductors: spectrum of a two-parameter random signal,” *J. Appl. Phys*., vol. 25, pp. 341-343, 1954.
[39]. M.J Buckingham, “Noise in electronic devices and systems”, *Horwood*, 1983.
[40]. K. Kandiah, and F.B. Whiting, “Low frequency noise in junction field effect transistors,” *J. Appl. Phys*., vol. 66, pp. 937-948, 1989.
[41]. J. Martín-Martinez *et al.*, "Statistical characterization and modeling of random telegraph noise effects in 65nm SRAMs cells," *IEEE 14th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*, pp. 1-4, 2017.
[42]. K. V. Aadithya *et al.*, "Accurate prediction of random telegraph noise effects in SRAMs and DRAMs," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 32*, no. 1, pp. 73-86, 2012.
[43]. D. Veksler *et al.*, "Random telegraph noise (RTN) in scaled RRAM devices," *IEEE International Reliability Physics Symposium (IRPS)*, pp. MY-10, 2013.
[44]. T. Matsumoto, K. Kobayashi, and H. Onodera, “Impact of random telegraph noise on CMOS logic circuit reliability,” *in Custom Integrated Circuit Conference*, pp.14–4, Sept. 2014.
[45]. C. Claeys *et al.*, "Random telegraph signal noise in advanced high performance and memory devices," *IEEE 31st Symposium on Microelectronics Technology and Devices (SBMicro)*, pp.1-6, 2016.
[46]. G. Ghibaudo, & T. Boutchacha, “Electrical noise and RTS fluctuations in advanced CMOS devices”, *Microelectronics Reliability*, 42(4), 573-582, 2002.
[47]. Tsai, & Ma, “The impact of device scaling on the current fluctuations in MOSFET's”, IEEE Electron Devices Transactions on, 41(11), 2061-2068, 1994.
[48]. Kirton, M. J., M. J. Uren, S. Collins, M. Schulz, A. Karmann, and K. Scheffer. "Individual defects at the Si: SiO2 interface." *Semiconductor science and technology 4*, no. 12,1116, (1989).
[49]. S. Dongaonkar *et al*., "Random telegraph noise (RTN) in 14nm logic technology: High volume data extraction and analysis," *IEEE Symposium on VLSI Technology*, pp. 1-2. IEEE, 2016.
[50]. Ma, Jigang, *et al.* "Investigation of preexisting and generated defects in nonfilamentary a-Si/TiO 2 RRAM and their impacts on RTN amplitude distribution," *IEEE Transactions on Electron Devices*, 65.3 (2018): 970-977.
[51]. R. Gao, J. Zhigang, S. M. Hatta, J. F. Zhang, J. Franco, Ben Kaczer, W. Zhang *et al*., "Predictive As-grown-Generation (AG) model for BTI-induced device/circuit level variations in nanoscale technology nodes," *IEEE International Electron Devices Meeting (IEDM)*, pp. 31-4. IEEE, 2016.

[52]. A. Neugroschel *et al*., "Applications of DCIV method to NBTI characterization." *Microelectronics Reliability* 47, no. 9-11 (2007): 1366-1372.
[53]. T. Grasser, *et al.* ,"The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability," *IEEE International Reliability Physics Symposium,* 2010.
[54]. T. Grasser, *et al.*, "Gate-sided hydrogen release as the origin of" permanent" NBTI degradation: From single defects to lifetimes," *IEEE International Electron Devices Meeting (IEDM),* pp. 20-1., 2015.
[55]. J. Brown et al, "A low-power and high-speed True Random Number Generator using generated RTN," in Symp. VLSI Technol.-Dig. Technol., pp. 95-96, 2018.
[56]. T. Grasser, *et al.,* "A two-stage model for negative bias temperature instability," *IEEE international reliability physics symposium (IRPS)*, pp. 33-44., 2009.
[57]. R. Gao, "Bias Temperature Instability Modelling and Lifetime Prediction on Nano-scale MOSFETs", PhD thesis, *Liverpool John Moore's University*, July 2018.
[58]. K. Kobayashi, *et al.*, "Random Telegraph Noise Under Switching Operation." *Noise in Nanoscale Semiconductor Devices*, pp. 285-333, Springer, Cham, 2020.
[59]. E. Simoen, and C. L. Claeys, "Random telegraph signals in semiconductor devices," vol.357. *Bristol: IOP Publishing,* 2016.
[60]. K. S. Ralls, et al. "Discrete resistance switching in submicrometer silicon inversion layers: Individual interface traps and low-frequency (1 f?) noise*," Physical review letters 52.3* (1984): 228.
[61]. Z. Shi, J. P. Mieville, and Michel Dutoit, "Random telegraph signals in deep submicron n-MOSFET's," *IEEE Transactions on Electron Devices*, 41, no. 7.,1161-1168, 1994.
[62]. P. Ren, et al., "New observations on complex RTN in scaled high-κ/metal-gate MOSFETs-The role of defect coupling under DC/AC condition" *IEDM* ,p. 778-781, 2013.
[63]. N. Tega, et al., "Increasing threshold voltage variation due to random telegraph noise in FETs as gate lengths scale to 20 nm" *VLSI*, p. 50-51, 2009.
[64]. Z. Zhang *et al.*, "Investigation on the amplitude distribution of random telegraph noise (RTN) in nanoscale MOS devices." *IEEE International Nanoelectronics Conference (INEC),* pp. 1-2, 2016.
[65]. K. K. Hung, *et al.*, "A unified model for the flicker noise in metal-oxide-semiconductor field-effect transistors." IEEE Transactions on Electron Devices 37.3 (1990): 654-665.
[66]. M. Schulz, and A. Karmann, "Individual, attractive defect centers in the SiO 2-Si interface of μm-sized MOSFETs," *Applied Physics A 52*, no. 2.,104-111, 1991.
[67]. R. Wang et al., "Too Noisy at the Bottom? -Random Telegraph Noise (RTN) in Advanced Logic Devices and Circuits," in IEDM Tech. Dig., Dec. 2018, pp. 388–391, doi: 10.1109/IEDM.2018.8614594.
[68]. K. Fukuda et al, "Random Telegraph Noise in Flash Memories－Model and Technology Scaling", in IEDM Tech. Dig., Dec. 2007, pp. 169–172, doi: 10.1109/IEDM.2007.4418893.
[69]. M. Tanizawa et al, "Application of a Statistical Compact Model for Random Telegraph Noise to Scaled-SRAM Vmin Analysis," in Proc. Symp. VLSI Technol., Jun. 2010, pp. 95–96, doi: 10.1109/VLSIT.2010.5556184.
[70]. Z. Zhang et al, "New Insights into the Amplitude of Random Telegraph Noise in Nanoscale MOS Devices," in Proc. Int. Rel. Phys. Symp., April 2017, pp.3C-3.1 – 3c-3.5, doi: 10.1109/IRPS.2017.7936288.
[71]. E. Simoen, W. Fang, M. Aoulaiche, J. Luo, C. Zhao, and C. Claeys. "Random telegraph noise: The key to single defect studies in nano-devices," *Thin Solid Films 613*, pp.2-5., 2016.
[72]. T. Grasser, "Noise in nanoscale semiconductor devices", *Springer Science & Business Media*, 2020.
[73]. Francesco Maria Puglisi, Andrea Padovani, Luca Larcher, and Paolo Pavan. "Random telegraph noise: Measurement, data analysis, and interpretation." *IEEE 24th International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, pp. 1-9. IEEE, 2017.
[74]. L.E. Baum, T. Petrie, G. Soules, N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains", *Ann. Math. Stat*. 41(1), 164–171, 1970.
[75]. A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm", *IEEE Trans. Inf. Theory*, 13(2), 260–269 (1967).

[76]. X. Chen, Y. Wang, Y. Cao, and H. Yang, "Statistical analysis of random telegraph noise in digital circuits." *19th Asia and South Pacific Design Automation Conference (ASP-DAC)*, pp. 161-166. IEEE, 2014.
[77]. F.M. Puglisi, P. Pavan, "Factorial hidden Markov model analysis of random telegraph noise in resistive random access memories", *ECTI Trans. Electr. Eng. Electr. Commun.* 12(1), 24–29, 2014.
[78]. Z. Ghahramani, Jordan, Factorial Hidden Markov Models. M.I. Machine Learn. 29, 245,1997.
[79]. H. Awano et al., "Multi-Trap RTN Parameter Extraction Based on Bayesian Inference," *14th International Symposium on Quality Electronic Design (ISQED)*, pp. 597–602, 4–6, 2013.
[80]. S. Chitrashekaraiah, et al., "Addressable test structures for MOSFET variability analysis", IEEE International Conference on Microelectronic Test Structures, pp. 31-35. IEEE, 2012.
[81]. Datasheet: "Keysight N6780 Series Source/Measure Units (SMUs) for the N6700 Modular Power System".
[82]. Ortiz-Conde, F. G. Sánchez, J. J. Liou, A. Cerdeira, M. Estrada, and Y. Yue, "A review of recent MOSFET threshold voltage extraction methods," Microelectron. Rel., vol. 42, no. 4, pp. 583-596, 2002.
[83]. G. Ghibaudo,"New method for the extraction of MOSFET parameter," Electronic Letters, vol.24, pp. 543-545, 1988.
[84]. M. Duan, J. F. Zhang, Z. Ji, W. D. Zhang, B. Kaczer, S. De Gendt, and G. Groeseneken, "New insights into defect loss, slowdown, and device lifetime enhancement," IEEE Trans. Electron Devices, vol. 60, no. 1, pp. 413-419, 2013.
[85]. M. Duan, J. F. Zhang, Z. Ji, W. D. Zhang, B. Kaczer, T. Schram, R. Ritzenthaler, G. Groeseneken, and A. Asenov, "New Analysis Method for Time-Dependent Device-To-Device Variation Accounting for Within-Device Fluctuation," IEEE Trans. Electron Devices, vol. 60, no. 8, pp. 2505-2511, 2013.
[86]. M. Duan, J. F. Zhang, Z. Ji, W. Zhang, Ben Kaczer, Tom Schram, Romain Ritzenthaler, Aaron Thean, Guido Groeseneken, and A. Asenov. "Time-dependent variation: A new defect-based prediction methodology." Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers, pp. 1-2. IEEE, 2014.
[87]. J. F. Zhang, et al., "DEFECTS FOR RANDOM TELEGRAPH NOISE AND NEGATIVE BIAS TEMPERATURE INSTABILITY", China Semiconductor Technology International Conference (CSTIC), ShangHai, March 13-14, 2016.
[88]. T. Nagumo et al, "Statistical Characterization of Trap Position, Energy, Amplitude and Time Constants by RTN Measurement of Multiple Individual Traps," IEDM Tech. Dig., Dec. 2010, pp. 628–631, doi: 10.1109/IEDM.2010.5703437.
[89]. R. Gao, et al., "As-grown-Generation Model for Positive Bias Temperature Instability," IEEE Trans. Electron Devices, vol. 65, no.9, pp.3662-3668, 2018, doi:: 10.1109/TED.2018.2857000.
[90]. M. Duan et al., "Key issues and techniques for characterizing Time dependent Device-to-Device Variation of SRAM," IEDM Tech. Dig., Dec. 2013, pp. 774–777, doi: 10.1109/IEDM.2013.6724730.
[91]. Kerber, A., Cartier, E., Pantisano, L., Rosmeulen, M., Degraeve, R., Kauerauf, T. & Schwalke, U. (2003) "Characterization of the Vt-instability un SiO2 HFO2 gate dielectrics" status: published, 41-45.
[92]. J. F. Zhang, Z. Ji, M. H. Chang, B. Kaczer, and G. Groeseneken, "Real Vth instability of pMOSFETs under practical operation conditions." p. 817.
[93]. Z. Ji, J. F. Zhang, M. H. Chang, B. Kaczer, and G. Groeseneken, "An Analysis of the NBTI-Induced Threshold Voltage Shift Evaluated by Different Techniques," IEEE Trans. Electron Devices, vol. 56, no. 5, pp.1086-1093, May, 2009.
[94]. Z. Ji, L. Lin, J. F. Zhang, B. Kaczer, and G. Groeseneken, "NBTI Lifetime Prediction and Kinetics at Operation Bias Based on Ultrafast Pulse Measurement," IEEE Trans. Electron Devices, vol. 57, no. 1, pp. 228-237, Jan, 2010.
[95]. A. Bal et al, "Trident: Comprehensive Choke Error Mitigation in NTC Systems," IEEE Trans. VLSI, vol. 26, no. 11, pp. 2195-2204, 2018, doi: 10.1109/TVLSI.2018.2863954.
[96]. A. Asenov et al, "RTS amplitudes in decananometer MOSFETs: A 3D simulation study," IEEE Trans. Electron Dev., vol. 50, no. 3, pp.839-845, 2003, doi: 10.1109/TED.2003.811418.
[97]. T. Grasser, "Stochastic charge trapping in oxides: From random telegraph noise to bias temperature instabilities," Microelectronics Rel., vol. 52, pp. 39-70, 2012, doi:10.1016/j.microrel.2011.09.002.

[98]. A. Manut et al, “Trigger-When-Charged: A Technique for Directly Measuring RTN and BTI-Induced Threshold Voltage Fluctuation Under Use-Vdd,” IEEE Trans. Electron Devices, vol. 66, no.3, pp.1482-1488, 2019, doi: 10.1109/TED.2019.2895700.

[99]. M. Nour et.al, “A stand-alone, physics-based, measurement-driven model and simulation tool for random telegraph signals originating from experimentally identified MOS gate-oxide defects,” IEEE Trans. Electron Devices, vol. 63, no. 4, pp. 1428–1436, Apr. 2016, doi: 10.1109/TED.2016.2528218.

[100]. A. Kerber, “Impact of RTN on stochastic BTI degradation in scaled Metal Gate / High-k CMOS technologies,” in Proc. Int. Rel. Phys. Symp., June 2015, pp.3B.3.1-3B.3.6, doi:10.1109/IRPS.2015.7112704.

[101]. E. Simoen, et al, “From a Device Physicist’s Dream to a Designer’s Nightmare,” ECS Transactions, vol. 39, no.1, pp.3-15, 2011, doi:10.1149/1.3615171.

[102]. H. Kurata et al, “Random Telegraph Signal in Flash Memory: Its Impact on Scaling of Multilevel Flash Memory Beyond the 90-nm Node,” IEEE J. Solid-State Cir., vol. 42, no. 6, pp. 1362–1369, June 2007, doi: 10.1109/JSSC.2007.897158.

[103]. Z. F. Ma et al, “A percolation study of RTS noise in deep submicron MOSFET by Monte Carlo simulation,” Chinese Phys. vol. 14, no. 4, pp. 808-811, 2005.

[104]. M. Luo et al, “Impacts of Random Telegraph Noise (RTN) on Digital Circuits,” IEEE Trans. Electron Devices, vol. 62, no. 6, pp.1725–1732, Jun. 2015, doi: 10.1109/TED.2014.2368191.

[105]. G. I. Wirth et al, “Statistical Model for MOSFET Bias Temperature Instability Component Due to Charge Trapping,” IEEE Trans. Electron Devices, vol. 58, no. 8, pp. 2743–2751, Aug. 2011, doi:10.1109/TED.2011.2157828.

[106]. K. Ito et al, “Modeling of Random Telegraph Noise under Circuit Operation-Simulation and Measurement of RTN-induced delay fluctuation,” in Proc. 12th Int'l Symp. on Quality Electronic Design, pp. 22-27, 2011, doi: 10.1109/ISQED.2011.5770698.

[107]. M. Tanizawa et al, “Application of a Statistical Compact Model for Random Telegraph Noise to Scaled-SRAM Vmin Analysis,” in Proc. Symp. VLSI Technol., Jun. 2010, pp. 95–96, doi: 10.1109/VLSIT.2010.5556184.

[108]. L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” Proc. IEEE, vol. 77, no. 2, pp. 257–286, 1989, doi: 10.1109/5.18626.

[109]. F. M. Puglisi et al., “A complete statistical investigation of RTN in HfO2-based RRAM in high resistive state,” IEEE Trans. Elec. Dev., vol. 62, no. 8, pp. 2606-2613, 2015, doi: 10.1109/TED.2015.2439812.

[110]. T. H. Both et al., “Modeling and simulation of the charge trapping component of BTI and RTS,” Microelectronics Rel., vol. 80, pp. 278-283, 2018, doi:10.1016/j.microrel.2017.11.009.

[111]. F. M. Puglisi, et al., “Monitoring stress-induced defects in HK/MG FinFETs using random telegraph noise,” IEEE Elec. Dev. Lett., vol. 37, no. 9, pp. 1211-1214, 2016, doi: 10.1109/LED.2016.2590883.

[112]. K. Takeuchi, et al.,“Single-Charge-Based Modeling of Transistor Characteristics Fluctuations Based on Statistical Measurement of RTN Amplitude,” in Symp. VLSI Technol.-Dig. Technol., pp. 54-55, 2009.

[113]. F. M. Puglisi, et al., “RTS noise characterization of HfOx RRAM in high resistive state,” Solid-State Electronics, vol. 84, pp: 160-166, 2013, doi: 10.1016/j.sse.2013.02.023.

[114]. G. Nicosia, et al.. “Investigation of the temperature dependence of random telegraph noise fluctuations in nanoscale polysilicon-channel 3-D Flash cells,” Solid-State Electronics, vol.151, pp. 18-22, 2019, doi: 10.1016/j.sse.2018.10.010.

[115]. Z. Ji et al., “A test-proven As-grown-Generation (A-G) model for predicting NBTI under use-bias,” Proc. Symp. VLSI Technol., pp. 36-37, Kyoto, Jun. 2015, doi:10.1109/VLSIT.2015.7223693.

[116]. J. F. Zhang, Z. Ji, and W. Zhang, “As-grown-generation (AG) model of NBTI: A shift from fitting test data to prediction,” Microelectronics Rel., vol. 80, pp. 109–123, Jan. 2018, doi:10.1016/j.microrel.2017.11.026.

[117]. K. Ota et al, “Experimental Study of Random Telegraph Noise in Trigate Nanowire MOSFETs,” IEEE Trans. Elec. Dev., vol. 62, no.11, p.3799-3804, 2015, doi: 10.1109/TED.2015.2471840.

[118]. B. S. Jo et al, "Characterization of Random Telegraph Noise Generated by Process- and Cycling-Stress-Induced Traps in 26 nm NAND Flash Memory," J. J. Appl. Phys., vol. 52, 04CA07, 2013, doi: 10.7567/JJAP.52.04CA07.

[119]. JEDEC, "Procedure for Wafer-Level DC Characterization of Bias Temperature Instabilities," JESD241, 2015.

[120]. R. Gao et al, "NBTI-Generated Defects in Nanoscaled Devices: Fast Characterization Methodology and Modeling," IEEE Trans. Electron Devices, vol. 64, no. 10, pp.4011-4017, 2017, doi:10.1109/TED.2017.2742700.

[121]. J. F. Zhang, "Oxide defects," Chapter 10 in Bias Temperature Instabilities for Devices and Circuits, pp. 253-285, Springer, 2014, doi: 10.1007/978-1-4614-7909-3.

[122]. M. Duan et al, "Key issues and solutions for characterizing hot carrier aging of nano-meter scale nMOSFETs," IEEE Trans. Electron Devices, vol. 64, no. 6, pp.2478-2484, 2017, doi:10.1109/TED.2017.2691008.

[123]. M. Duan et al., "Insight into electron traps and their energy distribution under positive bias temperature stress and hot carrier aging," IEEE Trans. Electron Devices, vol. 63, no. 9, pp. 3642–3648, Sep. 2016, doi: 10.1109/TED.2016.2590946.

[124]. J. F. Zhang et. al, "A comparative study of the electron trapping and thermal detrapping in SiO2 prepared by plasma and thermal oxidation, J. Appl. Phys., vol.72, pp. 1429-1435, 1992, doi:10.1063/1.351703.

[125]. J. F. Zhang et al, "Degradation of Oxides and Oxynitrides Under Hot Hole Stress," IEEE Trans. Electron Devices, vol. 47, no. 2, pp.378 - 386, 2000, doi: 10.1109/16.822284.

[126]. X. F. Zheng et al, "A New Multipulse Technique for Probing Electron Trap Energy Distribution in High-κ Materials for Flash Memory Application," IEEE Trans. Electron Devices, vol. 57, no. 10, pp. 2484-2492, 2010, doi:10.1109/TED.2010.2062520.

[127]. K. Sonoda, K. Ishikawa, T. Eimori, and O. Tsuchiya, "Discrete Dopant Effects on Statistical Variation of Random Telegraph Signal Magnitude," IEEE Trans. Electron Devices, vol. 54, no. 8, pp.1981-1925, 2017, doi:10.1109/TED.2007.900684.

[128]. B. Zimmer, O. Thomas, S. O. Toh, T. Vincent, K. Asanovic, and B. Nikolic, "Joint Impact of Random Variations and RTN on Dynamic Writeability in 28nm Bulk and FDSOI SRAM," in Proc. European Solid State Device Research Conference (ESSDERC), Sept. 2014, pp.98-101, doi:10.1109/ESSDERC.2014.6948767.

[129]. C. Y. P. Chao, H. Tu, T. M. H. Wu, K. Y. Chou, S. F. Yeh, C. Yin, and C. L. Lee, "Statistical Analysis of the Random Telegraph Noise in a 1.1 μm Pixel, 8.3 MP CMOS Image Sensor Using On-Chip Time Constant Extraction Method," Sensors, vol. 17, p.2704, 2017, doi: 10.3390/s17122704.

[130]. S. Ichino, T. Mawaki, A. Teramoto, R. Kuroda, S. Wakashima, T. Suwa, and S. Sugawa, "Statistical Analyses of Random Telegraph Noise in Pixel Source Follower with Various Gate Shapes in CMOS Image Sensor," ITE Trans. on Media Technology and Applications, vol. 6, no. 3, pp.163-170, 2018, doi: 10.3169/mta.6.163.

[131]. A.K.M. M. Islam and H. Onodera, "Worst-case Performance Analysis Under Random Telegraph Noise Induced Threshold Voltage Variability," in Proc. 28th Int. Symp. Power and Timing Modeling, Optimization and Simulation (PATMOS), pp.140-146, 2018, doi: 10.1109/PATMOS.2018.8464147.

[132]. M. Mehedi, K. H. Tok, J. F. ZHANG, Z. JI, Z. Ye, W. Zhang, and J. S. Marsland, "An assessment of the statistical distribution of Random Telegraph Noise Time Constants," IEEE Access, vol. 8, no. 10, pp.1496-1499, 2020, doi: 10.1109/ACCESS.2020.3028747.

[133]. Z. Çelik-Butler, S. P. Devireddy, H. Tseng, P. Tobin, and A. Zlotnicka, "A low-frequency noise model for advanced gate-stack MOSFETs," Microelectron. Rel., vol. 49 pp. 103–112, 2009, doii:10.1016/j.microrel.2008.12.005.

[134]. H. Qiu, K. Takeuchi, T. Mizutani, T. Saraya, J. Chen, M. Kobayashi, and T. Hiramoto, "Statistical Analyses of Random Telegraph Noise Amplitude in Ultra-Narrow (Deep Sub-10nm) Silicon Nanowire Transistors," in Proc. Symp. VLSI Technol., Jun. 2017, pp. 50–51.

[135]. M. Toledano-Luque, B. Kaczer, E. Simoen, Ph. J. Roussel, A. Veloso, T. Grasser, and G. Groeseneken, "Temperature and voltage dependences of the capture and emission times of individual traps in high-k dielectrics," Microelectronic Eng., vol. 88, pp. 1243-1246, 2011, doi: doi:10.1016/j.mee.2011.03.097.

[136]. D. McFadden, “Modeling the Choice of Residential Location,” Transportation Research Record, vol. 673, pp. 72–77, 1978.

[137]. T. Grasser, K. Rott, H. Reisinger, M. Waltl, J. Franco, and B. Kaczer, “A Unified Perspective of RTN and BTI,” in IEEE Proc. Int. Rel. Phys. Symp. (IRPS), April 2014, pp.4A.5.1-4A.5.7, doi: 10.1109/IRPS.2014.6860643.

[138]. X. Zhan, C. Shen, Z. Ji, J. Chen, H. Fang, F. Guo, and J. F. Zhang, “A Dual-Point Technique for the Entire ID–VG Characterization Into Subthreshold Region Under Random Telegraph Noise Condition,” IEEE Electron Device Lett., vol. 40, no. 5, pp.674-677, 2019, doi: 10.1109/LED.2019.2903516.

[139]. G. C. Chow, “Maximum-likelihood estimation of misspecified models,” Economic Modelling, vol.1, Issue. 2, pp.134-138, 1984, doi: 10.1016/0264-9993(84)90001-4.

[140]. S. Realov and K. L. Shepard, "Random telegraph noise in 45-nm MOS: Analysis using an on-chip test and measurement system," *in IEDM Tech. Dig.*, Dec., 2010, pp. 28.2.1-28.2.4, doi: 10.1109/IEDM.2010.5703436.

[141]. E. Wit, E. van den Heuvel, and J. W. Romeijn, “'All models are wrong...': an introduction to model uncertainty,” Statistical Neerlandica, vol.66, no.3, pp.217–236, 2012, doi:10.1111/j.1467-9574.2012.00530.x.

[142]. The MathWorks, I., 2019. Statistics and Machine Learning Toolbox, Natick, Massachusetts, United State. Available at: https://www.mathworks.com/help/stats/

[143]. J. Heyes, “Basic Tools, Normal Probability Plots,” ASQC Statistics Division Newsletter, vol. 13, no. 1, 1992.

[144]. C. Prasad et al., “Bias temperature instability variation on SiON/Poly, HK/MG and trigate architectures,” in Proc. IEEE Int. Rel. Phys. Symp., Jun. 2014, pp. 6A.5.1–6A.5.7, doi: 10.1109/IRPS.2014.6861101.

[145]. M. P. Li, “Jitter challenges and reduction techniques at 10 Gb/s and beyond,” IEEE Trans. Adv. Package., vol. 32, no. 2, pp. 290–297, May 2009, doi: 10.1109/TADVP.2009.2012432.

[146]. J. Zou et al., “New insights into AC RTN in scaled high-κ/metal-gate MOSFETs under digital circuit operations,” in Proc. Symp. VLSI Technol. (VLSIT), Jun. 2012, pp. 139–140, doi: 10.1109/VLSIT.2012.6242500.

[147]. H. Miki et al., “Voltage and temperature dependence of random telegraph noise in highly scaled HKMG ETSOI nFETs and its impact on logic delay uncertainty,” in Proc. Symp. VLSI Technol. (VLSIT), Jun. 2012, pp. 137–138, doi: 10.1109/VLSIT.2012.6242499.

[148]. C. Liu, K. T. Lee, H. Lee, Y. Kim, S. Pae, and J. Park, “New observations on the random telegraph noise induced Vth variation in nano-scale MOSFETs,” in Proc. IEEE Int. Rel. Phys. Symp., Jun. 2014, pp. XT.17.1–XT.17.5, doi:10.1109/IRPS.2014.6861194.

[149]. K. Ota, M. Saitoh, C. Tanaka, D. Matsushita, and T. Numata, “Systematic study of RTN in nanowire transistor and enhanced RTN by hot carrier injection and negative bias temperature instability,” in Symp. VLSI Technol. (VLSI-Technol.) Dig. Tech. Papers, Jun. 2014, pp. 1–2, doi: 10.1109/VLSIT.2014.6894417.

[150]. J. F. Zhang, A. Manut, R. Gao, M. Mehedi, Z. Ji, W. Zhang, and J. Marsland, “An assessment of RTN-induced threshold voltage jitter,” *IEEE 13th International Conference on ASIC (ASICON)*, pp. 1-4, 2019.

[151]. K. P. Cheung and J. P. Campbell, “On the magnitude of random telegraph noise in ultra-scaled MOSFETs,” in Proc. IEEE Int. Conf. IC Design Technol. (ICICDT), May 2011, pp. 1–4. doi: 10.1109/ICICDT.2011.5783191.

[152]. C. Chen, Q. Huang, J. Zhu, Y. Zhao, L. Guo, and R. Huang, “New understanding of random telegraph noise amplitude in tunnel FETs,” IEEE Trans. Electron Devices, vol. 64, no. 8, pp. 3324–3330, Aug. 2017. doi: 10.1109/TED.2017.2712714.

[153]. I. Zadorozhnyi, J. Li, S. Pub, H. Hlukhova, V. Handziuk, Y. Kutovyi, M. Petrychuk, and S. Vitusevich, “Effect of gamma irradiation on dynamics of charge exchange processes between single trap and nanowire channel,” Small, vol. 14, no. 2, Jan. 2018, Art. no. 1702516. doi: 10.1002/smll.201702516.

[154]. M. Mehedi, K. H. Tok, Z. Ye, J.F. Zhang, Z. Ji, W. Zhang, and J.S. Marsland, “On the Accuracy in Modelling the Statistical Distribution of Random Telegraph Noise Amplitude,” *IEEE Access*, 9, pp.43551-43561, 2021.

[155]. B. Kaczer et al., “The relevance of deeply-scaled FET threshold voltage shifts for operation lifetimes,” in Proc. IEEE Int. Rel. Phys. Symp. (IRPS), Apr. 2012, pp. 5A.2.1–5A.2.6, doi: 10.1109/IRPS.2012.6241839.

[156]. N. 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, 2008, pp. 541-546, doi: 10.1109/RELPHY.2008.4558943.

[157]. Z. Zhang, *et al*. "New approach for understanding "random device physics" from channel percolation perspectives: Statistical simulations, key factors and experimental results," *IEEE International Electron Devices Meeting (IEDM)*, 2016.

[158]. R. Gao, M. Mehedi, H. Chen, X. Wang, J. F. Zhang, X.L. Lin, Z.Y. He, Y.Q. Chen, D.Y. Lei, Y. Huang and Y. F. En, "A fast and test-proven methodology of assessing RTN/fluctuation on deeply scaled nano pMOSFETs," IEEE International Reliability Physics Symposium (IRPS), (pp. 1-5), 2020, doi: 10.1109/IRPS45951.2020.9129230.

List of Publications

[1]. **M. Mehedi**, K. H. Tok, J. F. Zhang, Z. Ji, Z. Ye, W. Zhang, and J. S. Marsland, "An assessment of the statistical distribution of Random Telegraph Noise Time Constants," *IEEE Access*, vol. 8, no. 10, pp.1496-1499, 2020, doi: 10.1109/ACCESS.2020.3028747.

[2]. **M. Mehedi**, K. H. Tok, Z. Ye., J. F. Zhang, Z. Ji, W. Zhang, and J. S. Marsland, "On the Accuracy in Modeling the Statistical Distribution of Random Telegraph Noise Amplitude," *IEEE Access*, 9, pp.43551-43561, 2021, doi: 10.1109/ACCESS.2021.3065869.

[3]. **M. Mehedi**, K. H. Tok, J. F. Zhang, Z. Ji, Z. Ye, W. Zhang, and J. S. Marsland, "An integrated method for extracting the statistical distribution of RTN time constants," *IEEE 14th International Conference on ASIC (ASICON)*, 2021.

[4]. J. Brown, J. F. Zhang, B. Zhou, **M. Mehedi**, P. Freitas, J. Marsland, and Z. Ji, "Random-telegraph-noise-enabled true random number generator for hardware security," *Scientific reports*, no. 0123456789,10(1), pp.1-13, 2020.

[5]. R. Gao, **M. Mehedi**, H. Chen, X. Wang, J. F. Zhang, X.L. Lin, Z.Y. He, Y.Q. Chen, D.Y. Lei, Y. Huang and Y. F. En, "A fast and test-proven methodology of assessing RTN/fluctuation on deeply scaled nano pMOSFETs," *IEEE International Reliability Physics Symposium (IRPS)*, (pp. 1-5), 2020, doi: 10.1109/IRPS45951.2020.9129230.

[6]. J. F. Zhang, M. Duan, **M. Mehedi**, K.H. Tok, Z. Ye, Z. Ji, and W. Zhang, "Defect loss and its physical processes," *IEEE 15th International Conference on Solid-State & Integrated Circuit Technology (ICSICT)*, (pp. 1-4), Nov. 2020, doi: 10.1109/ICSICT49897.2020.9278393.

[7]. A. Manut, R. Gao, J. F. Zhang, Z. Ji, **M. Mehedi**, W. D. Zhang, D. Vigar, A. Asenov and B. Kaczer, "Trigger-When-Charged: A Technique for Directly Measuring RTN and BTI-Induced Threshold Voltage Fluctuation Under Use-Vdd," *IEEE Trans. Electron Devices*, vol. 66, no.3, pp.1482-1488, 2019, doi: 10.1109/TED.2019.2895700.

[8]. J.F. Zhang, A. Manut, R. Gao, **M. Mehedi**, Z. Ji, W. Zhang, and J. Marsland, "An assessment of RTN-induced threshold voltage jitter," *IEEE 13th International Conference on ASIC (ASICON)*, 2019, doi: 10.1109/ASICON47005.2019.8983559.

Appendix A

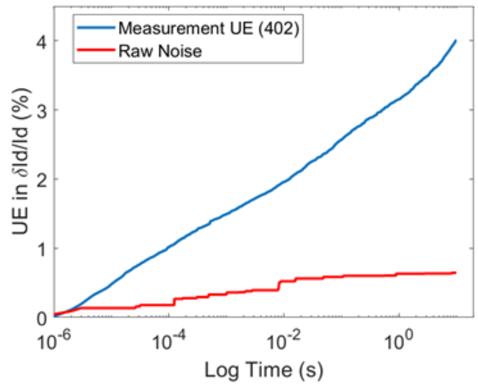
▪ Trap Extraction Methodology

Firstly, one looks for a clear step-like change in the measurement data and an example is given in Appendix A 2 (a). Then, as shown in the figure, locate the start and the end point within a shifted level for further averaging process. The start and end point can be selected by checking if the next upcoming 5 data points, ($\frac{\delta Id}{Id}_{ith+1}$, $\frac{\delta Id}{Id}_{ith+2}$, $\frac{\delta Id}{Id}_{ith+3}$, $\frac{\delta Id}{Id}_{ith+4}$, $\frac{\delta Id}{Id}_{ith+5}$) are within $+15\%\ and -15\%$ of current $\frac{\delta Id}{Id}_{ith}$. 15% is considered as the system noise level and it can be justified by the results in Appendix A 1. The envelope of system noise reaches $\frac{\delta Id}{Id}$ = 0.64%, which is 15.8% of the average envelope of 402 devices ($\frac{\delta Id}{Id}$ = 4.03%). Hence, 15% tolerance level is used.

To detect a step-like change, the differences between $\frac{\delta Id}{Id}_{ith}$ and its next 5 points are checked. The result of checking is a Boolean ('0' or '1') statement. If the difference is larger than 15%, it is a '1'. For example, for 5 checking points is used in Appendix A 2 (b), the results can either be '00000', '11111' or a mixture of '0' and '1'. The 5 checking points will be multiplied together as a single output of '0' or '1'. Thus, '00000' and mixture of '0' and '1' will lead to an output of '0'.

Output of 1 indicates that all 5 points are at least 15% away from $\frac{\delta Id}{Id}_{ith}$, so that it corresponds to a step like change. The $\frac{\delta Id}{Id}_{ith+1}$ will be the start point for the upcoming 'flat' region, therefore. If output is 0, the loop for start point or end point searching will continue until an output of '1' is detected.

Once the start and end point are detected, all points between the start and end point are summed up and averaged to obtain a mean value. The mean value will be plotted as a red line as shown in Appendix A 2 (b).

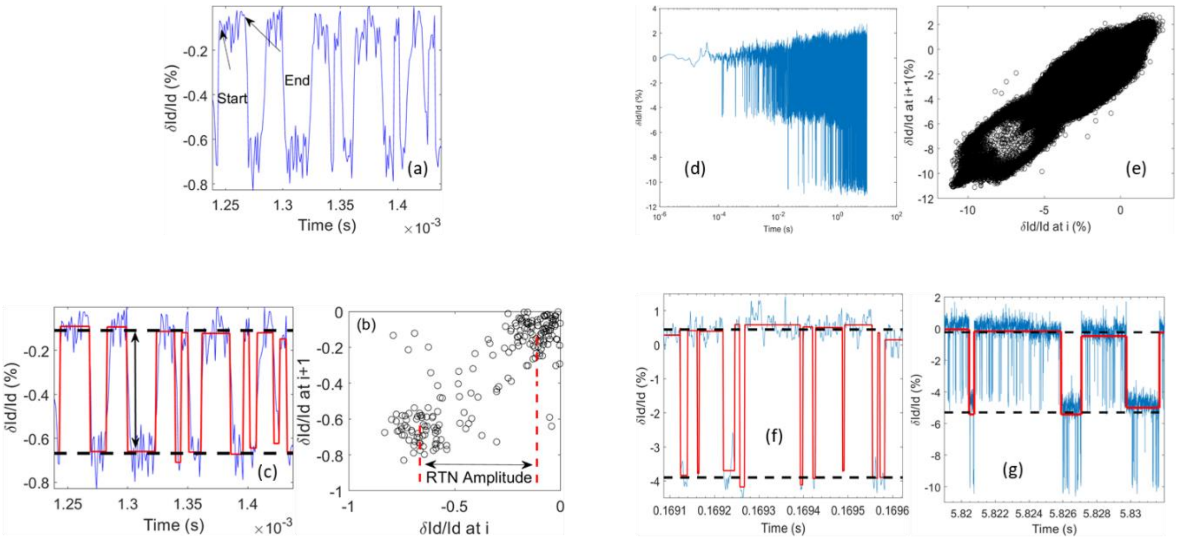


Appendix A 1 shows the average Up-Envelope (UE) of 402 devices in blue line and raw noise data in red.

The simplified signal is further processed to obtain the RTN signal with two discrete levels. The high levels of simplified signal are averaged to give the high-level of the RTN. Similarly, the low-level of simplified signal is averaged to obtain the low level of RTN. For example, the black dashed lines in Appendix A 2 (b) are the two discrete levels of the RTN. A signature of RTN is that Id switches between two levels multiple times. To justify the presence of RTN, a minimum of 3 switches between the two levels are used.

This method has been justified by using TLP. The results of the new method have been reasonably close to the result of TLP. For example, the result extracted is $\frac{\delta Id}{Id}$ =0.559%by the new method and $\frac{\delta Id}{Id}$ =0.558% is extracted from the TLP with the same data in Appendix A 2 (b) and Appendix A 2 (c). Thus, the method is justified and applied to the trap extraction of measurement data hereafter.

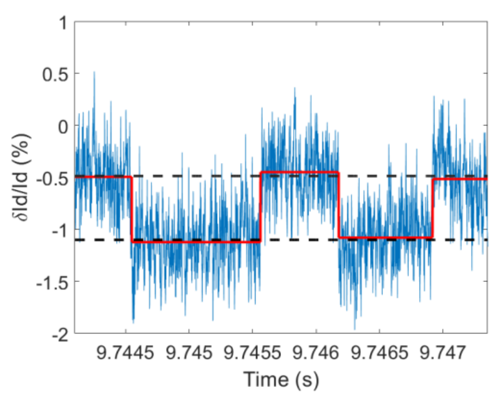
Appendix A 2 (d) gives one example where there are both fast and slow traps. TLP is not applicable in this case, as Appendix A 2 (e) shows that one cannot extract RTN amplitude by using the TLP for the signal of Appendix A 2 (d). The WTLP is not working either, as the density of cluster is hardly identified. In this case, the amplitude of the fast trap can be extracted first by using a sufficiently small-time window that the slow trap does not cause a switch, as shown in Appendix A 2 (f). With longer time window available, another slow trap can become active and available for amplitude extraction. To determine the amplitude of the slow trap, a large time window is used, as shown in Appendix A 2 (g). The extraction method for slower trap will be discussion in the upcoming sub section. Additionally, 3 traps are the maximum number of traps per device that can be reliably extractable by the measurement data applied.



Appendix A 2 (a) A Sample of RTN Trap found in devices. (b) Single Trap with Simplified Trap Modelled. (c) TLP result to prove our extraction method. (d) An example of two trap long measurement data. (e) TLP showed an unextractable result. (f) and (g) display the result of the 1st and 2nd Trap extracted from data in (d).

▪ Slow Trap Extraction

For slower and complex trap as in Appendix A 3, 'checking points' extraction method will not work directly because the signal corresponding to the two levels of slow trap can overlap and fall within +15% and -15%, caused by the fluctuation originating from the fast trap. This leads to the failure of 'checking points' extraction method in detecting any step like change in the data. Hence, to make 'checking points' extraction works for such overlapping data, the procedures are modified, as discussed below.



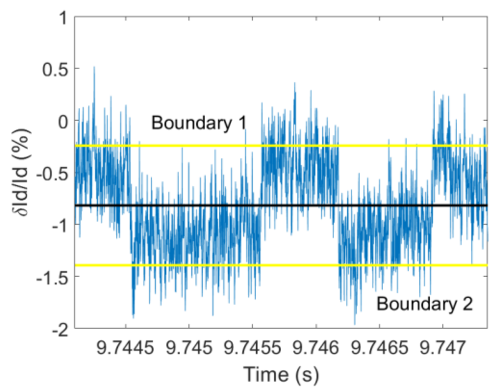
Appendix A 3 shows an extracted slow trap example.

Firstly, an average value of the whole dataset within the time window is calculated and it is -0.8181% $\frac{\delta Id}{Id}$ for this example, plotted as black line in Appendix A 4. Since this is the slower trap in this device, the fast trap which is previously extracted will be included in this slower trap extraction. 0.5% $\frac{\delta Id}{Id}$ fast trap is extracted previously, with 15% noise layer introduced to the fast trap, so that $0.5\% * 15\%$ gives 0.575%. This gives the estimation of fast trap with noises. Then, adding 0.575% to the average value gives 2 boundaries of,

i) -0.8181% + 0.575% = -0.2431% (Boundary 1)

ii) -0.8181% - 0.575% = -1.3931% (Boundary 2)

Both boundaries form the lines plotted in yellow in Appendix A 4, labelled as 'Boundary 1' and 'Boundary 2'. Each boundary will be used to perform an inspection over each data point, checking if the data points is above 'Boundary 1' or below 'Boundary 2'. This is done by using 'for loop' with the size of data as the looping iteration size.



Appendix A 4 shows the average line and fast trap boundaries with noise layer introduced.

Each boundary inspection will require one loop and there will be 2 loops to be done, therefore. First loop (Loop 1) checks if $data_{ith}$ (data at current time point in the loop) is above the 'Boundary 1'. An array named 'L1' is used to store the results of Loop 1. How Loop 1 works is shown below,

i) If $data_{ith}$ is above the 'Boundary 1', $L1_{ith}$ = '1'.

ii) If $data_{ith}$ is below the 'Boundary 1', $L1_{ith}$ = '0'.

At last, 'L1' is an array with the same size of the data set since it is checking every data point with Loop 1 and it contains of '0' or '1' at each time point.

Second loop (Loop 2) is introduced to check if $data_{ith}$ is below the 'Boundary 2'. An array named 'L2' is used to store the results of Loop 2. How Loop 2 works is shown below,

i) If $data_{ith}$ is below the 'Boundary 2', $L2_{ith}$ = '2'.

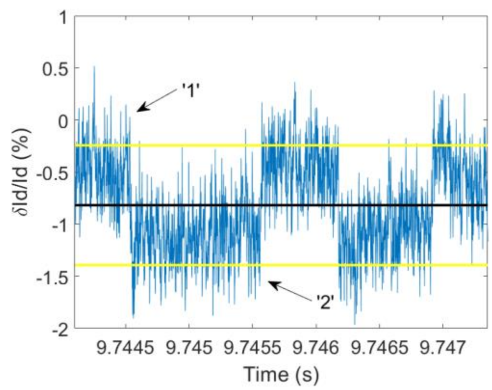
ii) If $data_{ith}$ is above the 'Boundary 2', $L2_{ith}$ = '0'.

'L2' array contains of '0' and '2' at each time point. Now, L1 and L2 will be combined to form an array, 'L_{result}' with mixtures of '0', '1' or '2' at each time point. Hereby, in 'L_{result}',

i) '0' represents data within 'Boundary 1' and 'Boundary 2'.

ii) '1' represents data above 'Boundary 1'.

iii) '2' represents data below 'Boundary 2'.



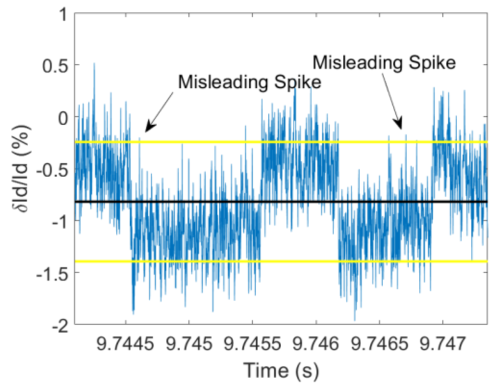
Appendix A 5 indicates the results region from for loops inspection.

'1' and '2' in L_{result} give indication of the step like change. Check starts from the first time point of $L_{result_{1th}}$ until the 'if statements' detect 3 continuous of '2' without '1' in between. '0' will be ignored as they are data within the boundaries. For example,

i) Array results, '0' '0' '0' '0' '0' '1' '1' '1' '0' '2' '1' '2' '2' is not the step like change.

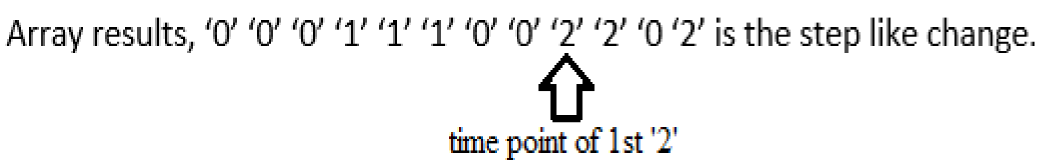
ii) Array results, '0' '0' '0' '1' '1' '1' '0' '0' '2' '2' '0 '2' '0' is the step like change.

A continuous detection of '2' is needed because there might be some misleading spike causing the $data_{ith}$ to break above 'Boundary 1' or below 'Boundary 2', thus, giving wrong indication of step like change. Example of misleading spikes is shown in Appendix A 6.



Appendix A 6 shows the misleading spike in the data.

Once 3 ‘2’ is detected, the time point of the first upper data can be determined. The time point is given by taking the time point of the 1^{st} ‘2’ detected in the 3 ‘2’ and minus by 1, illustration is given in Appendix A 7. This gives where the first upper data ends before step like change.



Appendix A 7 gives the clear illustration of how time point is picked.

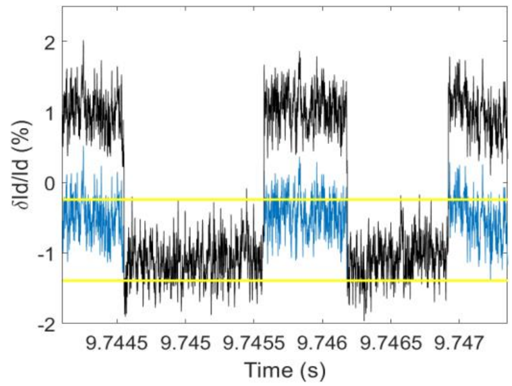
For instance, the time point of 1^{st} ‘2’ is 451_{th} out of 3200 data points (data size within the time window). The first upper of data starts from 1_{st} and ends at 450_{th} ($451_{th} - 1$).

To clearly visualize if the step like change is captured correctly, an upward shifting is done by offsetting the step like change by 3 times of the fast trap amplitude:

i) 0.5% * 3 = 1.5% (the offset value applied)

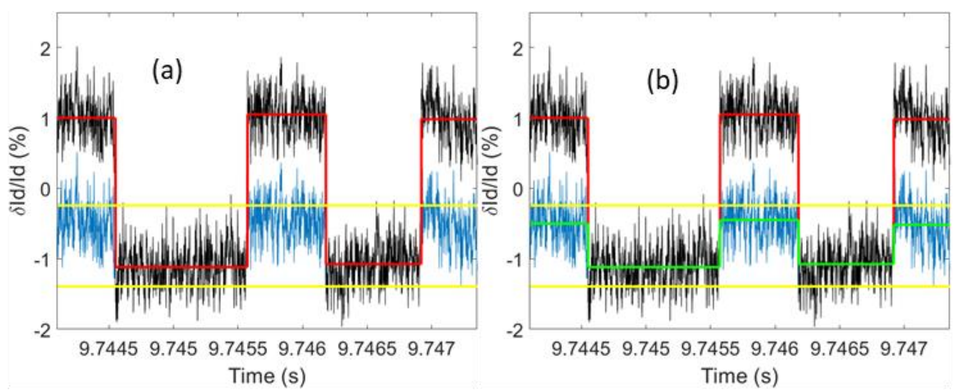
Also, this helps to prevent the ‘checking points’ extraction method used for the fast trap from failing in the overlapping data between steps like those in Appendix A 6. The data from 1_{th} to 450_{th} will be added by the offset value calculated above individually. Throughout these procedures, an offset data can be generated from the true data as plotted in black in Appendix A 8.

To clarify, only the upper edge of data is raised by offsetting. Lower edge data remains the same values as it is. Now, ‘checking points’ extraction method can be applied to extract the trap amplitude accurately with the clear black data.



Appendix A 8 shows the offset-ed data plotted in black.

With ‘checking points’ extraction method, the average of each shifted data is plotted as red in Appendix A 9 (a), but the red plotted line is the average of offset data. To extract the amplitude of true data, the upper shifted red lines must be taken off by the offset values of 1.5%. After taken off, the average is plotted as green line in Appendix A 9 (b). The green line represents the true extracted slow trap amplitude from the true data.



Appendix A 9 (a) and (b) show the average line for offset-ed and non-offset data.

Appendix A 10 (a) shows an example of a huge amplitude fast trap included. Fast trap amplitude that is previously extracted for this device is -4.2% $\frac{\delta Id}{Id}$. This amplitude acts as the centre line and $\pm 15\%$ of the amplitude applies to form the ‘Boundary 1’ and ‘Boundary 2’:

(i) Boundary ‘1’ = $-4.2\% * 0.85$ = -3.57%,

(ii) Boundary ‘2’ = $-4.2\% * 1.15$ = -4.83%.

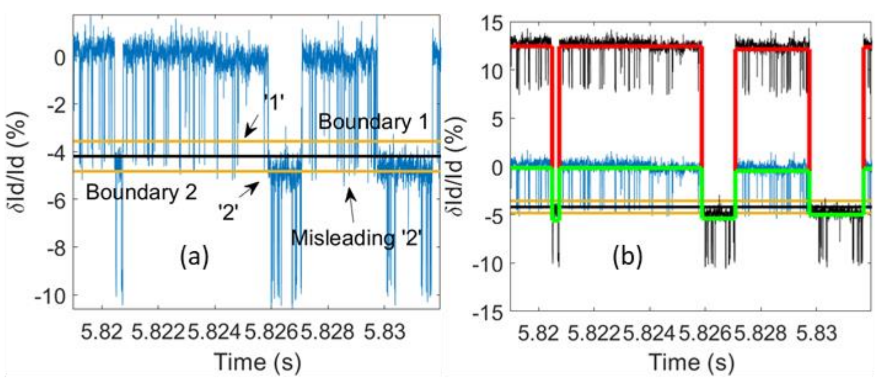
Both boundaries are plotted as yellow lines in Fig.6i. The procedures for Loop 1 and Loop 2 remain the same, $data_{ith}$ beyond ‘Boundary 1’ gives ‘1’ in ‘L1’ array and $data_{ith}$ below ‘Boundary 2’ gives '2' in 'L2'. $data_{ith}$ below 'Boundary 1' or above 'Boundary 2' gives '0' in 'L1' and 'L2'. L_{result} is obtained by combining 'L1' with 'L2'.

i) '0' represents data within 'Boundary 1' and 'Boundary 2'.
ii) '1' represents data above 'Boundary 1'.
iii) '2' represents data below 'Boundary 2'.

The criteria of step like change detection in this example is increased to 10 continuous time of '2' to be detected. This is increased due to the number of misleading '2' in Appendix A 10 (a) is larger. Then, the time point of the 1st '2' minus by 1 equal to the time point for first upper data. This gives the starting and ending time point for first upper data, this will be offset by the values of,

i) $-4.2\% * 3 = -12.6\%$ (the offset value applied).

Lastly, the offset-ed data is plotted as black data in Appendix A 10 (b). 'Checking points' extraction method is applied to extract the trap amplitude of black data and the result is plotted as red line in Appendix A 10 (b). The true trap amplitude is then extractable by taking off the offset value from red line. This forms the green line in Appendix A 10 (b).



Appendix A 10 (a) and (b) show the example of how huge amplitude slow trap can be extracted.