Statistical Analysis and Predictive Modelling of Random Telegraph Noise in Nanometre Devices for Low Power Applications

Kean Hong Tok

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

December 2023

ACKNOWLEDGEMENTS

First, I would like to thank my professor, Prof. J. F. Zhang for his guidance and continuous support throughout my PhD research project. Without his invaluable guidance, it would not have been possible to conduct this research.

Second, I would like to express my appreciation to my co-supervisors, Dr. John Marsland and Dr. Zhigang Ji for being supportive and for encouraging my PhD study.

I also would like to thank anyone who has been involved during my time in LJMU and worked in the same group, providing any kind of help for the completion of this project. Notable among them are Dr. James Brown, Dr Mehzabeen Mehedi, Dr Firas Hatem, Dr Pedro Freitas, and Dr. Zheng Chai.

I would like to express my appreciation to the faculty for funding my PhD tuition throughout the years. Also, I wish to thank the other support staff of the university. This includes those who helped in ordering electronic components required for LJMU nano-laboratory.

Finally, I would like to thank my family for offering me a great opportunity to pursue a PhD in LJMU, and for their support throughout these years.

ABSTRACT

As the Internet of Things is becoming the new normal in electronics, one of the core issues is low power consumption. The direct solution is to reduce the operation voltage towards the threshold voltage of Metal Oxide Silicon Field Effect Transistors (MOSFETs); however, this reduces noise toleration. The focus of this research is Random Telegraph Noise (RTN), which is one of the dominant noises in MOSFETs. It is well known that the impacts of RTN are inversely proportional to the size of the transistors. As the size of the transistor goes down, the impact of RTN increases and this affects the performance of circuits, such as causing jitters.

RTN is a stochastic and step-like fluctuation of drain or gate current of a device under given voltages. It occurs when an electron/hole is captured or emitted from or into a trap in the oxide layer. RTN has received attention widely, but understanding of it is incomplete, and its modelling remains a challenge. At present, there is no trustworthy RTN model of acceptable accuracy for circuit designers to use to optimize the performance of their circuits. One of the key weaknesses of early models is that they were developed based on test data gained from a short time window and their ability to predict future long-term RTN was not verified. The aim of this work is to develop a statistical model that can be used to predict the long-term impact of RTN on the device.

To develop RTN models, early works followed a bottom-up approach by extracting the properties of individual traps, including RTN amplitude per trap, capture and emission times. This requires selecting devices where the fluctuation can be clearly separated into the contributions of individual traps. In reality, however, there are many devices where such separation cannot be achieved. Excluding these devices when developing RTN models makes it questionable whether the data used for developing the model represents the statistical properties of real devices. Also, this is time consuming and does not provide direct information for circuit designers to evaluate their circuit stability. The needed direct information includes shifts in threshold voltage, ΔVth, and in drain current, $\Delta I_D/I_D$. An open question is: what is the probability that RTN can cause ΔVth in mV or $\Delta I_D/I_D$ in % to go beyond a certain level? A top-down modelling approach is proposed and used in this work. This approach does not select devices and extracts the cumulative distribution function (CDF) of RTN by measuring the impacts of all traps and integrating data from multiple devices into one dataset. This new approach will be referred to hereafter as the integral approach.

RTN has been widely characterized under a fixed voltage (DC); for example, in this project, Vg=0.5V and 0.9V have been applied as the gate voltage. The integral approach removes the burden of determining capture and emission time constants of individual traps by introducing the concept of 'Effectively Charged Traps' (ECT). A time-dependent number of ECTs is introduced to replace the time constant distribution and the kinetics of ECT against time are extracted, which can be extrapolated to predict the long-term RTN. The predictive ability of the integral approach is verified.

Rather than DC RTN, the majority of commercial circuits operate under AC conditions. It has been reported that AC RTN behaves differently from DC RTN. Early work has been carried out for AC RTN on characterizing the single trap properties, such as time constants and amplitude under different frequencies. In this work, the integral approach has been applied to analyse and model AC RTN. It is shown that the model developed by the integral approach can also be used to predict future RTN. The relationship is assessed between capture, emission time constants of a single trap and the number of ECTs against frequency.

A 6T SRAM circuit in read operation requires transistors to operate in either a linear or a saturation region. Early works have worked on saturation measurement but knowledge in the area is limited: no CDF information has been provided. To optimize circuit performance, RTN under saturation conditions must be measured, analysed, and modelled. Furthermore, both driving current and threshold voltage shift are needed in real circuit operation. Without them, the simulation of RTN for real circuit operation cannot be completed. RTN measurement under saturation is carried out and analysed by using the integral approach. ΔVth and $\Delta I_D/I_D$ have been found to follow different statistical distributions.

In summary, to optimize circuit performance and yield against RTN, designers need to know the probability of the RTN-induced ΔVth and $\Delta I_D/I_D$ reaching a certain level within a specific time window. This work tackles this challenge by developing an integral approach to extract an RTN model which can not only model short-term RTN, but also predict future RTN. It is hoped that the new model can eventually be embedded into an industrial standard simulator for circuit optimization.

LIST OF ABBREVIATIONS

Abbreviation	Definition
AC	Alternating Current
BTI	Bias Temperature Instability
CDF	Cumulative Distribution Function
Conv	Convolution
CMOS	Complementary Metal Oxide Semiconductor
CSR wafer	Cambridge Silicon Radio wafer
DAQ	Data Acquisition
DC	Direct Current
DDV	Device-Device Variation
DUT	Device Under Test
ECT	Effective Charged Trap
Env	Envelope
FHMM	Factorial Hidden Markov Model
GEV	Generalized Extreme Value
GND	Ground (signal-system)
GB	Giga-bytes
HCI	Hot Carrier Injection
HK/MG	High-κ Dielectric and Metal Gate
HMM	Hidden Markov Model
IV	Drain current (I_D)~Gate voltage (Vg)
IoT	Internet-of-Thing
ITRS	International Technology Roadmap for Semiconductors
JFET	Junction Field Effect Transistor
MLE	Maximum Likelihood Estimation
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
MS	Mega-Sample points
MB	Mega-bytes
NBTI	Negative Bias Temperature Instability

NMOS	Negative Metal Oxide Semiconductor
NI	National Instrument
OSC	Oscilloscope
PC	Personal Computer
PDF	Probability Density Function
PMOS	Positive Metal Oxide Semiconductor
RTN	Random Telegraph Noise
RTS	Random Telegraph Signal
SMU	Source Measurement Unit
SRAM	Static Random-Access Memory
SSE	Sum-Square-Error
TLP	Time Lag Plots
TSMC	Taiwan Semiconductor Manufacturing Company
TWC	Trigger-When-Charged
TDV	Time Dependent Variation
UE/LE	Upper/Lower Envelope
VB	Visual Basic
WTLP	Weighted Time Lag Plot

LIST OF SYMBOLS

Symbol	Description	Unit
μ	Mean/average value	
μ_0	Carrier mobility	
Ef/EF	Fermi level	eV
E_T	Trap energy level	
f	Frequency	Hz
g_m	Transconductance of an I_D-Vg curve	S
I_D	Drain current	A
I_{dcc}	Constant I_D value for Vth extraction	A
k	Boltzmann's constant	
N	Number of carriers per unit area	m^{-3}
N_A	Average Number of Effective Acceptor-like Traps	
N_D	Average Number of Effective Donor-like Traps	
n_A	Number of Acceptor-like Traps	
n_D	Number of Donor-like Traps	
Nt	Number of occupied oxide traps	
T	Temperature	^oC
t	Time	sec
Vd	Drain Voltage	V
Vdd	Drain supply voltage /Circuit operating voltage	V
Vg	Gate voltage	V
Vgs	Gate-to-source voltage	V
Vth	Threshold voltage	V
W/L	MOSFET channel width/length	m
ΔI_D	Drain current fluctuation	A
$\Delta I_{D,LIN}$	Drain current fluctuation in linear region	A
$\Delta I_{D,SAT}$	Drain current fluctuation in saturation region	A
δVth	Single trap amplitude	V
ΔVth	Threshold Voltage shift	V

ΔVth_A	Contribution from Acceptor-like Traps	V
ΔVth_D	Contribution from Donor-like Traps	V
$\Delta Vth_{,LIN}$	Threshold Voltage shift in linear region	V
$\Delta Vth_{th,SAT}$	Threshold Voltage shift in saturation region	V
σ	Standard deviation	
T	Average time constant	sec
τc	Average capture time	sec
τ_e	Average emission time	sec
$\tau_{CH} \setminus \tau_{CL}$	τ_C under high or low gate bias	sec
$\tau_{EH} \setminus \tau_{EL}$	τ_E under high or low gate bias	sec
η	Parameter of Exponential	
ϵ, θ	Location and Scale parameters for Lognormal	
ξ, α, β	Shape, Location and Scale parameters for GEV	

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

1.1 Background

The high-performance demands and growth of consumer products have driven significant progress and advancement in technology using complementary metal-oxide semiconductors (CMOS). At the same time, metal oxide semiconductor field effect transistors (MOSFETs) are facing instability issues. The most direct way to increase the performance of the circuit is downscaling the size of MOSFETs, allowing more devices to be fitted into the same area and increasing operation speed. In addition, the arrival of the Internet of Things (IoT) era requires power consumption to be minimized. As the aggressive scaling of transistor size continues, instability issues have become more severe [1-11]. Key instability issues include random telegraph noise (RTN), bias temperature instability (BTI), hot carrier injection (HCI), etc. These instabilities affect the MOSFET parameters, such as mobility degradation, current fluctuation, and threshold voltage shifting. For this reason, they cause degradation of device performance [1-11].

This research work focuses on the instability caused by RTN. RTN is a stochastic noise caused by trapping and de-trapping of charge carriers from/to the conduction channel [1-3], [6], [12-17]. RTN becomes increasingly important as MOSFETs become smaller; the impact of single-trap charge increases, affecting the operation of electronic circuits [1-3], [6], [12-17].

1.2 Key Issues

Efforts have been made in early works [7], [9-10], [18-19] to improve understanding and provide solutions for instabilities, especially on BTI. The understanding and impact of RTN instabilities in devices are also widely investigated [1-3], [6], [12-17], [20]. Although progress has been made, the understanding of RTN instabilities remains incomplete and many issues are unresolved. As IoT requires low power consumption, the easiest approach to achieve low power consumption is to lower the voltage applied. This approach allows effective power reduction. However, it leads to a significant issue where the gate voltage (Vg) is very close to the threshold voltage, Vth. This small difference between Vg and Vth reduces the tolerance of instability in Vth [3], [12], [21]. For instance, with a single RTN step-like voltage shift, the effective gate bias can be lower than Vth, which pushes the device into sub-threshold condition.

One of the shortcomings of early works on RTN is the relatively short time window used to measure RTN. Commercial devices are expected to work at the rated operation bias for 10 years. The experimental RTN measurement time used by early works [22-23], however, is far shorter than 10 years; the measurement time used by many works is 1 sec or less [24-26], although there are a few works that used longer times [22-23], [27-28]. With the RTN signal measured within this short time range, the accuracy and authenticity of the analysed results or statistical distribution are questionable. Long measurements of 20 sec or more are needed. One difficulty is the huge amount of data to be handled for simulators and laboratory instruments. For example, 1 s data at 1 MSa/s occupies roughly 4.5 MB (Megabytes) of memory in storage: a 1-hour RTN measurement (3600 sec) requires around 16.2 GB of memory for 1 device.

Early works often only measured devices that showed clear RTN, but this criterion is not realistic for devices that are commercially designed, as many devices do not have clear RTN signals. This means that the analysed results are based on only a subset of devices and are not true representative of the whole population. This can be solved by analysing the results from all measured devices and not deselecting devices that do not show clear RTN within the measurement window. This allows the true statistical properties of the RTN to be extracted. By doing this, trustworthy results can be obtained in real conditions.

Several statistical distributions have been proposed for the amplitude and time constants of RTN. The main issue of these results is the small number of data points used for extracting statistical distributions; they are typically less than 200 data points [29]. This reduces the accuracy of the statistical distribution as the larger the number of data points, the more accurate the statistical distribution will be [30]. As the gate voltage must be higher than the threshold voltage to turn on the MOSFETs, the question is how close a circuit designer can bring Vg to Vth. The answer largely depends on the impact of RTN. RTN causes random shifting, which can go beyond the overdrive voltage, Vg-Vth. This may accidentally shut down the MOSFETs, so that the circuit will not be operating as it should.

A typical RTN measurement in terms of ΔId or ΔVth can fluctuate in either a positive or negative direction. RTN in early works are typically analysed for ΔI_D or $\Delta Vth > 0$ [3], [19], [31]. To model RTN as close as possible to real circuit operation, ΔI_D or $\Delta Vth < 0$ cannot be ignored as they occur often. However, there has been little investigation in the past for the shift in both directions. This leads to a question: can the same statistical distributions be

applied to both negative and positive fluctuations or must they be modelled separately? The question of how to model both positive and negative fluctuations is unanswered by early works.

RTN modelling in early works followed a bottom-up approach. It was modelled by characterizing the individual trap properties [3], [19], [31-33], such as the amplitude and time constant of single traps. These traps were then grouped together into a set of data and fitted by a Cumulative Distribution Function (CDF) to determine the statistical distributions of amplitude and time constants. This approach is time-consuming and can run into difficulties, since extracting time constants of individual traps becomes difficult when multiple traps are present within a single device or when there are very few RTN events within the limited measurement window.

1.3 Motivation

The motivation of this project is to develop a model that provides the probability of RTN occurrence at any level of ΔI_D or ΔVth. At present, there is no well-accepted or trustable RTN model available for circuit designers to optimize their design in a digital circuit. RTN has been analysed in a few ways in the past; for example, Hidden Markov Model (HMM), Weighted Time Lag Plot (WTLP), etc. However, these approaches have their own limitations: HMM has a limitation on the number of traps and it requires clear RTN switching data to process; WTLP is an enhanced time lag plot (TLP) approach, but TLP only allows the extraction of a 2-level single trap. WTLP allows multiple-trap RTN to be extracted, but it can lead to confusing results on multi-cluster outputs. In addition, if the thermal noise or system noise is close to the RTN shift, TLP cannot be applied.

Since RTN fluctuates in both positive and negative directions, this work will tackle modelling of RTN in both directions. A model should not only fit the data well, but also be able to predict the RTN where test data is not available. The impact of RTN is expected to become greater as time increases. This project focuses on filling this knowledge gap. Beginning with a clear and in-depth understanding of RTN, experimental measurement conditions that maximize the RTN phenomenon will be identified, since it allows the extraction of trustworthy statistical results for long-term prediction with high accuracy. Based on these results, an industry-standard model will be developed to enable long term prediction of RTN.

1.4 Aim and Objectives

Based on the motivations mentioned above, the aims and objectives are summarized below:

Aim:

To develop an industry standard RTN model that provides the probability of RTN occurrence and to verify that the extracted model can predict long-term RTN.

Objectives:

1. In-depth understanding of RTN.
2. RTN Measurements without specific device selection.
3. Characterizing the impact of all traps on a device.
4. Combine RTN of multiple devices into one dataset (CDF) to enable statistical analysis.
5. Assess the accuracy of different types of statistical distributions.
6. Evaluate and verify the ability of the extracted model to predict long-term RTN without further test data available.

1.5 Contribution to Knowledge

The main contributions of this research work are:

1. A new RTN characterization method is developed using integral methodology. The method is constructed by measuring the impact of all traps within a device and grouping the RTN of multiple devices into one data set to extract the CDF. The developed methodology removes the burden of determining capture and emission time constants of individual traps by introducing the concept of 'Effectively Charged Traps' (ECTs). Importantly, the prediction of long-term RTN is tackled. A time-dependent number of ECTs is introduced to replace the time constant distribution. The experimental results confirm that the developed methodology has clear advantages in terms of overall fitting and error evaluation of multiple distributions. The ability to predict long-term RTN is verified by comparing the predicted CDF with actual measurements.

2. Alternating current (AC) RTN measurement is carried out and the properties of the measured RTN are extracted by the developed integral methodology. Single-trap RTN properties are evaluated against frequency. The statistical properties of the number and kinetics of ECT are investigated. Three distributions (Exponential, Lognormal and Generalized Extreme Value (GEV)) are assessed for the RTN amplitude. The experimental results prove that the integral methodology can be applied to AC RTN. Furthermore, the relationship is assessed between capture, emission time constants of single traps and the number of ECTs against frequency.

3. The impact of RTN is assessed on real circuits, such as SRAM, where transistors can operate under either linear or saturation conditions. As a result, RTN measurement under saturation is carried out and analysed by using the integral methodology. Moreover, in real circuit operation, both driving current and threshold voltage shift are needed. ΔVth extraction methods are evaluated, and the results confirm that conversion by transconductance (g_m) overestimates the impact of RTN. The results show the CDF per trap or $\Delta I_D/I_D$ and ΔVth follows different distributions. Once again, the predictive ability is verified, and the integral methodology can be applied under different Vg conditions.

1.6 Thesis Overview

This thesis consists of eight chapters and is organized as follows. Chapter 1 covers the background, key issues, and motivation of the research work. Aims and objectives are clearly defined. The contribution of knowledge achieved in this research work is described.

Chapter 2 gives a literature review on RTN modelling. Proposed distributions of RTN properties on amplitude and time constant are discussed.

Chapter 3 describes the devices and test facilities used in this research work. It includes the I_D-Vg characterisation and threshold voltage extraction methodology.

Chapter 4 covers the programme flow and details of the proposed integral methodology in this research work. The methodology is provided in steps and justifications are given for the chosen method.

In chapter 5, the integral methodology is used to extract the CDF of RTN properties. New concepts are introduced for removing the requirement of selecting individual RTN traps in

devices and the burden of determining capture and emission times of single traps. Also, the accuracy of RTN amplitude distribution per trap proposed by early works is assessed. For the first time, the ability to predict long-term RTN by the integral methodology is verified against test data.

Chapter 6 covers the assessment of AC RTN up to $3x10^4$ seconds. The dependence of AC RTN on time window and frequency is studied. The ability of the integral methodology to predict long-term AC RTN is assessed.

Chapter 7 studies the key parameters needed to model RTN for real circuit operation. Both driving current and threshold voltage, as well as transistors under linear and saturation region are assessed by using the integral methodology. The accuracy of ΔVth extraction by g_m is investigated. The number of ECTs under linear and saturation regions is discussed. The ability of the integral methodology to predict long-term RTN in linear driving current is assessed.

Finally, Chapter 8 concludes the contributions of this research work and recommends future work.

CHAPTER 2: LITERATURE REVIEW

In this chapter, the literature of early works on RTN modelling is reviewed to outline the approaches that they have followed and what has been achieved. The chapter is divided into two main sections. The first section covers the historical background and the parameters of RTN. The second section discusses the proposed RTN modelling techniques and what they can achieve, as well as the limitations of each technique.

2.1 RTN Properties

2.1.1 Origin of RTN

RTN is a stochastic noise which occurs when traps are capturing and emitting electrons or holes within the gate oxide [23], [34-37]. RTN behaviour was first observed by using a junction field effect transistor (JFET) in 1978 [38]. In 1984, RTN was identified in MOSFETs, originating in interface and bulk traps within the gate oxide [32]. Typical RTN behaviour is represented by drain current, ΔId, with capture and emission by a single trap in gate oxide which induces a two-level step-like fluctuation shown in Fig. 2.1(a-c) [33]. Fig. 2.1(b) shows electron trapping in an nMOSFET and Fig. 2.1(c) shows hole trapping in a pMOSFET. Apart from its effect on drain current, I_D, RTN is known to affect the threshold voltage of the transistor [39-40]. RTN trapping affects the transfer characteristic of the device $(I_D - V_G)$, as shown in Fig. 2.2, resulting in a fluctuation in the threshold voltage, ΔVth.

RTN has gained much attention as smaller devices are now widely used and a single trapped charge has a larger impact on smaller devices [34], [41], [42-46]. The randomness of capturing and emitting by traps has become an instability issue that affects the performance of circuits in terms of reliability and stability. For example, they can cause write/read failure in scaled memory cells like SRAM [47-53]. Increasing attention has been paid to the impact of RTN on SRAM. With the advancement of technology, SRAM has become one of the essential components in consumer electronic products, where it is commonly used for caches, due to the speed at which it can access stored data [54-55] and its low-power operation. Downscaling device sizes helps to meet the requirement to have more SRAM cells in one chip. However, RTN was mentioned as a key reliability factor for SRAM in 2009 in The International Technology Roadmap for Semiconductors (ITRS) [56]. It increases Vth variability, leading to a narrower noise margin [51-53], [56].

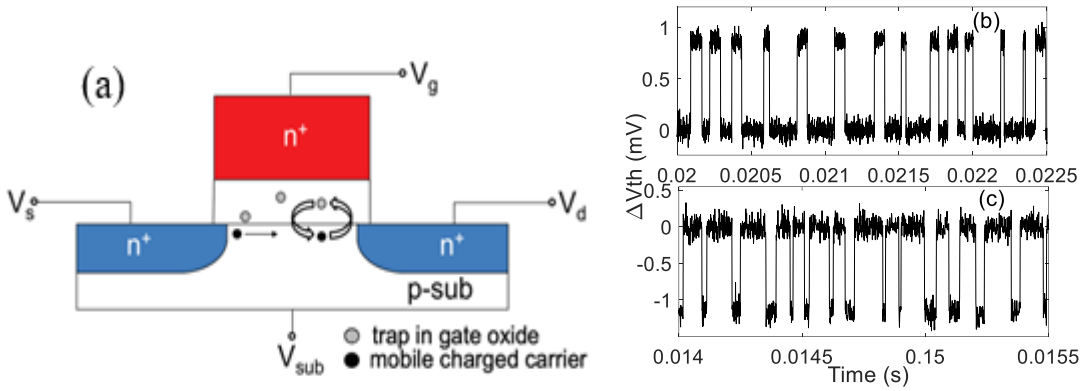


Fig. 2.1 (a) Illustration of capture and emission of single trap in gate oxide [33]. Typical RTN behaviour for nMOSFET (b) and pMOSFET (c).

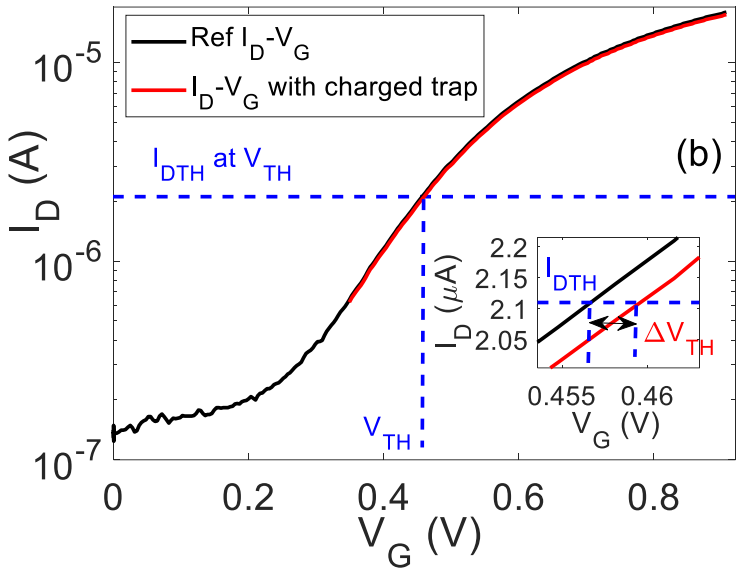


Fig. 2.2 An example of the ΔVth shift caused by RTN trap in I_D-Vg plot.

The characteristic properties of a two-level RTN trap can be summarized below:

1. The step-like amplitude shifting the I_D between the two levels of fluctuation, ΔI_D, as shown in Fig. 2.1. The amplitude can also be represented by ΔVth, as shown in Fig. 2.2.

2. The average time constants: the time for filling a trap is known as capture time, τc, and the time for neutralizing a trap is known as emission time, τe.

2.1.2 Time Constant

The time constant of a two-level RTN from a single trap can be easily extracted by taking the average of the characteristic capture time, tc, and emission time, te; these are commonly represented by τc and τe. It is to be noted that capturing an electron trap in nMOSFET causes a decrease in I_D as shown in Fig. 2.1(b), while for pMOSFET, capturing a hole trap leads to

smaller $|I_D|$ as shown in Fig. 2.1(c). Hence, it is common to refer to capture time when a trap is filled and emission time when a trap is neutralized. The probability of a two-level single trap RTN being captured or emitted can be evaluated as in Equation (2.1):

$$P_{c,e}(t) = \frac{1}{\tau_{c,e}} \exp\left(-\frac{t}{\tau_{c,e}}\right) \tag{2.1}$$

where $P_{c,e}$ represents the probability of a trap being filled or neutralized. The average time constant is $\tau_{c,e}$ and t is the interval time, considering an individual trap follows an exponential distribution [1]. $\tau c/\tau e$ is related to the difference between trap energy, E_T and Fermi level E_F as in Equation (2.2) [6], [57-59],

$$\frac{\tau c}{\tau e} = \exp\left(\frac{E_T - E_F}{kT}\right) \tag{2.2}$$

where k is Boltzmann constant, and T is temperature. RTN properties are well known to have a strong relationship with measurement conditions, such as gate voltage (Vg), drain voltage (Vd) and temperature. It has been shown that trap parameters, such as trap energy and its distance from the Si/SiO_2 interface can be extracted from these relations [1]. Fig. 2.3 shows the energy band diagram of a transistor. Based on Fig. 2.3, black circles represent the filled traps below E_F, grey circles represent the switching traps and white circles represent the neutral trap.

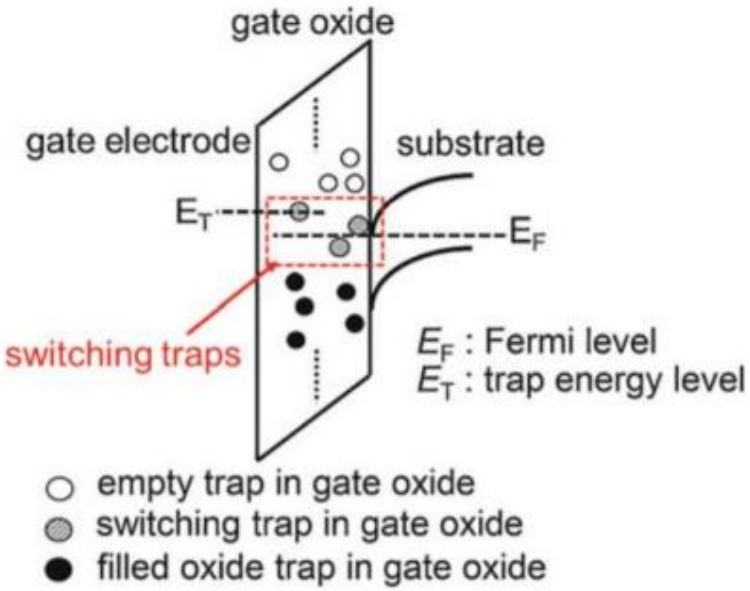


Fig. 2.3 An energy band diagram of nMOSFET [60].

To obtain accurate values of τc and τe, a sufficient number of transitions between two levels are needed. It has been stated that a minimum of 200 transitions within at least 20K data points are required to have an error less than 10% [1]. Moreover, this relates to the sampling rate of RTN measurements. A high sampling rate is needed to remove the chance of transition between levels being missed. However, a high sampling rate leads to another concern in RTN measurement, which is the amount of data to be collected. As a rule of thumb, the time window must be shortened to record the same amount of data with a higher sampling rate. This limits the time window of RTN measurements.

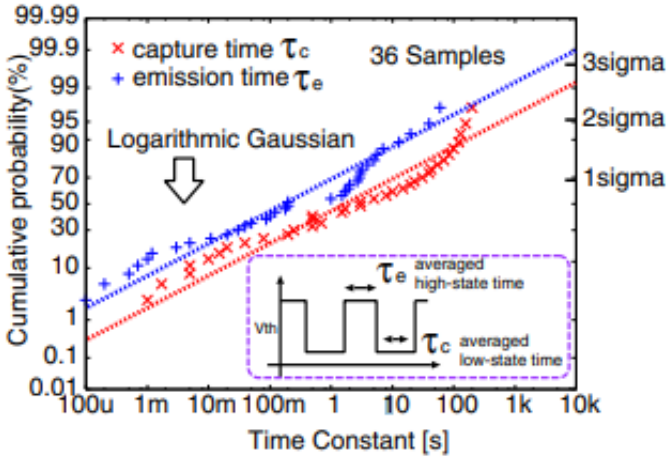


Fig. 2.4 Time constant distribution fitted by Lognormal distribution on 36 samples [29].

The trustworthiness of the extracted probability distribution of time constants is one of the main concerns. Two probability distributions for time constants have been proposed in the past, i.e., Log-uniform [1], [3] and Lognormal [29], [61]. However, the number of data points used to extract the statistical distribution has not been verified to be sufficient to determine the true distribution. For example, 36 samples were used as shown in Fig. 2.4, and it is not clear how accurate the cumulative distribution extracted from them is. As a result, modelling and predictions based on these extracted time constant distributions is questionable. An 'envelope approach' has been proposed, where the step-like changes in envelopes are characterized by the maximum and minimum of the data as shown in Fig. 2.5. The envelopes will remain the same if the following data does not have a new maximum or minimum.

The results of multiple traps from multiple devices are combined and analysed as a set of data to extract the statistical distribution of time constant as shown in Fig. 2.6 [21]. However, as

emission is minimized in this approach, it does not give the statistical distribution of emission time constants.

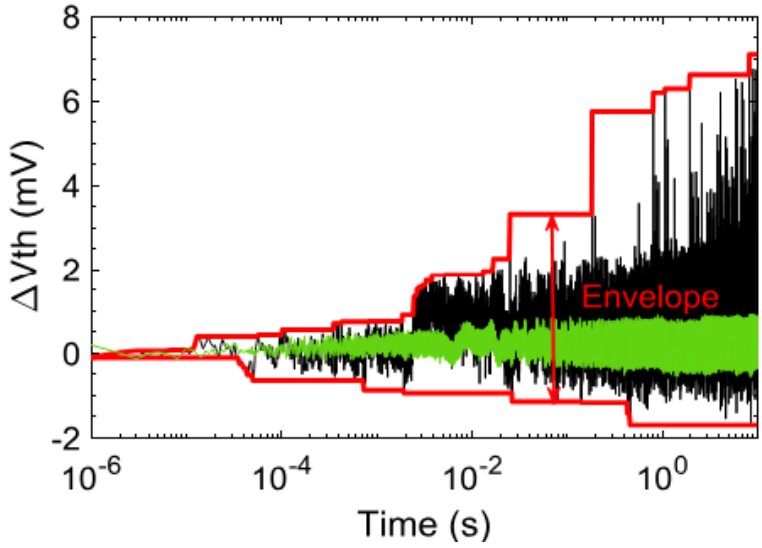


Fig. 2.5 Extraction of RTN Envelope from experimental data (black lines). The green trace represents a device of limited step-like change in ΔVth [21].

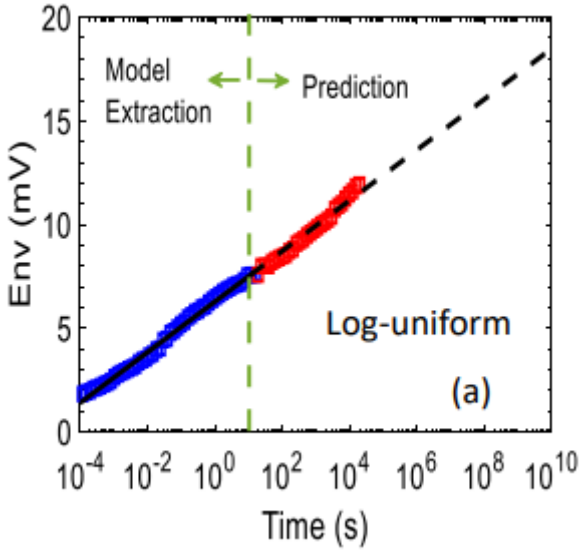


Fig. 2.6 Log-uniform distribution fitted on 'Envelope approach' for time constant distribution [21].

2.1.3 Amplitude Distribution

To model RTN, amplitude distribution is an essential parameter. The relative amplitude of RTN has grown rapidly due to downscaling of device sizes, which affects circuit stability. Also, the difference between gate and threshold voltage has been reduced steadily with the demand of low-power-operation circuits, making circuits more vulnerable to RTN impacts [30]. In addition, the number of traps in a device can vary hugely, so that the amplitude of RTN can have large device-to-device variations. It is, therefore, important to characterize the statistical distribution of RTN amplitude.

There are two mechanisms that contribute to RTN amplitude: a change in the number of carriers and a change in their mobility [1]. The normalized fluctuations in drain current can be evaluated in correlation to numbers and mobility fluctuations by [13],

$$\frac{\Delta I_D}{I_D} = \frac{1}{L\,x\,W}\left(\frac{1}{N}\frac{\partial N}{\partial N_t} \pm \frac{1}{\mu_0}\frac{\partial \mu_0}{\partial N_t}\right) \tag{2.3}$$

where L and W are the channel length and width, N is the number of carriers per unit area, N_t is the number of occupied oxide traps, and μ_0 is the carrier mobility. RTN amplitude can be obtained by taking the difference between the levels of RTN. RTN amplitude can be expressed in normalized drain current fluctuation, $\Delta I_D/I_D$, or threshold voltage shift, ΔVth. ΔVth is commonly converted from ΔId over transconductance, gm [27], [62]. Furthermore, gate voltage, Vg, can affect RTN amplitude; it is widely reported that the RTN amplitude increases with lower Vg [3], [62-63].

One of the concerns in RTN amplitude distribution is the accuracy of extracted single-trap RTN amplitude due to the random number of traps from one device to another. A complex RTN is formed by multiple traps, which greatly increases the difficulty of extracting the amplitude of each single trap. Several statistical distributions have been proposed by early works on the amplitude distribution of RTN, including Exponential [3], [64], [65], Lognormal [1], [3], [31], [65], and Gumbel [14], [29], [67] distributions. A comparison between exponential and lognormal has been carried out with RMS error calculated [66], which concluded that lognormal has a better fit [22]. Aside from monomodal distribution fitting, bimodal exponential and bimodal lognormal distributions have also been proposed [66], [68]. In 2021, Generalized Extreme Value (GEV) distribution was proposed as a better RTN amplitude distribution when compared with monomodal and bimodal exponential and lognormal distributions [30], [66],

[68]. Mehedi *et al.* [30] further claimed that GEV can fit the long statistical distribution tail well, which is important for SRAM design. In addition, Wang *et al.* reported that increasing Vg changes the RTN amplitude distribution [3].

Moreover, RTN amplitude is often modelled as positive $\Delta I_D/I_D$ or ΔVth (acceptor-like trap), representing electron traps for nMOSFET and hole traps for pMOSFET. However, in a typical RTN measurement, fluctuation occurs in both positive and negative directions, indicating that RTN traps can have either positive or negative charges. As a result, a correct distribution model should not only model positive fluctuation, but also model negative fluctuation. Yet, any correlation between positive and negative fluctuation has not been explored by early works. It is not known if positive and negative fluctuations can share the same distribution and parameters, or they must be modelled separately. This knowledge gap will be filled in this project.

2.2 Existing RTN Analysis Techniques

This section discusses the existing RTN analysis techniques for property extraction, detailing their advantages and limitations.

2.2.1 Voltage Histogram Method

One of the easiest approaches to extract the RTN amplitudes from a two-level RTN signal is plotting a histogram of the RTN signal [69-72]. This is known as voltage histogram method. The RTN amplitude of the two-level RTN is defined as the difference between the two normal distributions in the histogram in Fig. 2.7. It is also proven that it is able to extract multi-level RTN. A 4-level RTN data is shown in Fig. 2.8. After applying the voltage histogram methodology, 4 clear discrete levels of normal distribution can be observed as shown in Fig. 2.9 [73]. However, this method is unapplicable when RTN amplitude is close to system or thermal noise and also when multi-level RTN signals are relatively close to each other. This leads to unclear histogram plotting. Hence, this method is not reliable in many cases and a better tool for RTN property extraction is needed.

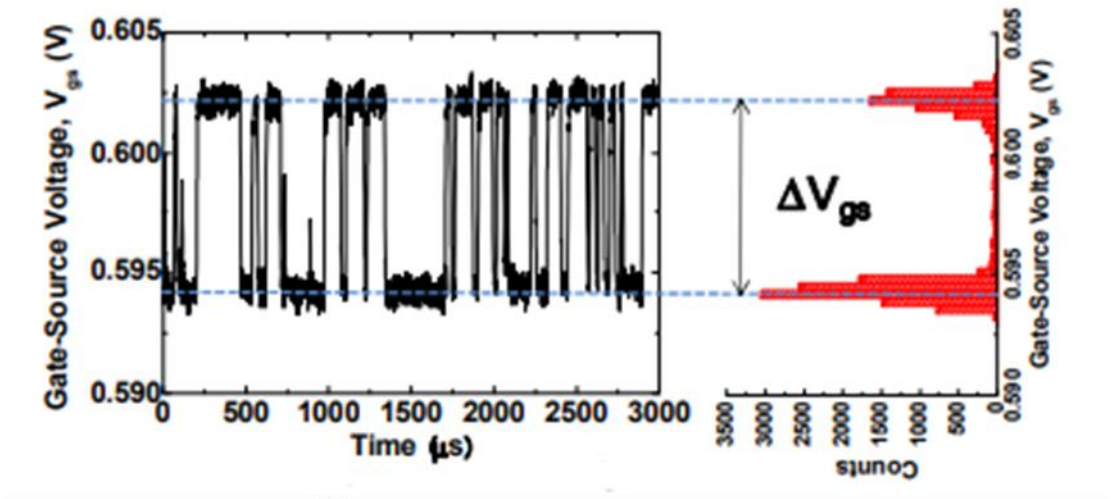


Fig. 2.7 RTN Amplitude extracted by voltage histogram method [69].

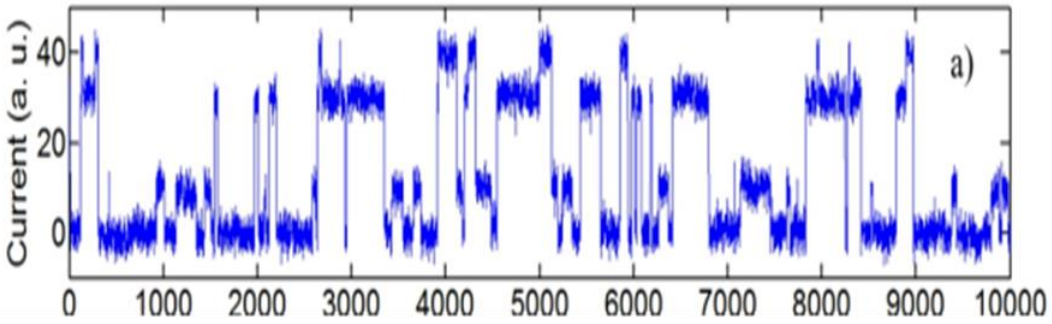


Fig. 2.8 RTN Amplitude extracted by voltage histogram method [69].

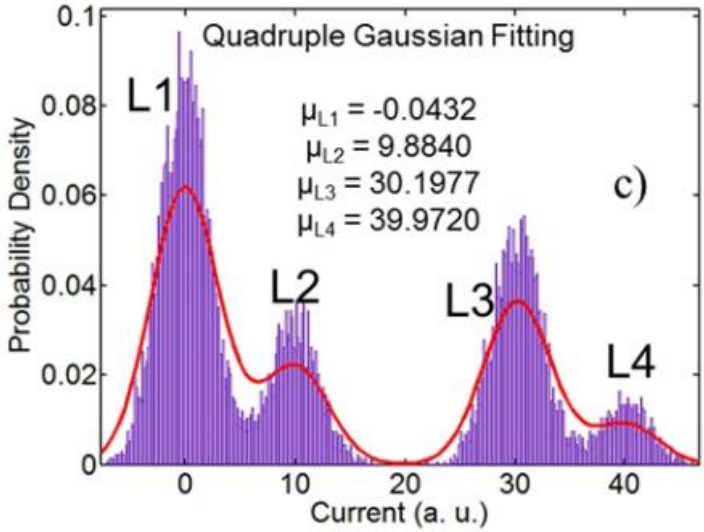


Fig. 2.9 4 level normal distribution fitting by voltage histogram method [73].

2.2.2 Time Lag Plot (TLP)

A clear two-level RTN signal is shown in Fig. 2.10(a) and the main task is to extract the properties of the RTN, despite system noise and thermal noise being included in the data. Early works [73-78] use a Time Lag Plot (TLP), which is also known as a lag plot, to analyse a two-level RTN signal in the time domain. TLP is used to check whether a set of data is randomly distributed or deterministic. It can be done by plotting the x-axis as the current i^{th} data at time interval, t_i, against the y-axis as the $i^{th}+1$ at time interval, t_{i+1} [79-80]. Fig. 2.10(b) represents the TLP graphical results of the two-level RTN signal in Fig. 2.10(a).

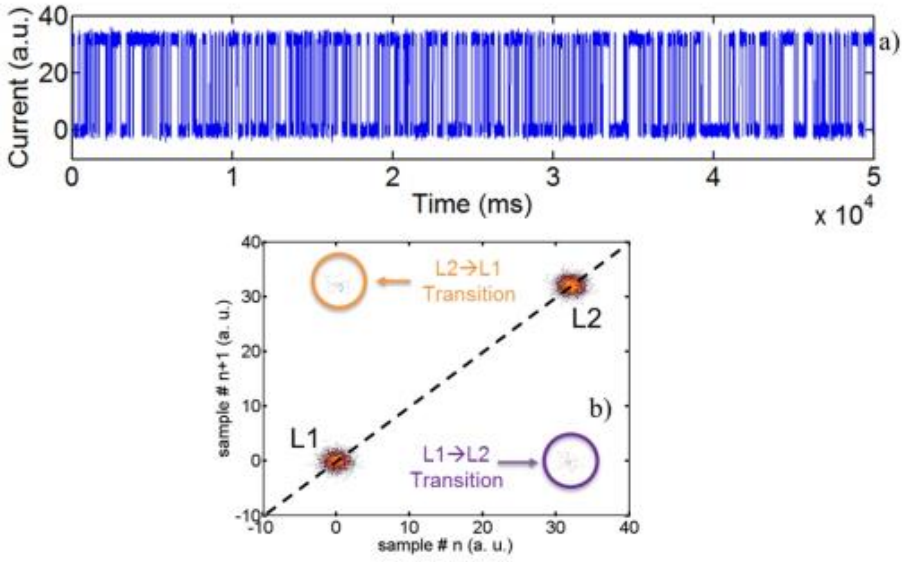


Fig. 2.10 (a) shows a two-level RTN signal. (b) TLP of the RTN data in (a) [73].

Based on Fig. 2.10(b), the amplitude of the RTN signal can be extracted from the difference between the centre point of each cluster, L1 and L2. TLP can be used only for RTN amplitude extraction, but not for time constant distribution. TLP only works for clear two-level traps. For a long-term window RTN measurement with complex RTN signals in Fig. 2.11, Fig. 2.12(a) shows that the cluster of TLP will overlap or no clear 2-level cluster is discernible, making TLP analysis impossible [79], [81]. Similarly, if an RTN amplitude is similar or close to the system and thermal noise level, the cluster of TLP will be overlapping. Hence, the accuracy of TLP is dependent on the number of traps and a trap amplitude that is much larger than the system noise and thermal noise level for a clear cluster.

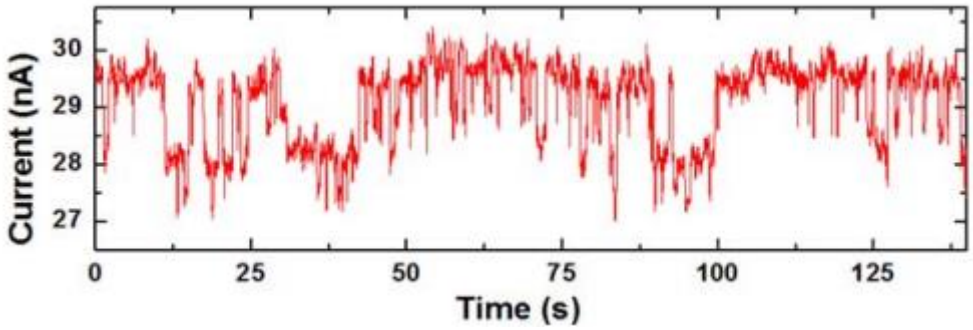


Fig. 2.11 An example of multi-levels RTN data [81].

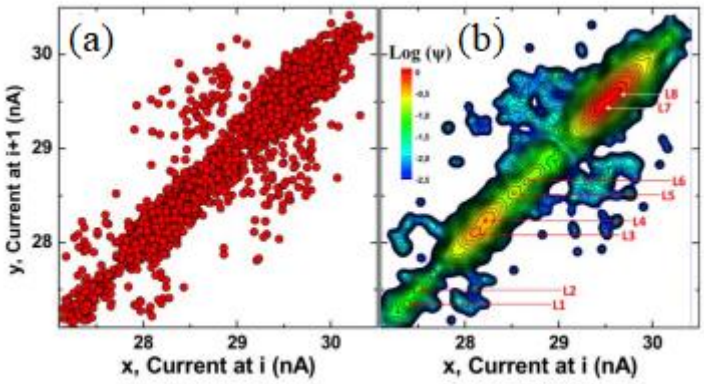


Fig. 2.12 RTN data in Figure 2.10 by TLP (a) and WTLP (b) [81].

2.2.3 Weighted Time Lag Plot (WTLP)

WTLP is proposed as an enhanced version of TLP to improve accuracy. It is designed to overcome the limitations of TLP by including the density of the cluster. This density helps to minimize the effects of other noises in the RTN signal, which in turn helps to identify clearer RTN levels in the cluster. It has been widely used in the past due to its easy implementation and noise allowance [76], [82-83]. Applied when multi-level RTN is spotted or an unclear signal is presented, WTLP can be used to identify whether any part of the signal has any shifting patterns. WTLP can be plotted the same as TLP with data, i^{th} at time, t_i and data, $i^{th} + 1$ at time, t_{i+1}. An example of a two-trap RTN signal is shown in Fig. 2.13(a) where TLP is applied in Fig. 2.13(b) and WTLP in Fig. 2.13(d). 4 different levels of current can be seen in the histogram of Fig. 2.13(c) and a density cluster by WTLP plotted in Fig. 2.13(d) gives the possibility to extract multi-level RTN traps.

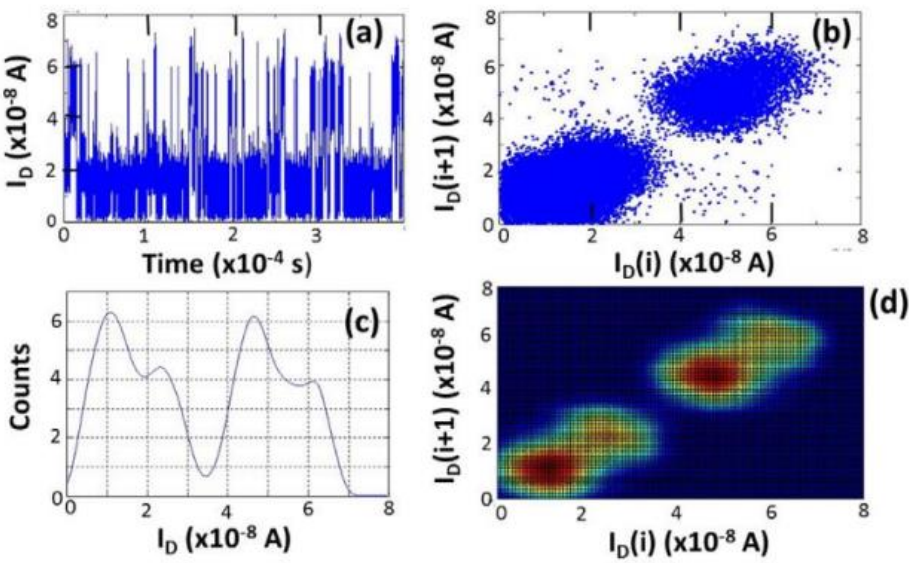


Fig. 2.13 (a) Time series generated by the hidden Markov routine in Matlab and containing a small and a large amplitude RTS; (b) corresponding simple TLP, showing only two clouds of data points; (c) amplitude histogram derived from the time domain signal and (d) coloured TLP, showing the presence of the two RTSs [83].

One of the limitations of WTLP is that the accuracy is highly dependent on the selection of centre point from the density cluster. WTLP can present confusing results on multi-cluster outputs. For example, in Fig. 2.12(b), the inter-correlated traps are hard to be separated and results can be controversial.

2.2.4 Hidden Markov Model (HMM)

HMM is the most popular and robust model used for RTN property extraction. HMM is a powerful tool in pattern recognition and statistical analysis where the input is modelled as a Markov Chain with hidden levels mixed in a noise signal [84]. In short, HMM is used to replicate the data or to generate an ideal RTN transition from noisy experimental data [15]. Fig. 2.14 shows a sample 2-level RTN trap with other noises included and fitted with an HMM waveform. The HMM waveforms represent the extracted RTN parameters in comparison to the test data. HMM can extract the amplitude and time constants of the single trap by the modelled waveform [66], [85]. The amplitude of the RTN trap is the difference between the fitted two levels. It is expected that the capture and emission time extracted from a single trap follow an exponential distribution.

However, HMM has limitations when dealing with multi-level RTN data [86] or noisy data such as when the noise levels are close to the RTN amplitude. This leads to a result in which some levels of RTN can be missed [87-88]. Another limitation of HMM is the input

requirement of the number of levels to be extracted. An incorrect assumption of the number of levels inputted into HMM leads to an incorrect HMM output.

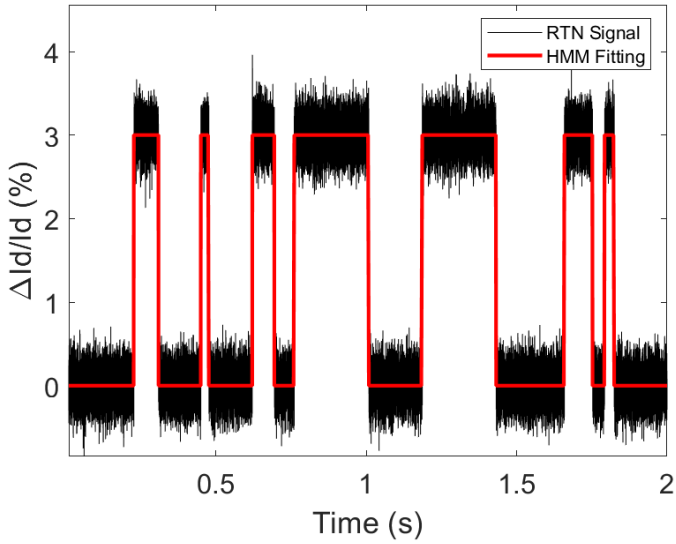


Fig. 2.14 2-levels RTN signal extracted by using HMM method.

2.2.5 Factorial Hidden Markov Model (FHMM)

FHMM is proposed as an improved version of HMM to tackle the limitations of HMM. FHMM decomposes multi-level RTN into multiple independent two-level RTN signals [89-91]. These multiple independent two-level RTN signals will be modelled with the HMM method to extract trap amplitude and time constants for each 2-level RTN. Fig. 2.15(a) shows a multiple-level RTN signal modelled by FHMM, resulting in the 3 individual two-level RTN signals in Fig. 2.15(b-d) [84]. It has been reported that HMM and FHMM provide greater accuracy compared to other analysing techniques [73].

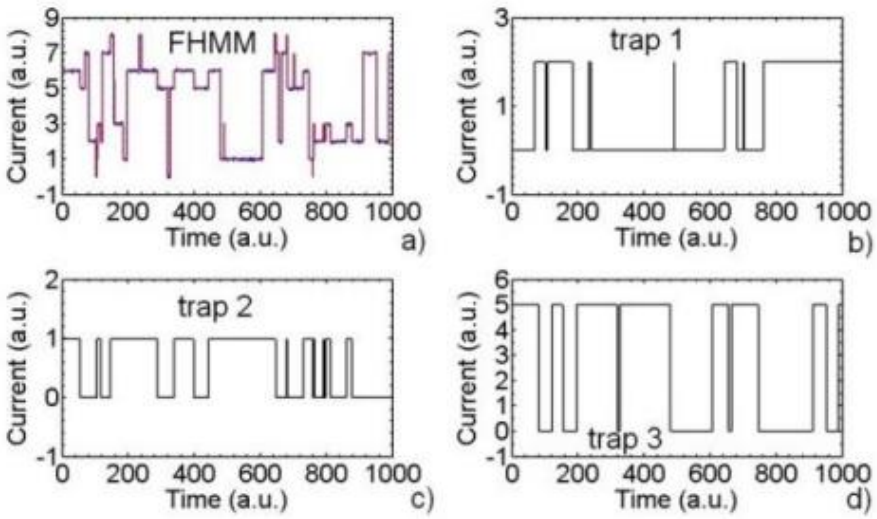


Fig. 2.15 (a) An eight-level RTN generated by three traps with superimposed additive gaussian noise (blue curve) and FHMM fitting (red curve). Accuracy is excellent. (b, c, d) Decomposed chains amplitudes and sequences are easily found, and trap characteristics can be inferred. [84]

HMM and FHMM share a similar limitation, which is the maximum number of data points of RTN measurement that can be used. A huge number of data points can cause FHMM and HMM processes to be too slow to obtain statistical results. Also, FHMM requires a pre-determined number of levels as an input to the model, which is not ideal for a noisy signal, where it can be difficult to pre-determine the number of levels. Despite being able to extract multi-level RTN, the accuracy of the extracted RTN properties decreases when the number of levels increases.

2.3 Summary

Although each of the proposed techniques has its advantages and a lot has been learnt from them, they have not led to a reliable and trustworthy RTN model. These techniques share a similarity: a Bottom-Up approach where single trap properties are modelled separately and then grouped as a set of data for statistical analysis. As RTN measurement data do not constantly show clear levels of RTN and the number of RTN traps in a transistor remains unknown, the limitations of the proposed techniques are summarised as follows:

1. Selection of only devices with clear RTN signals for extraction of properties, which is not a true statistical representation of real devices.
2. HMM provides high accuracy but a predefined number of levels is required as an input.
3. There are difficulties in extracting the time constant distribution.

These limitations lead to severe uncertainty over the accuracy of any RTN model derived from these techniques. To tackle these limitations, a top-down modelling technique is proposed and will be discussed in this project.

CHAPTER 3: DEVICES AND TEST FACILITIES

This chapter focuses on the devices, test facilities, and basic electrical characterisation methodology used in this work. The first section covers the laboratory instruments used and the process used for RTN measurements. The second gives the details of the wafer used for device measurement. The third describes the characterisation methodology. Lastly, the details of each measurement technique are given.

3.1 Laboratory Instruments and Flows of Measurement

An overview of the system setup is illustrated in Fig. 3.1. The instruments used in this work include a personal computer (PC) for software programming to control other instruments, a pulse generator (Agilent 81160A), an amplifier by Femto (DHPCA-100), a Parametric measurement mainframe unit (Agilent E5270A) with 4 source measure unit (SMU) slots, a home-made circuit for constant voltage supply, and a 4-channel oscilloscope (Agilent Infiniium MSO8104A). The probe station used in this research work is Cascade Summit with a controllable heating plate for desired temperature.

The measurement can be started by using a control program written in Visual Basic (VB) in the PC to send commands to instruments. The pulse generator, Agilent 81160A is used for gate voltage supply which is connected to the gate terminal of the Device Under Test (DUT). Drain voltage is set at a constant value, provided from the home-made circuit powered by a DC power supply. The drain current (I_D) is amplified and converted to voltage that is recorded by the oscilloscope, Agilent MSO8104A. Then, the data is saved and transferred to the PC. Before RTN measurement is started, a transfer characteristic (I_D-Vg) is recorded to make sure the connections and devices are working normally.

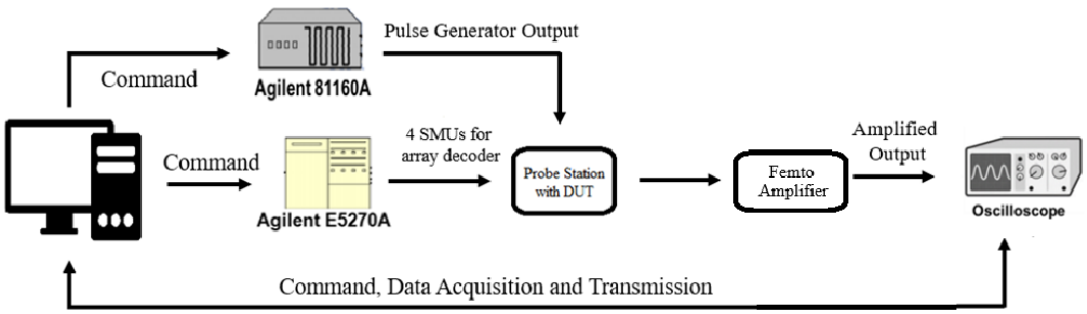


Fig. 3.1 Overview of the system setup.

3.1.1 Samples Used in this Research Work

The sample used for this research is a wafer from CSR plc (formerly Cambridge Silicon Radio), which was acquired by Qualcomm in August 2015. It is fabricated by Taiwan Semiconductor Manufacturing Company (TSMC) using its 28nm CMOS process with planar structure. It is a commercial process with proven quality and trustworthiness of the devices. The wafer has both array and individual device structures to optimise testing and analysis of ageing. Both nMOSFETs and pMOSFETs are used for this research work and it has metal gate and high-k/SiON stack. The device size used in this work is 90x27 nm in length and width. The average threshold voltage, Vth, is 0.45 V. The equivalent oxide thickness is 1.2 nm. The temperature applied for RTN measurement is 125ºC. An illustration of the CSR wafer is given in Fig. 3.2.

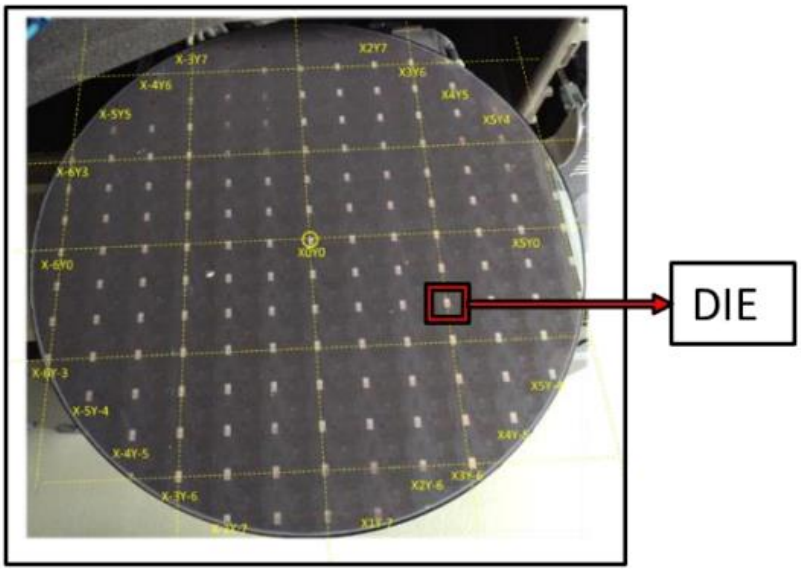


Fig. 3.2 CSR Wafer Sample [92].

For array devices, each die consists of several sizes, with 16 devices per size, where an individual device is at the centre of the die, as shown in Fig. 3.3. The array devices can be accessed by 4 SMUs from Agilent E5270A which act as a decoder of 4 bits from '0000' to '1111'. The decoder is activated by assigning 0.9 V; thus, '1' in the bit combination represents 0.9 V and '0' represents 0 V. An external supply of 0.9 V and a ground connection are needed to power up the decoder. This allows a large number of required devices to be measured for an accurate statistical analysis.

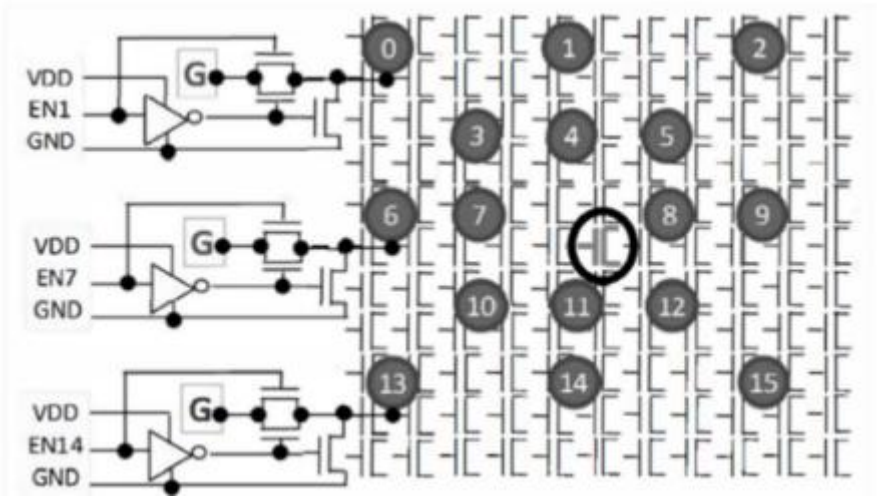


Fig. 3.3 The schematic representation of Decoder Enable lines connecting to transistors (numbered 0-15) and a centre transistor (circled) [93].

The connection of array must go through a probe card 78662 (Cantilever Probe Card) which is custom designed to probe these devices. In Fig. 3.4(b), the green printed circuit board (PCB) of the probe card with BNC connectors is designed to select a device size by connecting the corresponding gate, source, and drain on the connections of the die. Fig. 3.4 (a) shows the probe card for the probe pin connections to each die. Additionally, Table 3.1 gives the sizes of array transistors that are available in the CSR wafer and the connections to the PCB in Fig. 3.4(b).

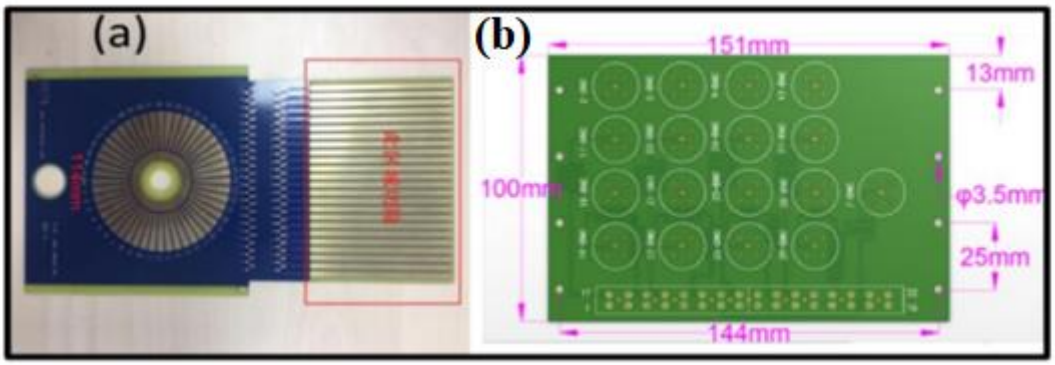


Fig. 3.4 The design of the probe card used in this research work [92].

Table 3.1 Sizes of array transistors in the CSR Wafer.

Device	Sizes		Pin Connections		
	Length (nm)	Width (nm)	Gate	Drain	Source
nMOS pMOS	27	90	23	1	2
	27	135	23	3	2
	27	180	23	4	5
	27	900	23	6	5
	36	900	23	8	9
	81	900	23	10	9
	225	900	23	11	12
	900	900	23	13	12

Additionally, the nMOS of the array devices can be switched on by directly applying the desired gate voltage, Vg, and drain voltage, Vd, as shown in Fig. 3.5(a). However, access to the pMOS is different in that the voltage drops between each terminal must be considered. It is due to the circuit designed for pMOS access of the array devices do not allow negative voltage to flow through. Hence, a special treatment is required to access the pMOS in the array. As shown in Fig. 3.5(b), to have -0.9 V and -0.1 V on the gate and drain respectively, 0.8 V needs to be applied to the drain, 0 V to the gate and 0.9 V to the source. This represents a voltage drop on the gate of -0.9 V, while the voltage drop on the drain is -0.1 V.

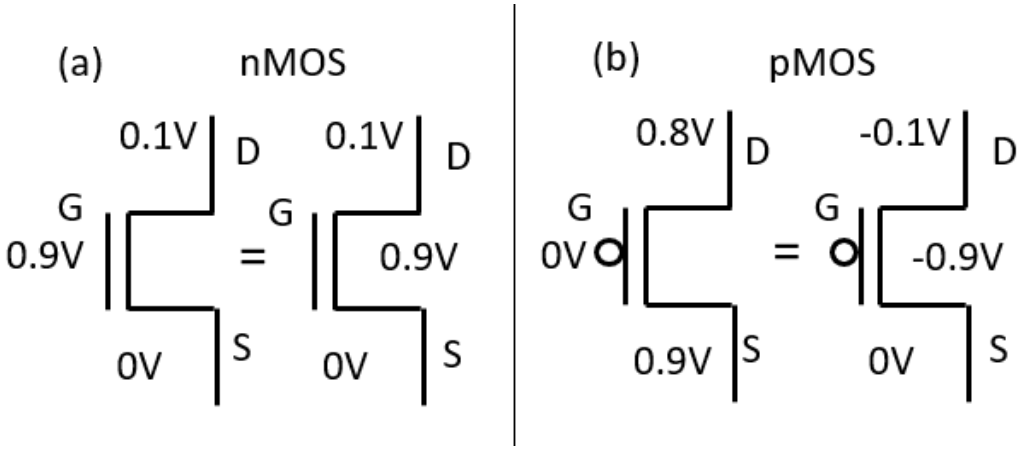


Fig. 3.5 An illustration of supply voltage to nMOS and pMOS for CSR devices.

3.2 Electrical Characterization of Transistors

This section discusses the measurements applied in this work such as I_D-Vg measurement and Vth extraction. It is important to measure the basic characteristics of a transistor correctly before proceeding to further tests.

3.2.1 Conventional I_D-Vg (IV) Characterization

IV measurement is an evaluation procedure for a device where bias voltages are applied to the terminals of the device while simultaneously measuring the current flowing through the terminals, commonly performed by using a commercial Source Measure Unit (SMU) [94]. An IV sweep is commonly performed by applying a varying Vg voltage under a constant drain bias, Vd. The typical waveform of a stepped IV sweep is given in Fig. 3.6. The typical measurement speed for a full IV sweep is approximately 1 to 10 seconds, depending on the step size applied. Hence, this method is commonly known as quasi-static or slow DC measurement.

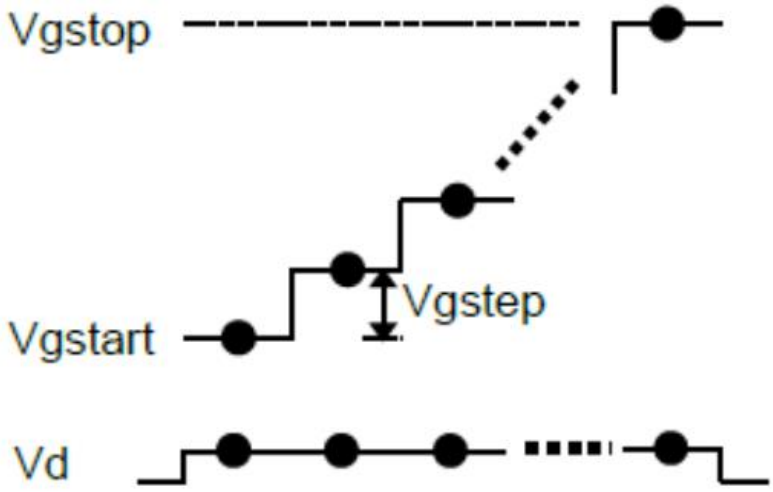


Fig. 3.6 Typical IV test waveform [95].

Pulse-IV measurement is a high-speed measurement (down to 3 µs) that enables full monitoring of the drain current, I_D, during a pulse on the gate of a transistor. A typical waveform of a Pulse IV measurement is shown in Fig. 3.7 [95]. The key advantage of pulse IV measurement is the very fast measurement speed. The high speed of the measurement is essential for cases where important aging or recovery mechanisms occur in the time frame that a traditional IV measurement would take to complete. The rapid speed of the pulse-IV

measurement allows for the assumption that no significant aging or recovery has occurred during the IV measurement, which would otherwise distort the results of after-stress measurements.

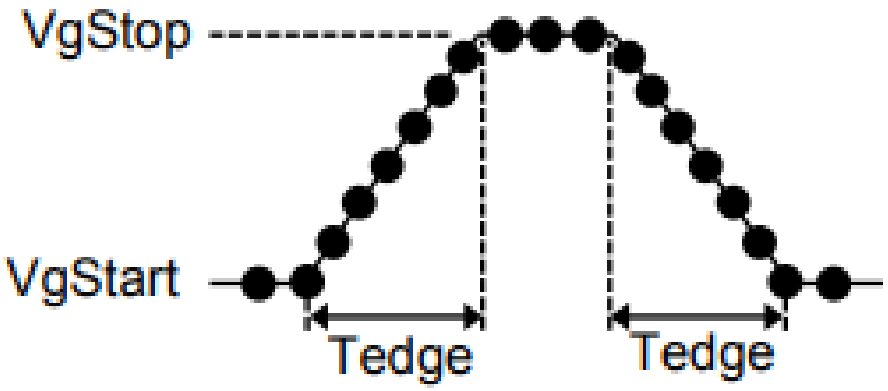


Fig. 3.7 Typical waveform of a Pulse IV measurement [95].

A single pulse IV is measured in 3 µs in this work. To have a sufficient number of points for a high accuracy of I_D-Vg extraction, 100 MSa/s is used. Within 3 µs, Vg is swept from 0 V to 0.95 V and the current, I_D, is recorded by an oscilloscope. The I_D-Vg data collected by the oscilloscope will be 300 data points, allocated uniformly from 0 to 0.95V Vg. To reduce the impact of system noise in the I_D-Vg measurement, the I_D-Vg measurement is measured 50 times [4], then the average of the 50 measurements is used as the reference/fresh I_D-Vg for the device. This will significantly reduce the impact of system noise to ~1 mV [4].

3.2.2 Threshold Voltage Extraction

One of the properties that can be extracted from an IV curve is the threshold voltage of the measured device. This should be compared with the datasheet given to verify that the measurement is correct. Many methods have been proposed [96] to extract the threshold voltage, Vth. Here, two widely used methods are discussed: the maximum transconductance ($Max - g_m$) method, and the constant current, Idcc, method. Max-g_m method is used in this research work for Vth extraction.

The Max-g_m method is a popular way to extract Vth from an IV sweep measurement [97]. The procedure to extract Vth by the maximum g_m is shown below:

1. Transconductance, g_m, is calculated by applying the first derivative to the slope of IV.

2. At the maximum of the g_m, the corresponding current is noted as I_D(max-g_m).
3. At I_D(max-g_m), a tangent line is drawn and the intercept point of the tangent line across the Vg (x-axis) is the threshold voltage, Vth of the device.

A typical waveform of Vth extraction is shown in Fig. 3.8 [95]. Apart from the max-g_m method, Idcc can be used to extract the threshold voltage, Vth. Idcc is widely recognized as $100nA * W/L$, where W and L are the channel width and length respectively [98]. The interception point from Idcc against the I_D-Vg curve at Vg is the extracted threshold voltage, Vth.

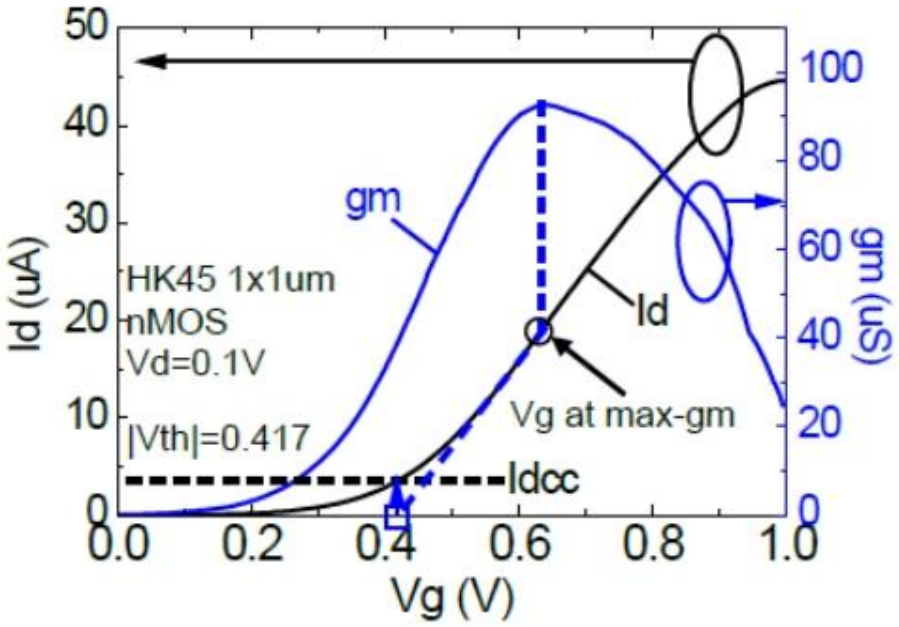


Fig. 3.8 Illustration of max-g_m and Idcc method for Vth extraction [95].

3.3 RTN Measurements

3.3.1 DC RTN Measurements

DC RTN measurement is carried out by recording the I_D of the device under constant gate and drain voltage. DC RTN allows extraction of the time constant, τc and τe, and the amplitude of a trap. The measurement pattern of a DC RTN measurement is given in Fig. 3.9. During the measurement, Vg is supplied by a pulse generator unit and Vd is supplied by a home-made circuit powered by a power supply unit. Vd is adjustable and 0.1 V is applied throughout all linear measurement. In an extension of this work, saturation measurement has been carried out under Vd = 0.9 V. I_D is observed by using an oscilloscope after amplification. The sampling rate used in this research work is 1 MSa/s which allows the detection of RTN traps with time constants down to the microseconds range.

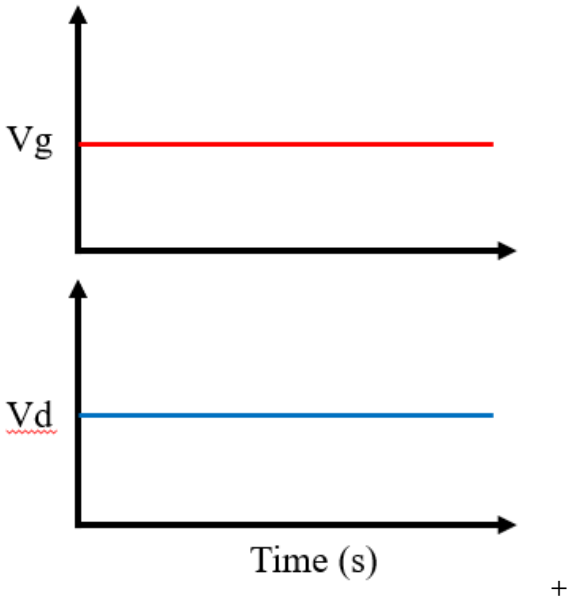


Fig. 3.9 Measurement Pattern of DC RTN Measurement.

The minimal time window measured is 10s and overnight measurement up to 64000 s is carried out by the methodology proposed by Mehedi *et al.* [30]. To overcome the hardware limitation of limited oscilloscope memory, the proposed overnight measurement is carried out by having a different storage rate of data point, compared to the sampling rate, in this case, Mehedi *et al.* stored every I_D at 0.1 s, upper (UE) and lower (LE) envelope at every 20 s while the sampling rate is 1 MSa/s, to overcome the oscilloscope reaching the memory limit, as shown in Fig. 3.10.

An improvement to instrumentation was made in this work, by replacing one of the oscilloscopes with a Digital Acquisition (DAQ) card, specifically a model manufactured by National Instruments (NI), NI – USB 6366. It can be programmed for data collection using the software supplied by NI.

The measurement starts by carrying out a pulse-IV and IV data is recorded. This is followed by preparing the oscilloscope for data recording and pulse generator for gate terminal bias changes on RTN measurement. The first 10 sec data is recorded and saved at 1 MSa/s by oscilloscope. Then, the RTN raw data after 10 sec are recorded and saved by DAQ at every 5 ms under 1 MSa/s until the measurement time hits 64k sec. 5 ms is selected because this is the minimum delay time for which the DAQ can save 1 data point from the display data before the data are updated; a time below 5 ms will lead to the possibility that overlapping data points will be saved.

The oscilloscope is used to detect the maximum and minimum points of the data displayed in the time screen of 2 sec at 1 MSa/s. Every 2 sec of data requires 5 sec of loading time before the data can be displayed in the oscilloscope's screen; hence, the max (UE) and the min (LE) are updated every 5 sec. If no higher or lower value of max and min is detected in the next 2 sec data, the max and min values will remain the same until there are values beyond the current saved values. The flow of modified overnight tests is given in Fig. 3.11.

In this work, an overall improvement on overnight RTN measurement is made by having the storage rate of raw I_D improved from 0.1 s to 5 ms and max, min from 20 s to 5 s. This improves the accuracy of continuous RTN data measurement with more data points available in the same measurement time.

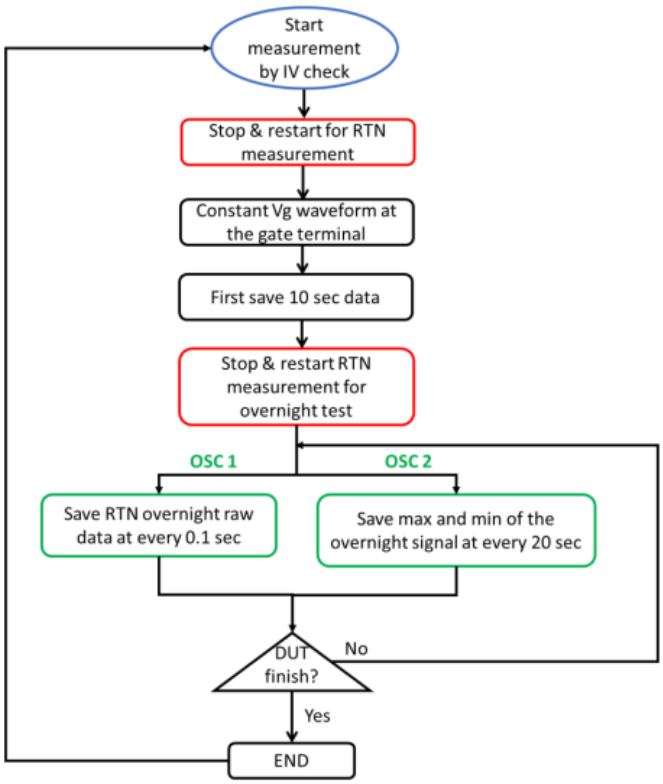


Fig. 3.10 Flowchart for the overall overnight test by Mehedi [92].

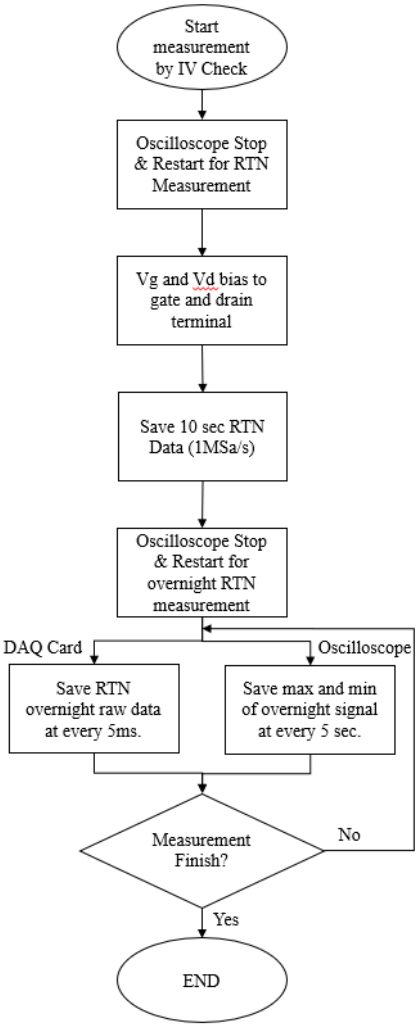


Fig. 3.11 Flowchart for the modified overnight test.

3.3.2 AC RTN Measurement

AC RTN measurement is carried out by alternating Vg between a high (VgH) and low (VgL) voltage under a constant Vd. The purpose of this measurement is to analyse RTN under conditions closer to realistic digital circuit operation, as these digital circuits are typically operated under AC conditions. A typical waveform of this measurement is given in Fig. 3.12. VgH values applied are 0.5 V and 0.9 V, where 0.5 V was chosen to be close to the threshold voltage, Vth, of 0.45 V for low-power applications and 0.9 V was chosen because it is the rated operating voltage for this technology. VgL is 0 V and Vd is 0.1 V throughout all the measurements. Frequencies selected on Vg in this work are 10 kHz, 100 kHz and 1 MHz.

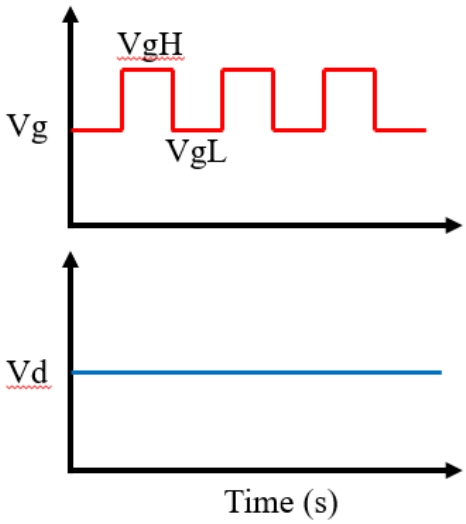


Fig. 3.12 Measurement Waveform of AC RTN Measurement.

I_D was recorded throughout the time window; however, the duration of VgL (0 V) is removed during data processing, as shown in Fig. 3.13. Hence, the time window for I_D under VgH is dependent on the duty factor of the frequency selected on Vg. In this work, 50% duty factor is selected, so that the time window for I_D under VgH is half of the total measurement time. The I_D current is amplified then measured using an oscilloscope. The sampling rate used in this research work is 1 MSa/s. The overnight measurement and data saving method is the same as DC RTN as shown in Fig. 3.11, except that the Vg bias in the AC RTN test is continuously switching.

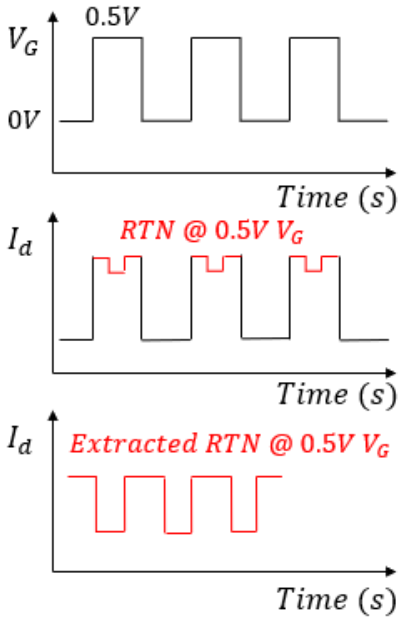


Fig. 3.13 I_D data processing on AC RTN Measurement.

3.4 ΔVth extraction by Pulse IV for ΔVth ($Vg = Vth$)

It has been proposed that the error of ΔVth estimated by $\Delta I_D/g_m$ can be substantial if Vg is well above Vth [4]. For measurement under Vg = 0.9V, ΔVth can be estimated from a stress pulse IV at stress time. The Vg waveform for measuring ΔVth is given in Fig. 3.14. Vg is stepped down from 0.9 V to 0.3 V, which acts as an IV sweep. To minimize de-trapping, the p-IV is sweeping at 0.9 V to 0.3 V instead of 0 V. Only 1 measurement of I_D-Vg is taken, unlike the reference I_D-Vg where 50 measurements are used. This is because the impact of thermal noise on ΔVth will be corrected through modelling.

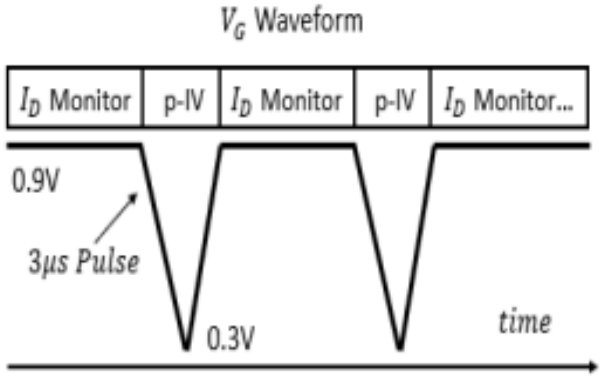


Fig. 3.14 Vg waveform for measuring ΔVth after stress.

The extraction of RTN-induced ΔVth under Vg = 0.9V is summarized below:

1. The Max-g_m method is used to extract the threshold voltage, Vth, of the device on the measured Fresh pulse-IV.
2. From the extracted Vth, the Idcc is determined by intercepting Vth across the IV curve towards the I_D (Y-axis).
3. The difference in Vg is determined between Fresh (Reference) and the I-V after RTN at Idcc (at Vth), as shown in Fig. 3.15.
4. This difference in Vg is the RTN-induced ΔVth at that time point.

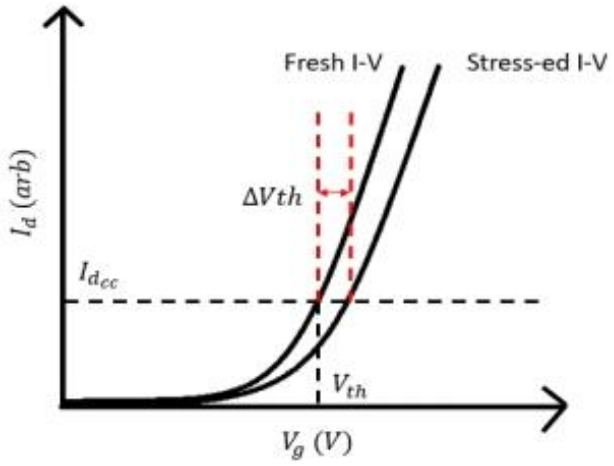


Fig. 3.15 ΔVth extraction with charged traps.

The technique represented in Fig. 3.15 contains two pulse-IV measurements for each iteration. First, the falling edge, 0.9 V to 0.3 V is observed, followed by the rising edge, 0.3 V to 0.9 V. The results are then compared to verify the method. The falling edge and rising edge can be compared to see if any trapping or de-trapping happens during the shifting down of Vg. Theoretically, both results should be very similar to each other. Two examples of the falling and rising edges are compared. As can be seen from Table 3.2, the difference between falling and rising edges is negligible. This indicates that de-trapping is negligible during a pulse-IV measurement.

Table 3.2 Comparison of ΔVth estimated by Falling and Rising Edge.

Comparison in ΔVth	Falling Edge (mV)	Rising Edge (mV)
Example 1	1.86	1.8
Example 2	3.79	3.84

3.5 Summary

In this chapter, the test instruments and test samples used in this research work are covered. An overview of the measurement setup is given. Next, the basic electrical characterization methodology is discussed, and a fast pulse-IV measurement is introduced. The methods for ΔVth extraction are also discussed.

Two methods are introduced for RTN measurement, DC RTN and AC RTN measurements, and experimental conditions are given. Additionally, an improvement is discussed for proposed

methods of overnight measurement. Lastly, the procedure for ΔVth extraction at $Vg = Vth$ is described.

CHAPTER 4: RTN MODEL BY INTEGRAL METHODOLOGY

This chapter focuses on the integral methodology. The first section covers the background of RTN modelling by early works. The second gives the structure of the integral methodology, including the inputs, outputs, and key function of the model. Lastly, the optimization of computation time and justification is also included.

4.1 Introduction

In early works, RTN modelling attempts were mainly focused on the characteristics of individual traps [3], [19], [31-33]; for example, the capture time, emission time and amplitude. This approach, however, has not led to a trustworthy model that can be used by circuit designers to analyse RTN for optimizing circuit performance. The main concern from a circuit designer's point of view is the total impact of RTN on current and threshold voltage, instead of the capture time, emission time and amplitude of each trap. Hence, this chapter focuses on developing a new methodology that can directly give the probability of a certain ΔVth to circuit designers or end users.

4.2 The Integral Methodology

This section covers the details of the integral methodology, starting with an overview, followed by a detailed flowchart of the procedure. Details such as the structure of the method, the inputs, key functions, and the outputs are discussed. The method for optimizing the computation time is presented. The integral method is implemented by a Matlab-based program.

4.2.1 Structure of the Integral Methodology

In the early works, individual traps were characterized by 3 important parameters,

- The impact of traps, i.e., the amplitude of parameter shift induced by individual traps,
- The time constants of individual traps,
- Number of traps per device.

The main aim of this RTN model is to estimate the total impact of multiple traps combined at any given time point. To facilitate the statistical analysis of RTN without obtaining these properties of individual traps, the concept of Effective Charged Traps (ECT) is introduced. At a given time point, the RTN-induced ΔVth measured from multiple devices will form a dataset that has a statistical distribution. It is assumed that this statistical distribution originates from a

group of traps that are always charged. In another words, ECT is the equivalent set of fixed charges that give the same statistical distribution of ΔVth as measured from multiple real devices where traps dynamically capture and emit charges. Although the ΔVth of one device at a given time can be stochastic, the statistical distribution of the set of ΔVth obtained from multiple devices will be stable when the number of devices is sufficiently large, as shown by Fig. 4.1 below. As a result, one can assume that this statistical distribution originates from a set of ECTs that are always charged.

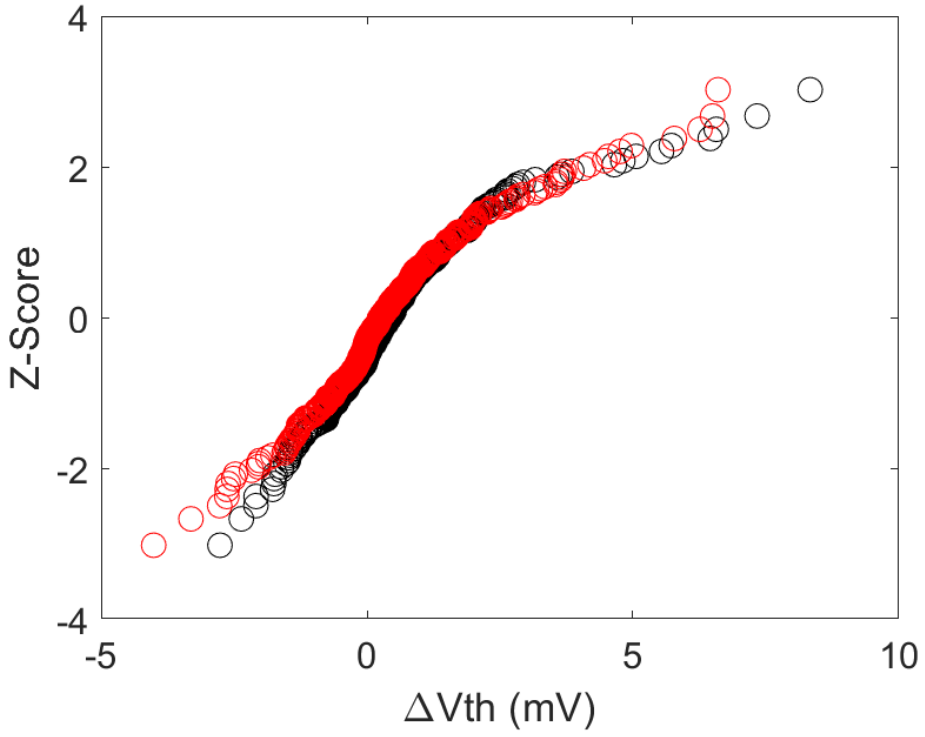


Fig. 4.1 ΔVth is measured twice and there is little difference between the two set of data.

By assuming there are both acceptor-like and donor-like traps in a device, the RTN-induced ΔVth becomes,

$$\Delta Vth_{RTN} = \Delta Vth_A + \Delta Vth_D \tag{4.1}$$

where ΔVth_A represents the contribution of acceptor-like traps and ΔVth_D represents the contribution of donor-like traps. For nMOSFETs, $\Delta Vth > 0$ is used for acceptor-like traps and $\Delta Vth < 0$ for donor-like traps. As ECTs are always charged, it enables modelling RTN without obtaining the capture and emission times for each trap. The model based on ECT requires the number of traps per device, n, and the statistical distribution of impact per trap, i.e., δVth.

Early works [1], [3], [19], [25], [31-33], [96], [99] proposed 3 types of distributions for δVth:

1. Exponential Distribution
2. Lognormal Distribution
3. Generalized Extreme Value (GEV) Distribution

These 3 distributions will be implemented into the integral methodology and compared between each other for the selection of distributions. These probability distribution functions (pdf) for Acceptor-like traps can be generally represented by

$$pdf_A(\delta Vth_A) \tag{4.2}$$

As mentioned earlier, the total impact of RTN on a device at a time point, ΔVth_A, is the key information needed. When there are n_A acceptor-like traps in a device, ΔVth_A can be evaluated from

$$\Delta Vth_A = \delta Vth_1 + \delta Vth_2 + \ldots. + \delta Vth_{n_A} \tag{4.3}$$

Mathematically, if Z = X + Y and X and Y are two independent random variables with pdf_X and pdf_Y, respectively, the pdf of Z is the convolution of pdf_X and pdf_Y [100]:

$$pdf(z) = \int_{-\infty}^{\infty} pdf_X(x) pdf_Y(z-x) dx \tag{4.4}$$

As a result, if there are two acceptor-like traps in a device, we have

$$pdf_{A2}(\Delta Vth_{A2}) = \int_{-\infty}^{\infty} pdf(\delta Vth_1) pdf(\Delta Vth_{A2} - \delta Vth_1) d(\delta Vth_1) \tag{4.5}$$

If there are 3 traps, the pdf of ΔVth_{A3} should be a convolution of $pdf_{A2}(\Delta Vth_{A2})$ with $pdf(\delta Vth_3)$:

$$pdf_{A3}(\Delta Vth_{A3}) = \int_{-\infty}^{\infty} pdf_{A2}(\Delta Vth_{A2}) pdf(\Delta Vth_{A3} - \Delta Vth_{A2}) d(\Delta Vth_{A2}) \tag{4.6}$$

Similarly, if we have n_A acceptor-like traps, we can convolute one trap at a time until all traps are included. The pdf of ΔVth_A from n_A traps can be represented by

$$pdf_A(\Delta Vth_A) = Conv\left[pdf(\delta Vth_1), pdf(\delta Vth_2) \ldots pdf(\delta Vth_{n_A})\right] \tag{4.7}$$

As shown in Equation (4.1), there are two types of traps to be modelled: acceptor-like and donor-like. Equation (4.7) gives the pdf acceptor-like traps, ΔVth_A. For donor-like traps, ΔVth_D is negative, but both the Exponential and Lognormal distribution require their

parameters to be positive. To overcome this difficulty, the absolute value of δVth_D will be used,

$$|\Delta Vth_D| = |\delta Vth_1| + |\delta Vth_2| + \ldots + |\delta Vth_{n_D}|. \tag{4.8}$$

The pdf of $|\Delta Vth_D|$ can be evaluated similarly:

$$pdf_D(|\Delta Vth_D|) = Conv\left[pdf(|\delta Vth_1|), pdf(|\delta Vth_2|) \ldots pdf\left(\left|\delta Vth_{n_D}\right|\right)\right] \tag{4.9}$$

Acceptor-like traps reflect positive fluctuation and donor-like traps reflect negative fluctuation in ΔVth. Thus, the total impact of all traps in a device can be evaluated by:

$$\Delta Vth_{RTN} = \Delta Vth_A - |\Delta Vth_D| \tag{4.10}$$

Also, the pdf of ΔVth_{RTN} can be expressed as:

$$pdf[\Delta Vth_{RTN}(n_A, n_D)]$$
$$= \int_{-\infty}^{\infty} pdf_A[\left(\Delta Vth_{RTN}(n_A, n_D) + \delta Vth_{n_D}\right] pdf_D\left(\delta Vth_{n_D}\right) d(\delta Vth_{n_D}) \tag{4.11}$$

Equation (4.11) gives the probability that a device will have ΔVth_{RTN}, if every device has n_A acceptor-like traps and n_D donor-like traps. In reality, different devices will have different numbers of traps and it is well accepted that the number of traps per device follows a Poisson distribution. This leads to:

$$pdf(\Delta Vth) = \sum_{n_A=0}^{\infty} \sum_{n_D=0}^{\infty} \frac{e^{-N_A} N_A^{n_A}}{n_A!} \frac{e^{-N_D} N_D^{n_D}}{n_D!} pdf\left(\Delta Vth_{RTN}(n_A, n_D)\right), \tag{4.12}$$

where N_A and N_D are the average number of acceptor-like and donor-like traps per device, respectively.

If only one type of trap is considered, for example, acceptor-like traps, the equivalent equation can be transformed into:

$$pdf(\Delta Vth) = \sum_{n_A=0}^{\infty} \frac{e^{-N_A} N_A^{n_A}}{n_A!} pdf\left(\Delta Vth_{RTN_A}(n_A)\right) \tag{4.13}$$

Equation (4.13) shows a summation symbol of the Poisson distribution multiplied by $pdf\left(\Delta Vth_{RTN_A}(n_A)\right)$. It equals

$$Sum\left(\frac{e^{-N_A}N_A^0}{0!}pdf\left(\Delta Vth(0)\right)+\frac{e^{-N_A}N_A^1}{1!}\left(pdf\left(\Delta Vth(1)\right)\right.\right.$$
$$\left.\left.+\frac{e^{-N_A}N_A^2}{2!}pdf\left(\Delta Vth(2)\right)+..+\frac{e^{-N_A}N_A^{n_A}}{n_A!}pdf\left(\Delta Vth(n_A)\right)+\cdots\right)\right. \quad (4.14)$$

Examples of Matlab commands for the methodology can be found in the Appendix. A Tutorial example is also given at the end of this chapter. Table 4.1 gives the lower and upper boundaries applied in the numerical integration in Equation 4.5 and Equation 4.11. Derivation of Equation 4.4 and justification of these boundaries can be found in the Appendix.

Furthermore, thermal noise is a common noise source in circuits, and it must be taken into account. Thermal noise is commonly known to follow a Normal distribution with a zero average value. Its effect can be added into Equation (4.11) by carrying out a convolution,

$$pdf(\Delta Vth) = Conv[pdf(\Delta Vth_{RTN}), pdf(\Delta Vth_{Thermal})]. \quad (4.15)$$

Fig. 4.2 shows the flow chart of the integral methodology. Fig. 4.3 illustrates how to apply the integral methodology to a dataset with required inputs and the detailed working operation of the model.

Table 4.1 The lower and upper boundaries of each distribution for convolution.

Distributions	Boundaries for Numerical Integration					
	Acceptor		Acceptor + Donor			
	Lower-Boundary	Upper-Boundary	Lower-Boundary		Upper-Boundary	
Exponential	0	ΔVth_A	$\Delta Vth < 0$	$\Delta Vth > 0$	$\Delta Vth < 0$	$\Delta Vth > 0$
			0	ΔVth	Inf	Inf
Lognormal	0	ΔVth_A	$\Delta Vth < 0$	$\Delta Vth > 0$	$\Delta Vth < 0$	$\Delta Vth > 0$
			0	ΔVth	Inf	Inf
GEV	t_A $t_A = \alpha_A - \frac{\beta_A}{\xi_A}$	$\Delta Vth_A - t_A$	$\Delta Vth < (t_A - t_D)$	$\Delta Vth > (t_A - t_D)$	$\Delta Vth < (t_A - t_D)$	$\Delta Vth > (t_A - t_D)$
			t_A	$\Delta Vth + t_D$	Inf	Inf

Once the statistical distribution per trap is selected, its parameters will be fitted with the measured dataset. Maximum Likelihood Estimation (MLE) is used for the fitting [30, 101-102]. The parameters to be extracted are:

1) Average number of effective Acceptor-like traps per device, N_A
2) Average number of effective Donor-like traps per device, N_D
3) Parameters for δVth distribution:
    a. Exponential distribution
        i. $\eta_A \; and \; \eta_D$
    b. Lognormal distribution
        i. $\epsilon_A, \theta_A \; and \; \epsilon_D, \theta_D$
    c. Generalized Extreme Value distribution.
        i. ξ_A, α_A, β_A and ξ_D, α_D, β_D
4) The standard deviation of thermal noise, σ.

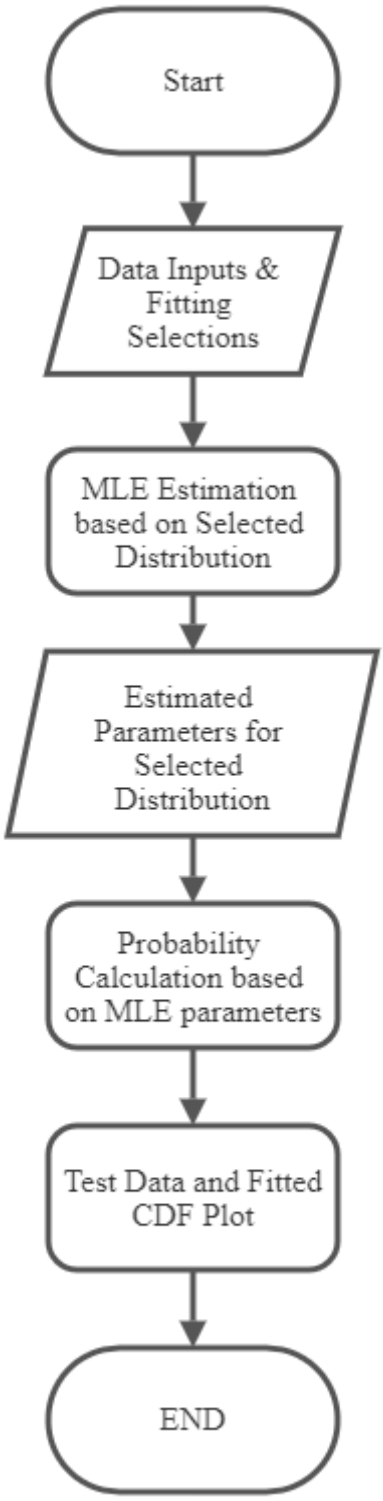
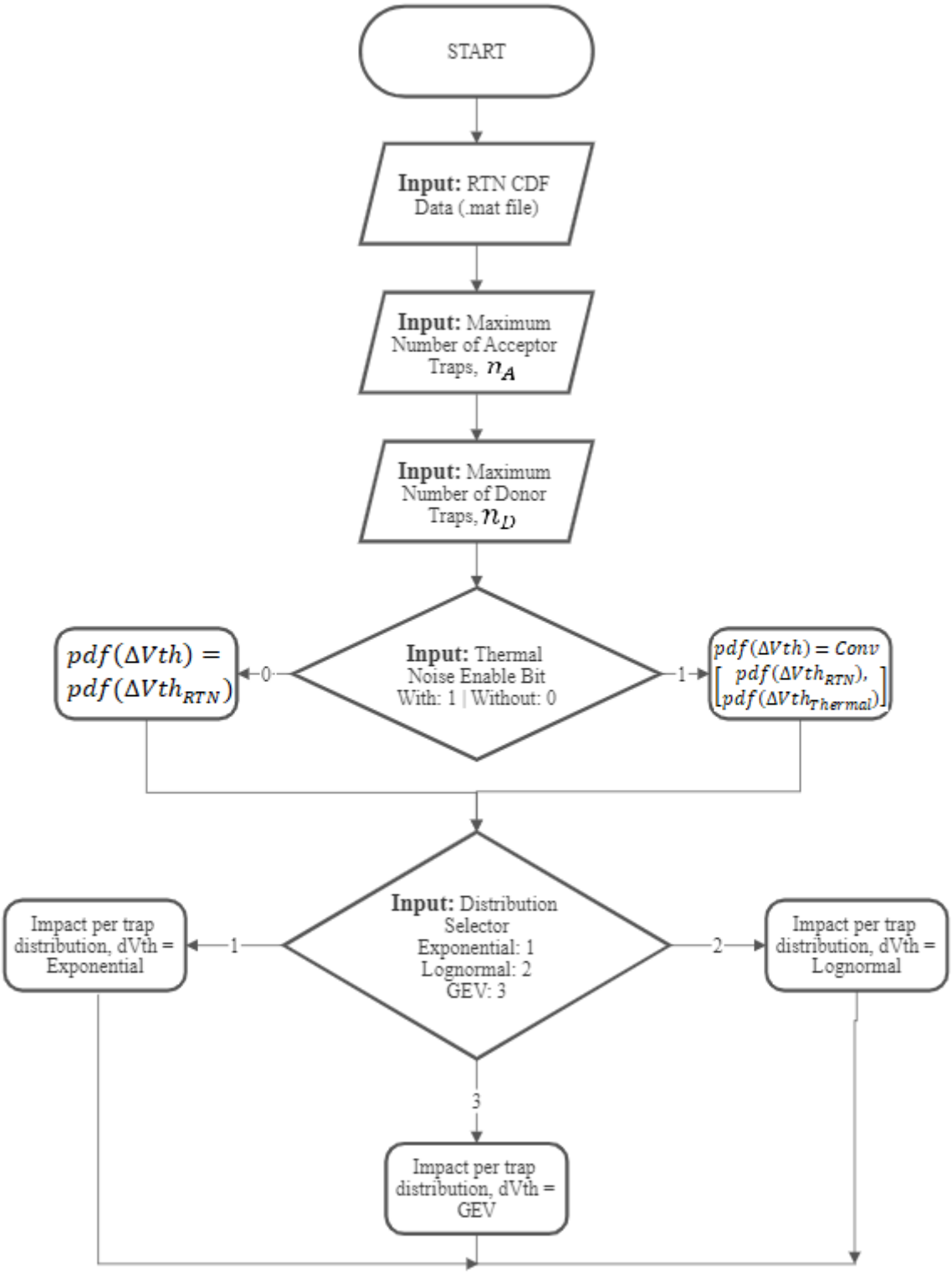


Fig. 4.2 A flow chart of the integral methodology.


START
Input: RTN CDF Data (.mat file)
Input: Maximum Number of Acceptor Traps, n_A
Input: Maximum Number of Donor Traps, n_D
Input: Thermal Noise Enable Bit With: 1 | Without: 0
0
1
$pdf(\Delta Vth) = pdf(\Delta Vth_{RTN})$
$pdf(\Delta Vth) = Conv\left[pdf(\Delta Vth_{RTN}), pdf(\Delta Vth_{Thermal})\right]$
Input: Distribution Selector Exponential: 1 Lognormal: 2 GEV: 3
1
2
3
Impact per trap distribution, dVth = Exponential
Impact per trap distribution, dVth = Lognormal
Impact per trap distribution, dVth = GEV

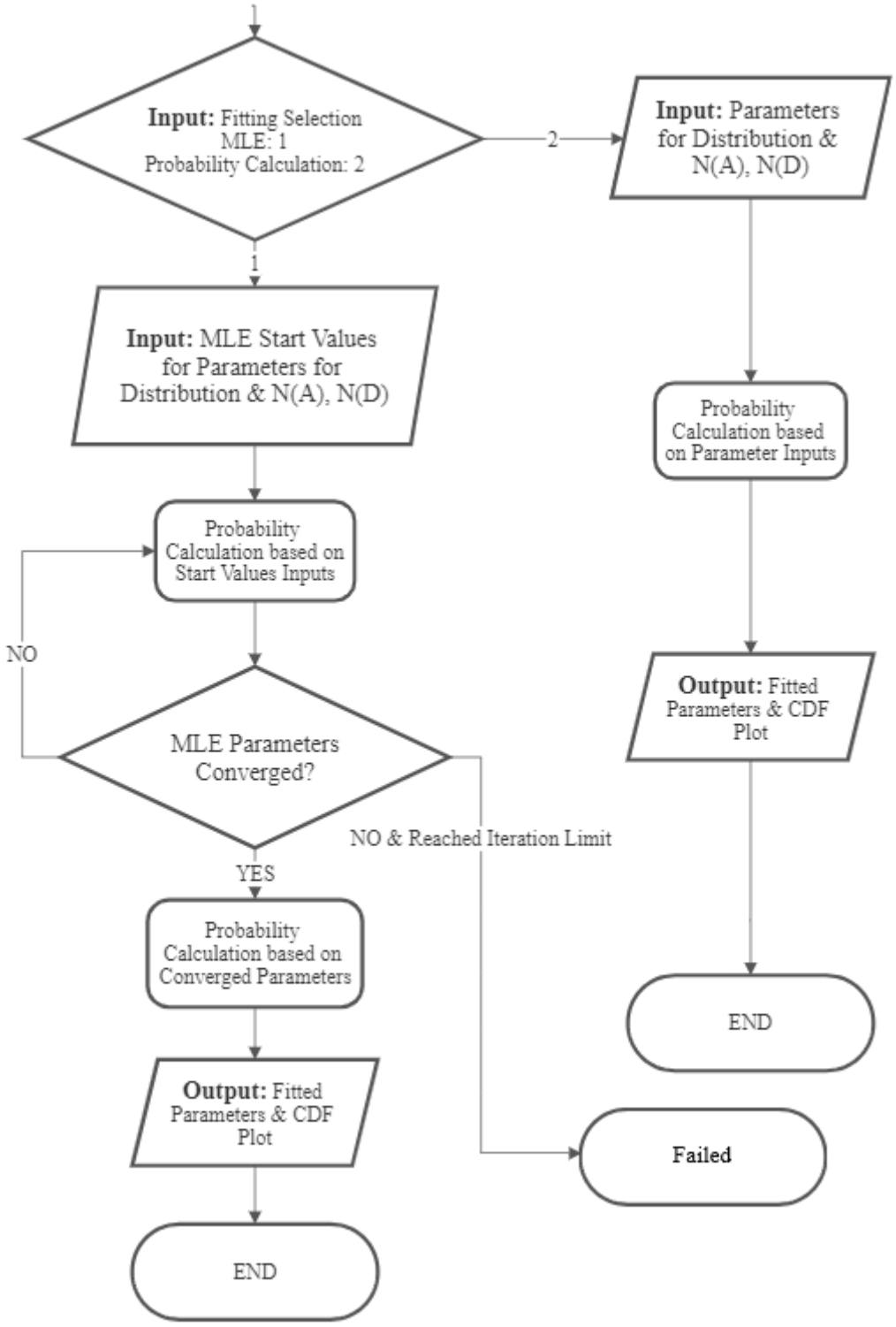


Fig. 4.3 A detailed illustration of the Matlab program with required inputs.

The analytical formulas for the pdf of the three distributions are given in Table 4.2:

Table 4.2 The PDF of δVth and parameters to be fitted for each PDF.

	PDF of δVth	Parameters to be fitted	
Exponential	$\frac{1}{\eta} e^{-\frac{\delta Vth}{\eta}} \quad \delta Vth > 0$ $0 \quad \delta Vth < 0$ $\delta Vth \in [0, \infty]$	η_A	η_D
Lognormal	$\frac{1}{\delta Vth \theta \sqrt{2\pi}} e^{\left(-\frac{(\ln(\delta Vth)-\epsilon)^2}{2\theta^2}\right)} \quad \delta Vth > 0$ $0 \quad \delta Vth < 0$ $\delta Vth \in [0, \infty]$	ϵ_A θ_A	ϵ_D θ_D
GEV	$\frac{1}{\beta}(k)^{\xi+1} e^{-k} \quad \delta Vth > t$ $0 \quad \delta Vth < t$ where, $k = \left(1 + \xi\left(\frac{\delta Vth - \alpha}{\beta}\right)\right)^{-\frac{1}{\xi}},$ $t = \alpha - \frac{\beta}{\xi}.$ $\delta Vth \in [t, \infty]$	ξ_A α_A β_A	ξ_D α_D β_D
Thermal	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\Delta Vth}{\sigma}\right)^2} \quad \Delta Vth = Real\ Number$ $\delta Vth \in [-\infty, \infty]$	σ	

The numerical convolutions are performed in pdf. They are then integrated to give the CDF, which is then fitted with the measured CDF.

4.2.2 Inputs for the Model

First, an RTN dataset presented in CDF form is needed, as shown in Fig. 4.4. Every single point of the dataset represents a ΔVth measured from a device, which includes the impacts of all traps. By combining these ΔVth from multiple devices in ascending order into one dataset, they can be presented as an empirical CDF plot by calculating the probability as [103]

$$CDF(\Delta Vth, i) = (i - 0.5)/N \tag{4.16}$$

where i = 1 has the lowest ΔVth and i = N (= 402, in this work) has the highest ΔVth. To illustrate the tail region clearly, the same dataset can be plotted in Z-score in Fig. 4.5. The Z-score can be obtained by looking at a Z-Score table or using a normal inverse cumulative calculator based on the CDF values from Equation (4.16).

To prevent the Z-score approaching infinity when CDF approaches 1, The '-0.5' is used in Equation (4.16), so that CDF does not reach 1, when i=N. It should be pointed out that the accuracy of the Z-score calculated in this way becomes poor, as i approaches N, as '-0.5' was artificially added.

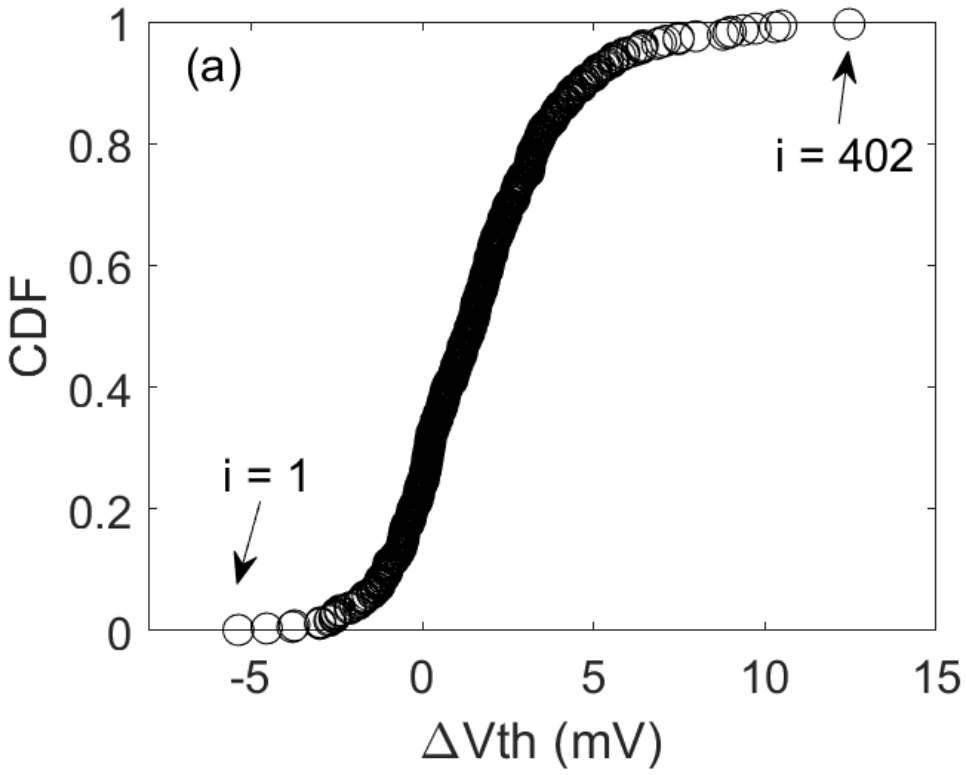


Fig. 4.4 A CDF of RTN dataset from multiple devices in ΔVth.

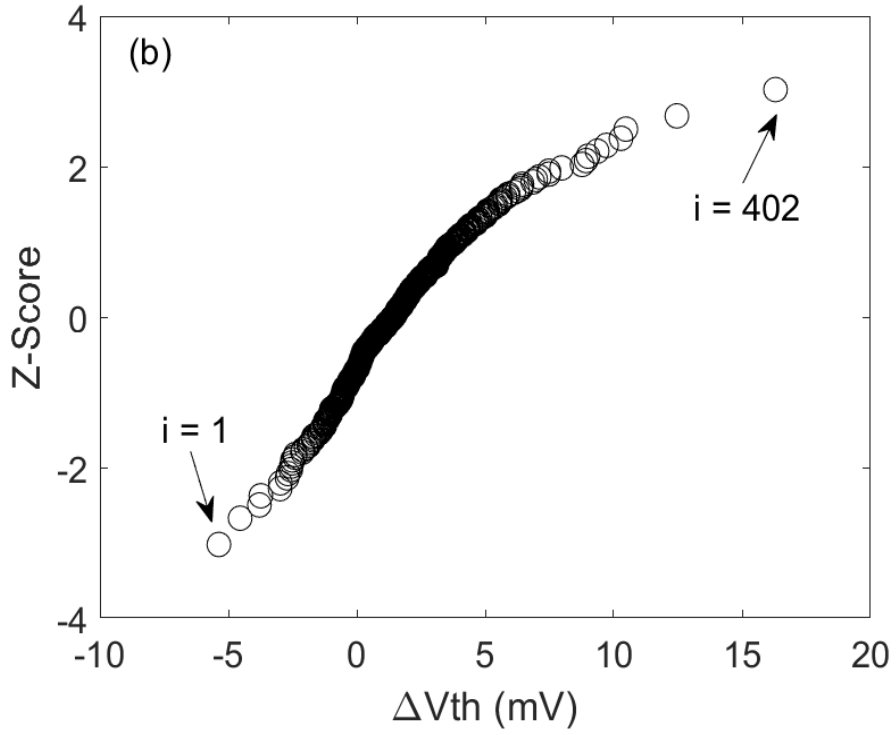


Fig. 4.5 A RTN dataset from multiple devices in ΔVth plotted in Z-Score.

The required inputs are shown in Fig. 4.6 and listed below:

i. CDF of RTN Dataset to be fitted: the name of the file containing the dataset should be given.
ii. Maximum number of Traps for Acceptor, $n_{A,}$
iii. Maximum number of Traps for Donor, $n_{D,}$
iv. Thermal Noise Enable Bit: 0 | 1,
v. Distribution Selector: Exponential | Lognormal | GEV,
vi. Maximum Likelihood Estimation | Probability Calculation Selector
vii. Parameter Start values for Maximum Likelihood Estimation or Known parameters for probability calculation.

```
Please enter the data filename: 60DevicesExample
Please enter the maximum number of Traps for Acceptor (nA): 4
Please enter the maximum number of Traps for Donor (nD): 4
Thermal Noise: With (1) | Without (0): 0
Please select distribution: Expo (1) | Lognormal (2)| GEV (3): 1
Please select MLE Fit (1) | Probability Calculation (2): 1
Please enter the start values in [mu(A) mu(D) N(A) N(D)]: [0.65 0.65 1 1.2]
Elapsed time is 455.086147 seconds.
```

Fig. 4.6 An example of required inputs in Matlab.

Fig. 4.6 gives an example for the inputs required in Matlab for the integral methodology.

- Maximum number of Traps for Acceptor (n_A) / Donor (n_D) decides the number of items to be calculated for each ΔVth in Equation 4.12. In principle, they should be infinite, but they must be finite in practice. It is found that they must at least double the average number of traps per device, as justified in Section 4.2.5. In Fig. 4.6, the maximum number of traps is 4 for both Acceptors (n_A) and Donors (n_D).
- The flexibility of including thermal noise has been included. Thermal noise fitting can be enabled or disabled by giving '1' or '0', respectively, to input (iv).
- The impact per trap distributions in this model include Exponential, Lognormal and GEV distributions, which can be selected by giving '1', '2' or '3', respectively, to input (v).
- For input (vi), one can select between parameter fitting by MLE or the probability calculation with a known set of parameters.
- In input (vii),
    - For an MLE extraction, start values of the parameters are required to initialize the MLE process.
    - For probability calculation, the parameters input by users will be directly applied to the equation to calculate the probability.

In Fig. 4.6, Exponential distribution and MLE estimation are selected. The start values of the Exponential distribution for the MLE estimation are,

I. $\eta_A = 0.65$,
II. $\eta_D = 0.65$,
III. $N_A = 1$,
IV. $N_D = 1.2$.

The $\eta_A \; and \; \eta_D$ are the average δVth per acceptor-like and donor-like trap, respectively, as given in Table 4.2. The parameters needed for the Lognormal and GEV distributions are also given in Table 4.2. The $N_A \; and \; N_D$ are the start values of the average number of acceptor-like and donor-like traps per device, respectively.

4.2.3 Outputs of the Model

For any model it is important to verify that it is working as expected. This can be done by using the RTN model to fit a known parameter set of data which is generated by Monte Carlo simulation. The hypothetical dataset in Fig. 4.7 is generated by using,

1. $Average\ number\ of\ acceptor-like\ traps, N_A = 2.5$
2. $Average\ number\ of\ donor-like\ traps, N_D = 1.5$
3. $Average\ \delta Vth\ per\ acceptor-like\ trap = 1$ mV
4. $Average\ \delta Vth\ per\ donor-like\ trap = 0.5$ mV

The exponential distribution was used for the impact per trap. With the start values for parameters in Fig. 4.6 directly fed into the integral model, the fitted probability by MLE is plotted as the red line in Fig. 4.7. The calculated probability fits the data very well. The extracted parameters by MLE are:

1. $Fitted: Average\ number\ of\ acceptor-like\ traps, N_A = 2.5$
2. $Fitted: Average\ number\ of\ donor-like\ traps, N_D = 1.4$
3. $Fitted: Average\ \delta Vth\ per\ acceptor-like\ trap = 0.9\ mV$
4. $Fitted: Average\ \delta Vth\ per\ donor-like\ trap = 0.6$ mV

These values agree well with the real values for generating the dataset and validate that the proposed integral methodology can be used to extract the model parameters.

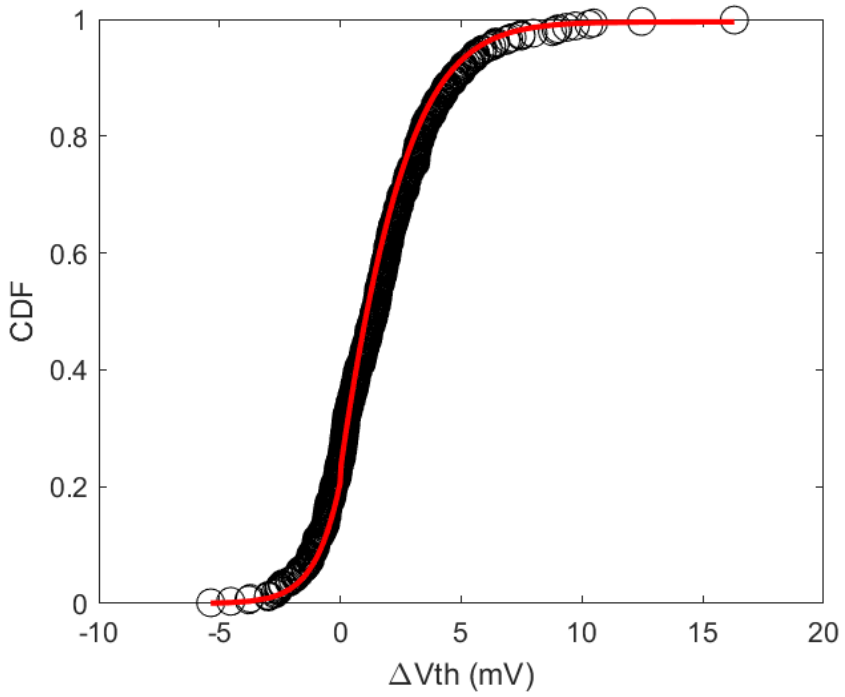


Fig. 4.7 Verification of the RTN integral model by using a Monte Carlo generated set of data.

From the extracted parameters, the arithmetic mean, and standard deviation of each distribution can be evaluated from [104-106],

1) Exponential

$$Arithmetic\ Mean = \eta$$

$$Standard\ Deviation = \eta$$

2) Lognormal

$$Arithmetic\ Mean = \exp\left(\epsilon + \frac{\theta^2}{2}\right)$$

$$Standard\ Deviation = e^{\epsilon+\frac{1}{2}\theta^2}\sqrt{e^{\theta^2}-1}$$

3) GEV

$$Arithmetic\ Mean = \alpha + \frac{\beta[\Gamma(1-\xi)]}{\xi}$$

$$Standard\ Deviation = \sqrt{\frac{\left(\Gamma(1-2\xi) - \left(\Gamma(1-\xi)\right)^2\right)\beta^2}{\xi^2}}$$

4.2.4 Key Functions of the Model

The main task of the integral methodology is to extract the average number of acceptor and donor-like traps, N_A and N_D, and the parameters for the statistical distributions in Table 4.2 for a measured RTN dataset at a specific time point. By extracting their values for datasets measured at different time points, the N_A and N_D can be plotted against time to establish their kinetics. This mimics aging modelling and effectively transforms modelling traps of different time constants into modelling the kinetics of ECT.

Simulation results with an exponentially distributed impact per trap are used as an example. An illustration of how to transform this distribution into a time-dependent number of ECTs is given in the steps below:

1. Apply integral method to the CDF of the RTN dataset at different time points, as shown in Fig. 4.7.
2. Plot Fitted Average Number of Traps (ECT) per device, N_A and N_D, against Time as shown in Fig. 4.8.
3. Fit N_A and N_D with different kinetics: Power Law, Lognormal, and Log-uniform between 10^{-4} to 10^{1} sec as shown in Fig. 4.9.
4. The fitted kinetics are then extrapolated to predict the future RTN as shown in Fig. 4.10. The prediction by different kinetics is compared with the N_A and N_D extracted from test data at a longer time window to select the kinetics that give the best prediction.

The prediction capability is verified and discussed in Chapters 5, 6 and 7. As can be seen in Fig. 4.7, ΔVth can be either negative or positive and it can be fitted accurately by the integral method.

Another powerful feature of this integral method is that it can calculate the probability of a particular ΔVth if N_A and N_D and parameters for the statistical distributions in Table 4.2 are given. This removes the need for performing time-consuming Monte Carlo simulations.

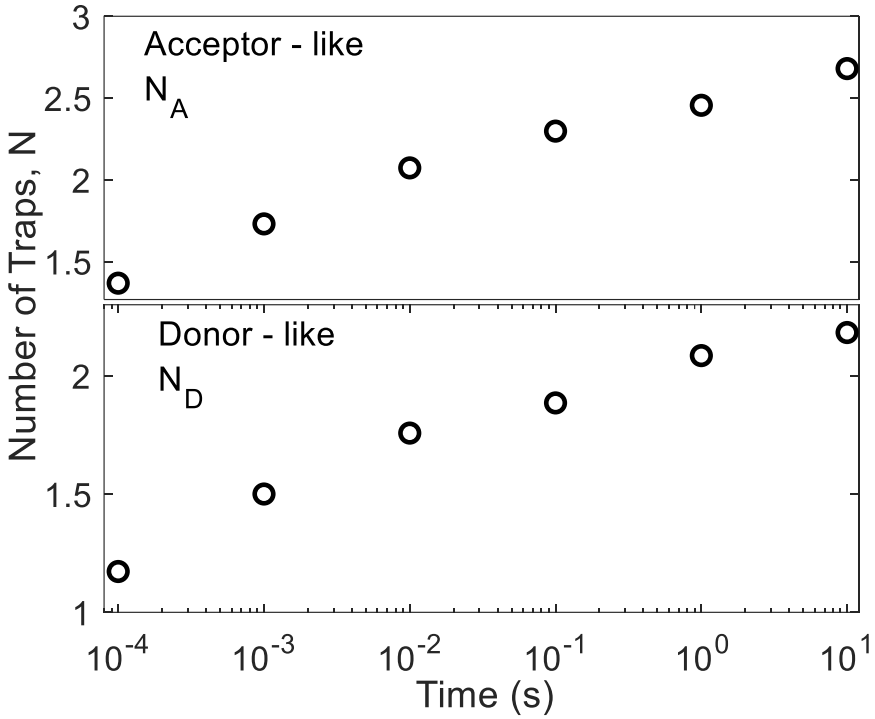


Fig. 4.8 The N_A and N_D extracted from a set of RTN measurements at different time points.

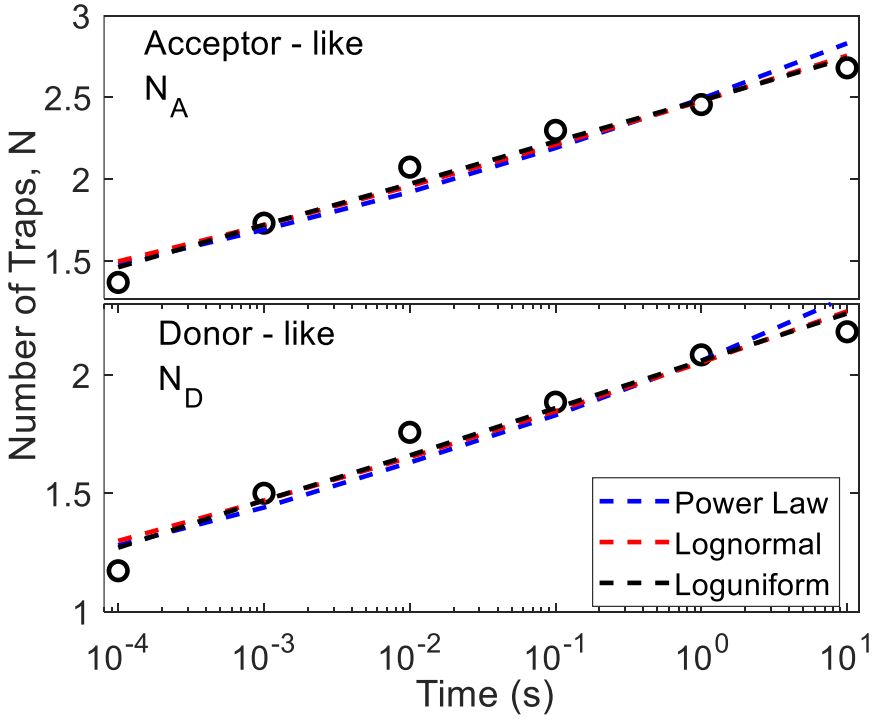


Fig. 4.9 N_A and N_D from 10^{-4} to 10^{1} sec are fitted by 3 kinetics; Power-Law, Lognormal, and Log-uniform.

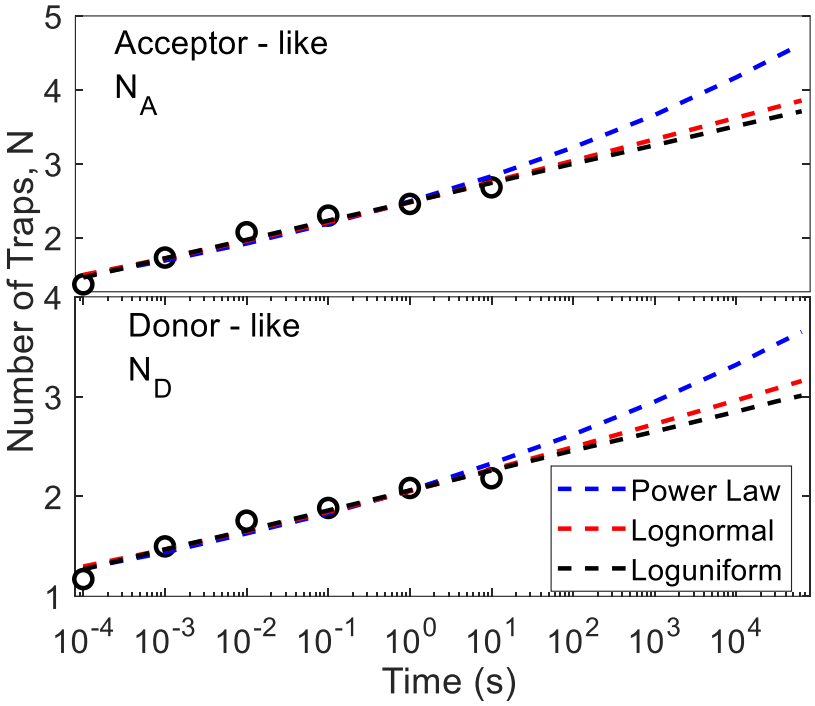


Fig. 4.10 The kinetics fitted between 10^{-4} to 10^{1} sec are extrapolated to a longer time window to predict RTN in longer time windows.

4.2.5 Optimization of Computation Time and Justification

For an MLE extraction, start values are required for the parameters to be fitted. As there are many mathematical calculations per fitting, this can become very time consuming. To reduce the overall fitting time, lower and upper boundaries can be set for the start values of parameters. For example, if start values are 1, lower and upper boundaries can be set at 0.5 and 1.5. The MLE extraction parameters will not go below or beyond the lower and upper boundaries. It is to be noted that, by fitting the equation through the MLE method, there is only one set of unique parameters to be extracted if the parameter estimation is convergent. The closer the start values are, the faster the MLE estimation will be. Thus, an exact computing time cannot be given, but the typical time will vary between 7 hours and 11 hours, approximately.

If the parameters' boundaries do not cover the parameters' values, the MLE will not converge. Another reason that can cause MLE not to converge is the number of iteration of MLE. If the start values are too far from the unique values, the number of iterations may not be enough. A common practice for start value selection is to first extract parameters at the shortest time of the available RTN datasets; in this case, the time is 10^{-4} sec. Fitting the data at 10^{-4} sec with wide boundaries will take a long time; however, once the parameters for 10^{-4} sec are obtained, the values of fitted impact per trap, δVth, can be used as the start value for δVth at the next

time point, since the average impact per trap should be time insensitive. The average number of traps per device should increase with time; hence, the start values can be set higher than the fitted values at a previous timepoint. For example, if MLE output for $N_A\ at\ 10^{-4}\ sec$ is 0.5, the start value for a timepoint later than 10^{-4} sec must be higher than 0.5, which helps in reducing the fitting time. By applying this practice, only the first set of data needs to use the wide boundary conditions which require a long fitting time (11 hours), and the following datasets can use narrower boundary conditions which only need approximately 7.5 hours.

The summation in Equation (4.14) must be terminated at a sufficiently high n_A and n_D to give an acceptable accuracy. In this work, the highest number of n_A and n_D is set at 12 and it is found that this allows accurate computation of devices with an average of 4 traps, as will be shown later. However, not every dataset has an average number of traps up to 4. The higher the n_A or n_D, the greater the number of convolutions and computation time needed.

To reduce the fitting time, the maximum number of traps is made adjustable. The higher the number, the larger the convolution needs to be. If they are far beyond the average number of traps, the probability calculated at a high number of traps contributes little and can be ignored. Hence, the relationship between the maximum number of traps and the average number of traps must be found. By using flexible $n_A\ and\ n_D$, the fitting time for N_A and $N_D = 3$ is given in Fig. 4.11 by using different maximum $n_A\ and\ n_D$.

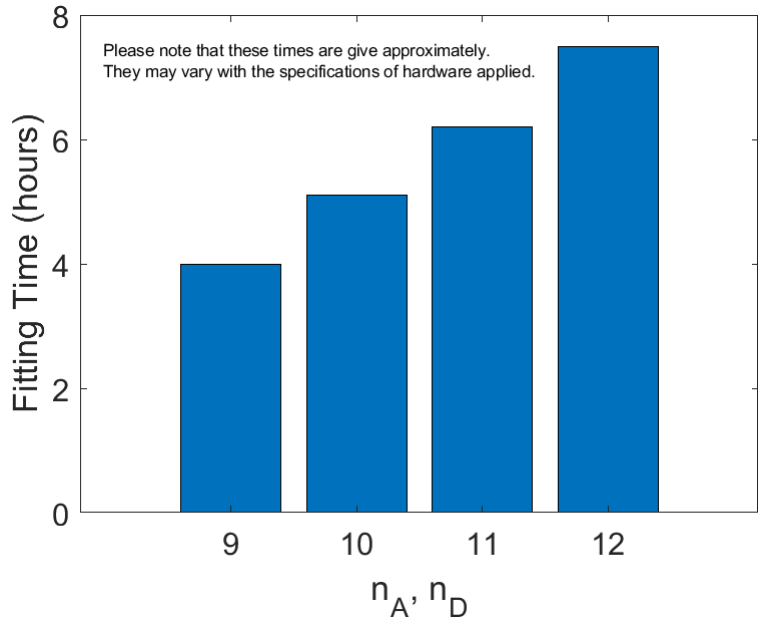


Fig. 4.11 The relationship between the maximum n_A, n_D and fitting time.

To further improve the fitting time, a look-up table based on each distribution is established to be used as an approximation of numerical integration. The CDFs in the look-up table are calculated based on the methodology discussed above (numerical integration) for Acceptor or Donor-like traps and saved as a reference for every specified value of parameters of distributions from ΔVth = -10 mV to 10 mV. This range of ΔVth is selected because it covers the spread of measured ΔVth.

Exponential: The look-up table for exponential distribution is stored up to η = 0.8 mV with 0.05 mV interval steps. 0.8 mV is selected as the highest value because this is the highest value fitted from experimental data in this work.

Lognormal and GEV: The look-up table for Lognormal distribution is a table for 2 parameters, ϵ and θ. GEV has 3 parameters, $\xi, \alpha \; and \; \beta$. GEV requires multiple look-up tables. For each given ξ, a table is established with α and β changing from 0.1 to 1 with 0.1 interval steps. As illustrated in Fig. 4.12, for each specified $\xi, \alpha \; and \; \beta,$ the probability that a given ΔVth will occur when there is n_A number of acceptor-like traps in a device is pre-calculated and stored in the table for future use.

During Maximum Likelihood Estimation, the numerical convolutions in Equations (4.4)-(4.7) have to be repeatedly carried out for each set of estimated parameters, which is time consuming. The look-up table is used to avoid this repeated computation, as it allows the required results to be extracted from the table.

The look-up table uses the maximum number of acceptor-like traps of $n_A = 15$. The maximum n_A = 15 is selected because the current experimental data has the maximum average of 4 traps per device, which requires at least $n_A = 12$. This relationship between average number of traps per device and maximum n_A will be discussed later. To be conservative, $n_A = 15$ is applied. Acceptor-like and Donor-like traps share the same look-up table.

With the probabilities of n_A and n_D known from 0 to 15, the convolution of Acceptor and Donor-like traps in Equation (4.11) can be calculated reasonably quickly. Fitting time between numerical integration and look-up table is compared in Fig. 4.13. The fitting time is significantly reduced from 240 minutes to 16 minutes by applying the look-up table. These times, however, should be seen as a rule-of-thumb, as they can vary with the specifications of the users' computers.

For parameter values or ΔVth that are not in the look-up table, interpolation between two closest values available is applied as an estimation. A comparison of fitted CDF between numerical integration and look-up table is given in Fig. 4.14(a). Fig. 4.14(b) gives an error comparison between each fitted CDF and test data item to justify the interpolation with the look-up table: the error is similar. The normalized difference between two fitted CDFs in Fig. 4.14(a) is 1.47×10^{-5}, which further justifies the use of interpolation for estimation.

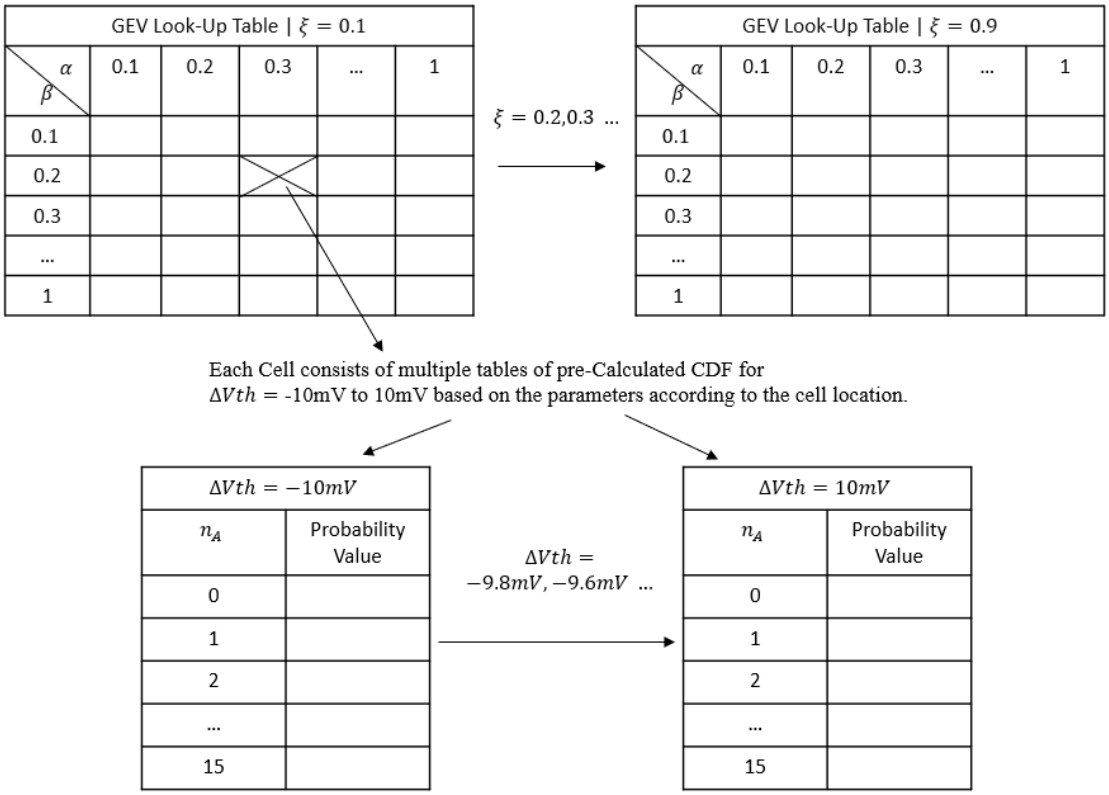


Fig. 4.12 Illustration of GEV Look-Up Table with 3 fitting parameters.

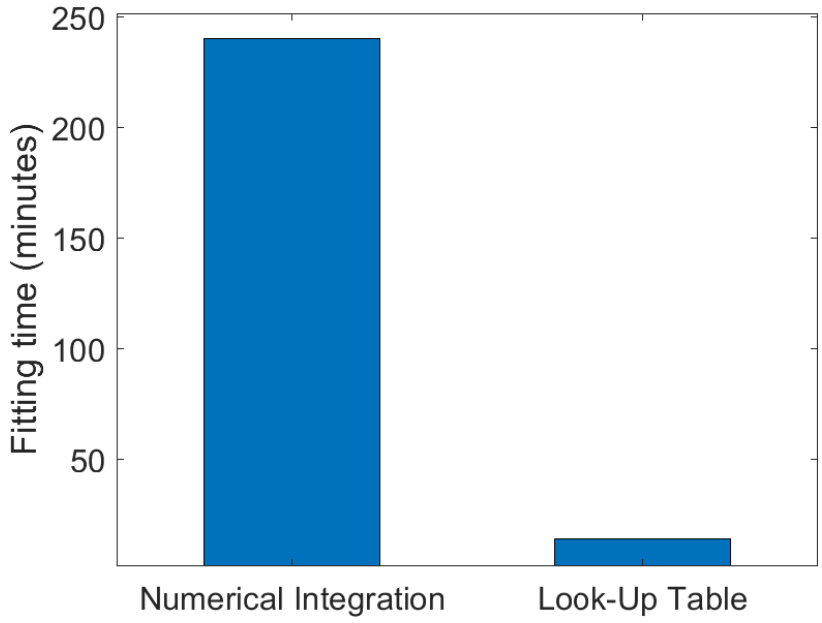


Fig. 4.13 Fitting time comparison between Numerical Integration and Look-Up Table.

A major advantage of this integral method is that it enables the direct and rapid calculation of the probability for a particular ΔVth to occur for a given set of statistical parameters. The computation time is insensitive to the ΔVth value. A large ΔVth will occur rarely and correspond to a large multiple of the standard deviation, σ. A computation time comparison between Static Monte Carlo Simulation and Integral Method against multiple σ is made in Fig. 4.15. The higher the σ multiples, the longer the time required for the Monte Carlo Simulation, as more hypothetical cases have to be generated. At 6σ, the computation time of static Monte Carlo simulation is 4 orders higher than that of the integral method.

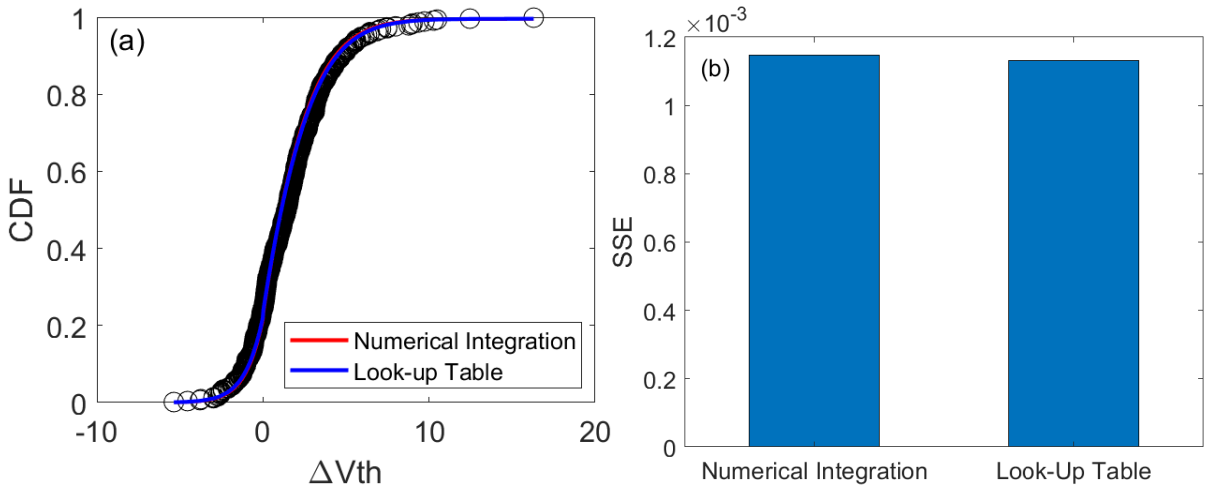


Fig. 4.14 (a) Fitted CDFs of Numerical Integration and Look-up Table. (b) Error comparison between two fitted CDFs and test data.

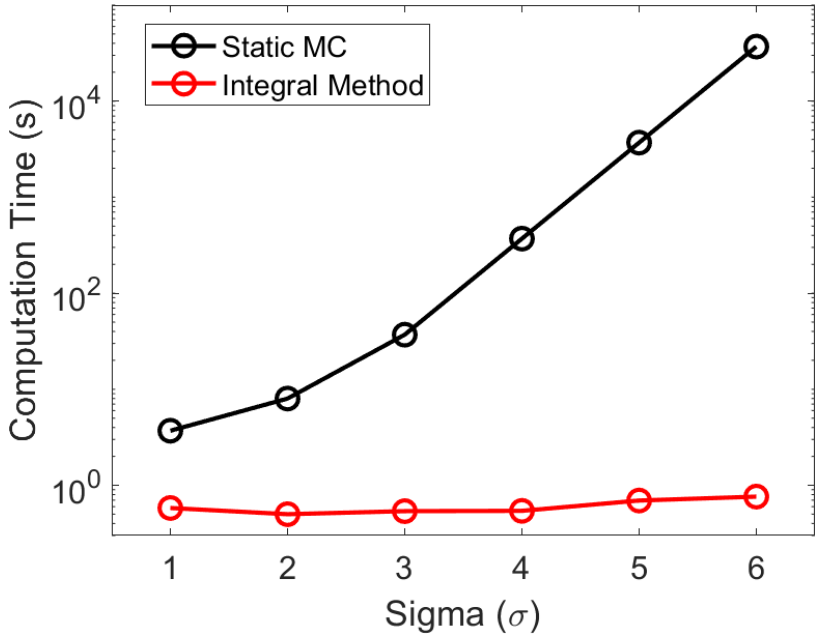


Fig. 4.15 Computation time comparison between Static Monte Carlo Simulation and Integral Method against σ.

To justify the setting of the maximum number of acceptor or donor-like traps per device, the dependence of fitting errors on them is examined. For a dataset with the average number of acceptor-like traps of 2 ($N_A = 2$) per device, Fig. 4.16 shows that errors reduce initially, but further reduction is small when the maximum number of traps is set at 4. The result clearly showed that the maximum number of traps should be at least double (i.e., a factor of 2) the average number of traps. To be conservative, a factor of 3 is applied throughout the model in this work.

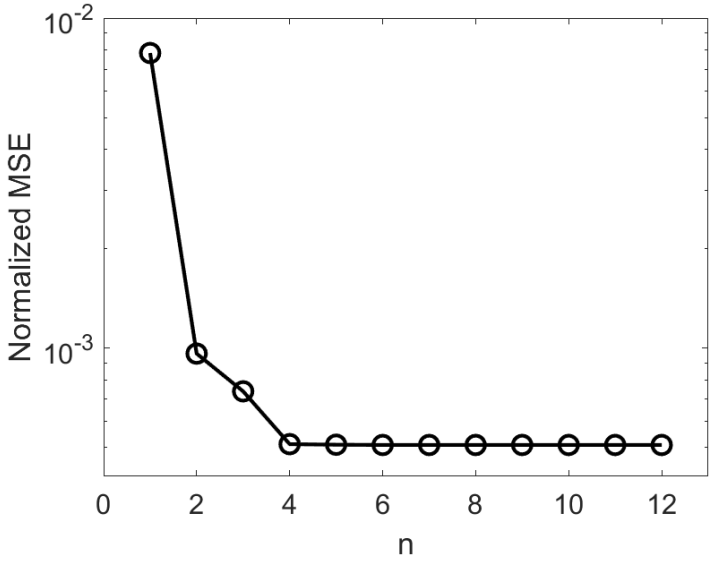


Fig. 4.16 The relationship between the maximum n_A and n_D against the normalized MSE on a dataset with average 2 traps per device.

4.3 Conclusion

In this chapter, the process of RTN modeling has been described, including how this methodology can be applied and step-by-step guide for creating the RTN model. The method for parameter extraction based on the RTN modeling equation has been given. This model has been justified by fitting to a theoretically generated statistical set of data with a known average number of traps per device, N_A and statistical distribution of δVth. For the first time, a methodology is proposed that provides the capability of estimating the probability for a certain ΔVth to occur, which provides a powerful new tool for circuit designers to optimize circuit designs.

CHAPTER 5: AN INTEGRAL METHODOLOGY FOR PREDICTING LONG TERM DC RTN

This chapter discusses the application of integral methodology on RTN measurement data and long-term RTN prediction. The first section gives the motivation for developing a trustworthy RTN model. The second gives the methodology and measurement method on RTN measurement and data processing. Follows by discussion on the concept of integral methodology and how it can be used on RTN data. Lastly, the CDF prediction of long-term RTN has been applied and verified.

5.1 Introduction

As discussed in Chapter 1, noise in MOSFETs adversely affects the performance of circuits [1-3], [12], [19], [21], [29], [31], [64], [102], [107-108]. IoT edge units are particularly vulnerable to it because the requirement of low power drives the operation voltage towards threshold level and reduces noise tolerance [12], [21].

Efforts have been made to optimize circuit design and performance [1-3], [12], [19], [21], [29], [31], [64], [102], [107-108]. When devices are relatively large, there are many traps, and they collectively form 1/f noise. Thus, the impact of a single trap is typically too small to be clearly observed, and modelling was mainly in the frequency domain [105]. However, as the downscaling of device size reached the nanometre scale, there are only a few traps in a device. The relative impact of a single trap increases and can be observed as Random Telegraph Noise (RTN) [1-3], [12], [19], [21], [29], [31], [64], [102], [107-108], where drain current exhibits step-like changes. The stochastic nature of trap distribution results in large device-to-device variation (DDV) of RTN [1-3], [19], [64]. When compared with the time-independent DDV from other sources, such as discrete random dopants and line edge roughness [2], whose outliers can be screened out by pre-shipping tests of chips, the time-dependent DDV by RTN is difficult to screen out, as trapping may not occur within the limited test time window.

It has been reported that capturing one charge carrier by a trap can induce a shift of threshold voltage up to tens of millivolts [19] and a fluctuation of current up to 10% [24]. This level of instability is comparable with that typically used to define the ageing-induced device lifetime [109]. It adversely impacts circuit performance by causing, for example, jitters and malfunctioning of SRAM operation [3]. For future quantum computing at low temperature, this fluctuation can reach 50% [110].

To tackle these RTN-induced challenges, many recent efforts have been made to model RTN in the time domain [1], [3], [27], [30], [61], [111-113]. To assess the impact of RTN on a circuit, a circuit designer needs the probability of device parameters at a given level within a time window. This can be accomplished through simulation, if one knows the number of traps in a device, their capture/emission times, and the shift induced by each trap [1], [3], [27], [30], [61], [112-113].

To determine the statistical distributions of these parameters, a bottom-up approach was often followed: individual traps were characterized first and then grouped together to fit an assumed cumulative distribution function (CDF) [3], [19], [27], [30], [61]. Most attention has been paid to the CDF of RTN amplitude. The proposed CDFs include Exponential [3], [19], [99], [102], Log-normal [1], [3], [31], and Generalized Extreme Value (GEV) [30]. These early efforts laid the foundation for current work.

For the average capture/emission times of a trap, the proposed CDFs include log-normal [19], [61] and log-uniform [1], [21], [114]. When compared with the amplitude, there is less data available on the capture /emission times [27], [61]. The number of capture/emission times obtained by the bottom-up approach is often less than 100 [27], [61], which is not large enough to reliably determine the CDFs. Without the CDFs of time constants, one could not predict RTN for different time windows. Some early works [24], [27] measured RTN with a time window of tens of seconds, while it is well known that RTN-induced fluctuation increases with a longer time window. In addition, most circuits are not designed to operate for a short time, such as a few seconds. A number of compact models [111-113] were proposed assuming that RTN magnitude and time constants follow certain statistical distributions. However, it is not verified that these models can be used to predict long-term RTN.

5.2 Motivation for the Trustworthy RTN Model

On top of these reliability issues, the traditional method of extracting the properties of individual traps requires the selection of specific devices with clear RTN behaviour. The issue with this approach is that there is a significant chance that a device does not have RTN and device selection results in an overestimation of the RTN impact by excluding devices with no RTN. Since bottom-up approaches have been widely attempted in the past and did not establish a trustworthy RTN model, a top-down modelling approach is proposed and discussed in this work.

The aim of this work is two-fold: to develop an integral methodology for extracting the CDF of RTN and, for the first time, to verify that the extracted model can predict long-term RTN. This will be achieved through:

- Characterizing the impact of all traps on a device collectively, rather than individually. This removes the need in some early works to select devices with individually analysable traps [35], [38].
- Integrating the RTN of multiple devices into one dataset to enable statistical analysis.
- Using the concept of "effective charged traps (ECT)", which removes the formidable burden of determining the capture and emission time constants of individual traps.
- Replacing the fixed number of traps per device, used by early works, with the time-dependent number of ECT. This mimics the modelling of ageing, where trap number increases with stress time through generation [109]. It transforms the distribution of time constants into the kinetics of ECT, simplifying the prediction of long term RTN.
- Introducing acceptor-like and donor-like traps to enable modelling of both positive and negative parameter shift, which is widely observed experimentally but rarely modelled.

5.3 Methodology and Measurement

Drain current, I_D, was measured against time under a gate bias, Vg, of 0.5 V and drain bias, Vd, of 0.1 V, as reported in [21], and one example is given in Fig. 5.1 (a). The Vg was chosen to be close to the threshold voltage, Vth, of 0.45 V for low-power applications. The devices used are nMOSFETs with a channel length of 27 nm and channel width of 90 nm, fabricated by a 28 nm commercial CMOS technology. To catch fast traps, a sampling rate of 1 MSample/sec was used and all tests were carried out at 125ºC.

The reference I_D, Iref, for each device was obtained from the average value of the first 10 measurement points. The more points used to calculate Iref, the less the thermal noise effect. However, the more points used to calculate Iref, the later the RTN characterization can be started: 10 points were selected as a balanced trade-off between these assumptions. The impact of the number of points used to calculate Iref on the measured statistical distribution is shown in Fig. 5.2, where the distribution with one point is compared with that with 10 points. The difference in their average ΔVth is less than 2%.

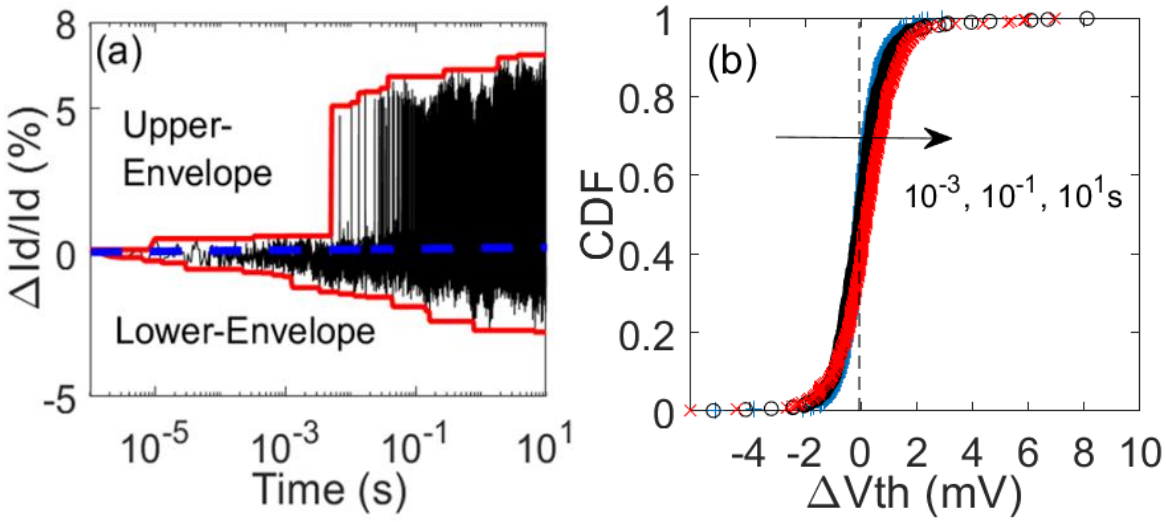


Fig. 5.1 (a) An example of measured data on one device, where $\Delta I_D>0$ represents a reduction of Id from its reference value. The black lines are the $\Delta I_D/I_D$ values, which can be either positive or negative. The red lines represent the upper and lower envelopes of the fluctuation. (b) The CDF of $\Delta Vth=\Delta I_D/g_m$ at different times taken from 402 devices. For each device at a given time, the ΔI_D value was taken from a measurement like the one in (a). The distribution is not symmetric in $\Delta Vth>0$ and $\Delta Vth<0$.

The relative shift of I_D was evaluated from $\Delta I_D/I_D$=(Iref-I_D)/Iref, so that a reduction of I_D magnitude gives a positive $\Delta I_D/I_D$. The corresponding threshold voltage shift is estimated from $\Delta Vth=\Delta I_D/g_m$, where g_m is transconductance and measured for each device from the slope of a pulse (3 µs) I_D-Vg at Vg=0.5 V [4], which was taken before the RTN test. The impact of RTN trapping-detrapping during the IV measurement on g_m was not taken into account here. As the RTN induced $\Delta I_D/I_D$ is in the order of a few percent, the corresponding uncertainty in g_m should also be a few percent. This leads to an uncertainty of a few percent in $\Delta Vth=\Delta I_D/g_m$, which is considered as insignificant, when compared with the large statistical device-to-device variation studied in this work, as shown in Fig. 5.1(b).

Based on the I_D versus time measurement, early works [24], [27] often selected the devices that have analysable RTN signals for extracting the amplitude and time constants of individual traps and discarded devices where the fluctuation was too complex for such analysis. In this work, however, we focus on the collective impact of all traps on I_D and do not experimentally decompose the I_D fluctuation into contributions of individual traps. This removes the need for device selection and increases test efficiency.

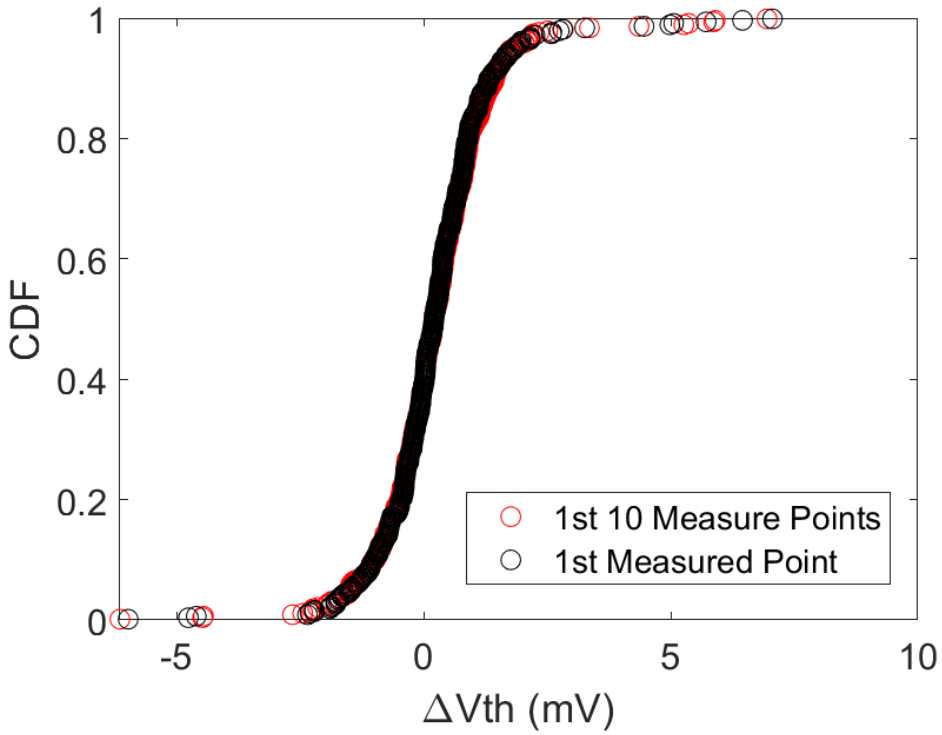


Fig. 5.2 The number of points used to calculate Iref has little impact on the measured statistical distribution.

The test in Fig. 5.1(a) was repeated on 402 devices and their ΔVth at a given time was integrated into one dataset for statistical analysis, as shown in Fig. 5.1(b). As time increases, the amplitude of fluctuation rises in Fig. 5.1(a), so that the |ΔVth| at a fixed CDF becomes larger in Fig. 5.1(b). Knowing this distribution will allow one to determine the probability of Vth at a specified level and in turn assess its impact on circuits. The challenge is how to model this distribution and how to predict it for longer times when test data are not available.

5.3.1 The concept of Effective Charged Traps (ECT)

The concept of Effective Charged Traps was proposed in Chapter 4 and will be further elaborated here. The probability that a trap will be charged depends on the ratio of its average emission and capture time [64]. However, the statistical distribution of time constants is difficult to determine experimentally, and this difficulty increases along with the time window. To simplify tests, attempts were made to focus on either the capture [21], [115] or emission [19].

Fig. 5.1(a) shows how to extract the envelope of the fluctuation, which is reached when multiple traps were simultaneously charged. Fig. 5.3 shows the average of the envelopes, which can be used to estimate the number of active traps for a given time window. As different devices hit their envelopes at different times, Fig. 5.3 shows that the average of measured ΔVth is

substantially smaller than their average envelopes. In Section 5.5.1 estimation of the percentage of traps that are charged is investigated.

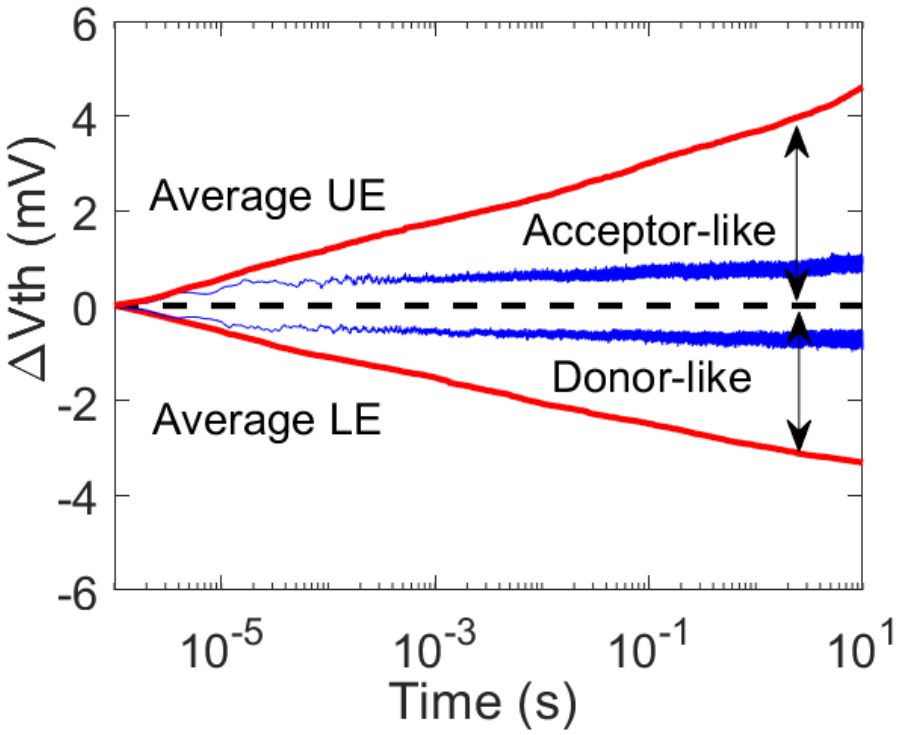


Fig. 5.3 A comparison of the average ΔVth>0 and ΔVth<0 of 402 devices (blue lines) with their average upper and lower envelopes. Using the envelopes will overestimate the impact of RTN.

The envelopes have been used to estimate the worst impact of RTN [21], [115-116], but they clearly overestimate this impact. To optimize circuit design, however, one needs to model the probability of ΔVth at any level between the two envelopes, which is the objective of this work.

To effectively take emission into account without evaluating the emission time of individual traps, the concept of 'effective charged traps (ECT)' is proposed in Chapter 4. The inspiration for this concept comes from the use of effective trap density for assessing the impact of ageing on devices [109], [117]. When the local current density beneath a trapped charge in the gate dielectric is the same, its impact on the current depends on its distance from the dielectric/substrate interface. It is, however, difficult to determine this distance. To overcome this difficulty, one can assume all traps being at the interface so long as they produce the same ΔVth. The equivalent density of traps at the interface is referred to as effective trap density [117].

Similarly, to simulate the impact of RTN on devices and circuits, what is needed is the CDF in Fig. 5.1(b), rather than the detailed physical processes that this CDF originates from. For a

given RTN-induced ΔVth at a time in Fig. 5.1(b), one can assume that it originates from a set of traps that are always charged. These 'effective charged traps' ignore emission, but will produce the same CDF as, i.e., be statistically equivalent to, that measured experimentally where emission occurs. By using the collective impact of a set of ECT to model CDF, as detailed in Section 5.4 this integral approach does not require characterization of the capture and emission time of individual traps, greatly reducing modelling and testing time.

5.3.2 Acceptor-like and Donor-like Traps

Fig. 5.1(a) shows that ΔI_D =(Iref-I_D) can be either positive or negative, resulting in the corresponding positive and negative ΔVth in Fig. 5.1(b). Early works [3], [19], [27], [102] typically only modelled ΔVth>0, since the popular Exponential and Log-normal distributions require ΔVth>0. As the distribution of ΔVth in both ΔVth>0 and ΔVth<0 is considerable and they are asymmetric, it is essential to model the RTN in both directions. Fig. 5.4 plots the upper envelope against the lower envelope at 10 sec, as defined in Fig. 5.1(a). There is little correlation between them, so that it is essential to model the fluctuation in both positive and negative directions, as the fluctuation in one direction cannot be used to infer that in the other direction.

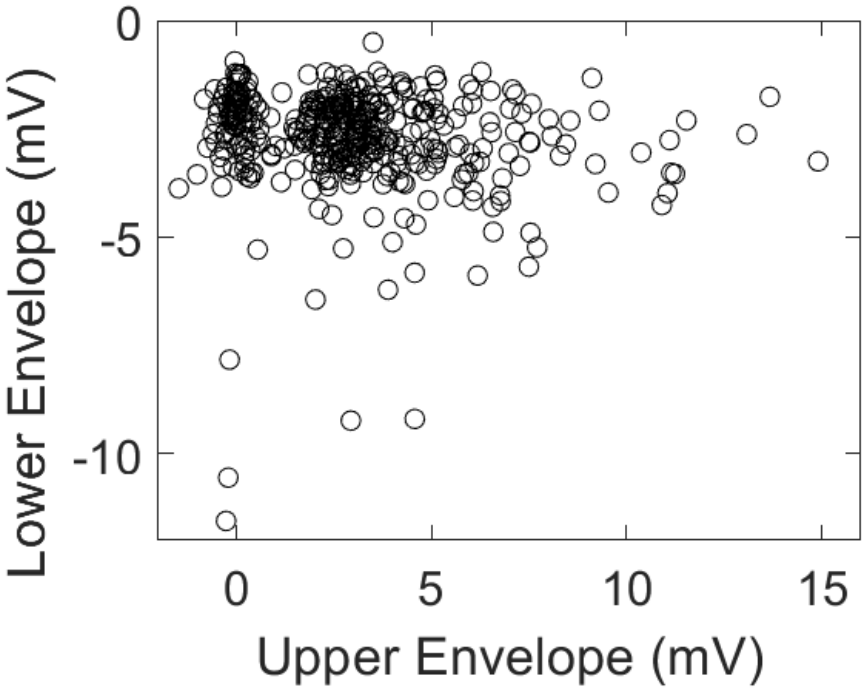


Fig. 5.4 The lower-envelope is plotted against the upper-envelope at 10 sec for 402 devices. There is little correlation between these two.

The shift in a positive direction, i.e., ΔVth>0, for nMOSFETs is widely interpreted as traps capturing electrons from the conduction channel. Such traps can be either neutral or negatively charged and will be referred to as acceptor-like traps [117]. There can be different explanations for ΔVth<0. One of them is that some traps were pre-filled with electrons when Iref was measured. These pre-filled electrons could be emitted during measurement, resulting in ΔVth<0 [99]. Another is that there are donor-like traps, which can be either neutral or positive [117], and positive charges can result in ΔVth<0.

The measured data in this work cannot determine which explanation is correct and such determination is not required for achieving the aim of this work: proposing an integral method for predicting the long term RTN. For the convenience of presentation, we will hereafter use acceptor-like traps for ΔVth>0 and donor-like traps for ΔVth<0 [117]. This allows ΔVth to be modelled in both directions through

$$\Delta Vth = \Delta Vth_A + \Delta Vth_D, \tag{5.1}$$

where $\Delta Vth_A > 0$ and $\Delta Vth_D < 0$ represent the contribution from acceptor-like and donor-like traps, respectively.

5.4 Modelling Procedure at a given time

To model the RTN-induced CDF in Fig. 5.1(b) by a set of ECTs at a given time, one needs to determine the number of traps per device, n, and the statistical distribution of threshold voltage shift caused by one trap, δVth. It should be noted that δVth is the shift per trap, which is different from the shift per device, ΔVth.

Statistically, it is well accepted that n follows a Poisson's distribution [3], [102]. Agreement has not been reached on the distribution of δVth and the popular assumptions include Exponential [3], [19], [99], [102] and Log-normal [1], [3], [31]. In addition, Generalized Extreme Value (GEV) distribution has been proposed recently [30]. All three distributions will be used in this work.

For a given δVth probability distribution function (pdf), the pdf of n_A number of acceptor-like traps can be evaluated from:

$$pdf_A(\Delta Vth_A) = Conv\left[pdf(\delta Vth_1), pdf(\delta Vth_2) \ldots pdf\left(\delta Vth_{n_A}\right)\right], \tag{5.2}$$

where ΔVth_A is the combined shift caused by n_A acceptor-like traps. When an analytic formula is not available, the Convolution (Conv) in Equation (5.2) can be carried out numerically trap-by-trap:

first between pdf(δVth_1) and pdf(δVth_2), and the result is then convoluted with pdf(δVth_3). This continues until $pdf(\delta Vth_{n_A})$ is convoluted.

To overcome the difficulty that Exponential and Log-normal pdf requires $\delta Vth > 0$, Equation (5.3) evaluates the magnitude of δVth for n_D donor-like traps by

$$pdf_D(|\Delta Vth_D|) = Conv\left[pdf(\delta Vth_1), pdf(|\delta Vth_2|) \ldots pdf(\left|\delta Vth_{n_D}\right|)\right] \tag{5.3}$$

The pdf[$\Delta Vth=\Delta Vth_A-|\Delta Vth_D|$] can be evaluated from

$$pdf[\Delta Vth(n_A, n_D)] = \int_{-\infty}^{\infty} pdf_A[\Delta Vth(n_A, n_D) + x] pdf_D(x) dx. \tag{5.4}$$

The pdf[$\Delta Vth(n_A,n_D)$]) in equation (5.4) is the probability that a device will have a ΔVth, if every device has n_A acceptor-like and n_D donor-like traps. After taking into account that both acceptor-like and donor-like traps follow Poisson distributions, the pdf(ΔVth) becomes:

$$pdf(\Delta Vth_{RTN}) = \sum_{n_A=0}^{\infty} \sum_{n_D=0}^{\infty} \frac{e^{-N_A} N_A^{n_A}}{n_A!} \frac{e^{-N_D} N_D^{n_D}}{n_D!} pdf[\Delta Vth(n_A, n_D)], \tag{5.5}$$

where N_A and N_D are the average number of effective acceptor-like and donor-like traps per device, respectively.

Finally, the measured ΔVth contains both RTN and thermal noise, which follows a Normal distribution. The thermal noise is taken into account by using:

$$pdf(\Delta Vth) = Conv[pdf(\Delta Vth_{RTN}), pdf(\Delta Vth_{Thermal})]. \tag{5.6}$$

5.5 Results: Extracting Model Parameters

Based on the equations (5.1) – (5.6), the model parameters are extracted by the Maximum Likelihood Estimation (MLE) [30]. They include the statistical distribution parameters of δVth in Table 5.1, the average number of acceptor-like and donor-like traps per device, N_A and N_D, and the standard deviation of thermal noise.

As an example, Fig. 5.5(a) shows that the fitted CDFs with test data at 10 sec by assuming Exponential, Log-normal and GEV distributions, respectively, with thermal noise fitted. To quantify the discrepancy between the data and the fitted CDF, Fig. 5.5(b) gives the sum of square error (SSE) per device. The SSE reduces in the order of Exponential, Log-normal, and GEV.

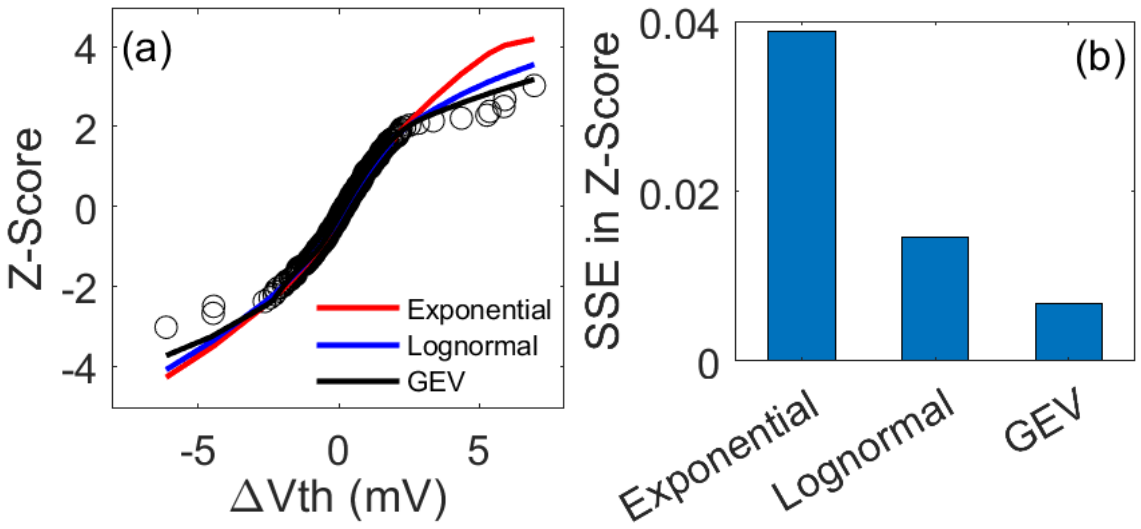


Fig. 5.5 (a) The CDF fitted based on the effective charged traps at 10 sec. The symbols are test data, and the lines are fitted. (b) compares the sum of squared errors (SSE) per device for the three statistical distributions.

Similar to Fig. 5.5 at 10 sec, the test data at other times are also fitted. Fig. 5.6(a) and Fig. 5.6(b) show the extracted average number of acceptor-like (N_A) and donor-like (N_D) traps, respectively, per device against time. As expected, N_A and N_D increase with time, as longer time activates slower traps. There are more acceptor-like traps than donor-like traps, resulting from the skewed distribution towards positive ΔVth in Fig. 5.1(b). GEV gives the highest numbers, while Exponential has the lowest. At 10 sec, GEV has N_A =2.3 and N_D = 2.

The extracted average δVth induced by one trap, μ, is given in Fig. 5.6(c) and Fig. 5.6(d) for acceptor-like and donor-like traps, respectively. In agreement with early work [27], [114], μ is independent of the time. The acceptor-like traps have larger μ than the donor-like traps. GEV gives the largest μ and Exponential gives the lowest. It should be noted that the average μ is in a range of 0.5 ~ 0.8 mV, indicating that there are small δVth that can be difficult to measure directly.

The other parameters for δVth distributions are not sensitive to time, either, since δVth is the shift per trap. Their values are given in Table 5.1.

Table 5.1 The pdf formula and their average parameter values extracted between 10^{-4} and 10 sec. The parameters η, α, β, and σ have the unit of mV.

	PDF of δVth	Acceptor	Donor
Exponential	$\frac{1}{\eta}e^{-\frac{\delta Vth}{\eta}}$	$\eta = 0.56$	$\eta = 0.48$
Lognormal	$\frac{1}{\delta Vth\theta\sqrt{2\pi}}e^{\left(-\frac{(\ln(\delta Vth)-\epsilon)^2}{2\theta^2}\right)}$	$\epsilon = -0.43$ $\theta = 0.12$	$\epsilon = -0.71$ $\theta = 0.16$
GEV	$\frac{1}{\beta}(k)^{\xi+1}e^{-k}$ $k = \left(1+\xi\left(\frac{\delta Vth-\alpha}{\beta}\right)\right)^{-\frac{1}{\xi}}$	$\xi = 0.35$ $\alpha = 0.43$ $\beta = 0.34$	$\xi = 0.42$ $\alpha = 0.59$ $\beta = 0.19$
Thermal	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\Delta Vth}{\sigma}\right)^2}$	$Exponential, \sigma = 0.11$ $Lognormal, \sigma = 0.13$ $GEV, \sigma = 0.13$	

.

5.5.1 Prediction of Effective Charged Traps (ECT)

The RTN signal induced by a trap has two key properties: the amplitude and time constants. They are independent and both of them are randomly distributed. This gives two sets of independent statistical distributions: one for amplitude and the other for time constants.

As discussed earlier, capture and emission time is difficult to determine, but using the concept of ECT for modelling RTN converts the distribution of traps' time constants into a time-dependent number of ECTs, as shown in Fig. 5.6(a) and Fig. 5.6(b). This transforms the prediction of long-term RTN to finding the kinetics of ECTs, similar to predicting device ageing [107]. For ageing, power law is the well-known kinetics [107]. For the time constant distribution of RTN, two distributions were proposed: a uniform distribution against logarithmic time (Log-uniform) [1], [21] and a Log-normal distribution [29], [61]. These three forms of kinetics will be tested against the experimental data next.

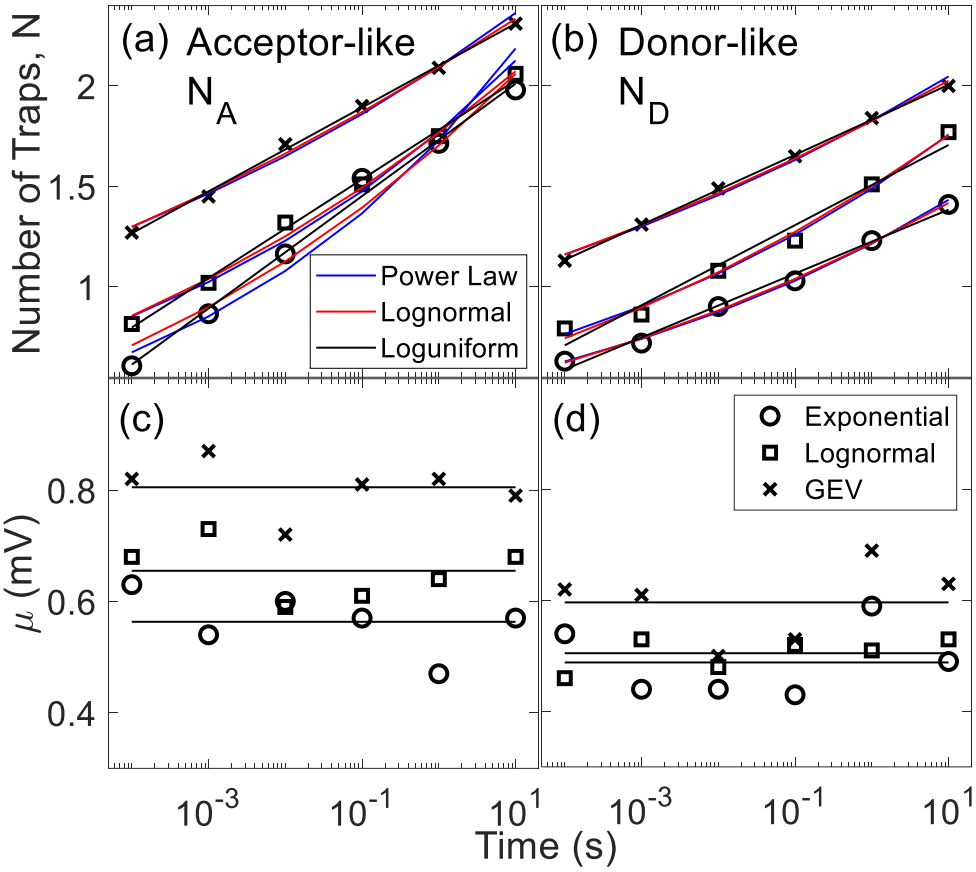


Fig. 5.6 The extracted average number of acceptor-like (a) and donor-like (b) traps per device at different times, based on different δVth distributions. The lines in (a) and (b) are the fitted kinetics. The extracted average δVth per trap, μ, is given in (c) for acceptor-like traps and in (d) for donor-like traps. The lines in (c) and (d) are the mean values.

Fig. 5.6 shows that all three kinetics can fit test data reasonably well over five orders of magnitude in time between 10^{-4} and 10 sec with the Root-Mean-Square-Error (RSME) of a few percent as shown in Fig. 5.7. As a result, good fitting with test data is not sufficient to justify a model. If a model is correct, it should be able to not only fit test data, but also predict the long-term RTN where test data is not used for fitting.

To further verify these kinetics, the RTN tests were extended from 10 to 6×10^4 sec. In Fig. 5.8, the symbols are extracted from test data. The data scattering beyond 10 sec is larger, since 402 devices were used between 10^{-4} and 10 sec and only 51 devices were used for the time-consuming tests of 6×10^4 sec.

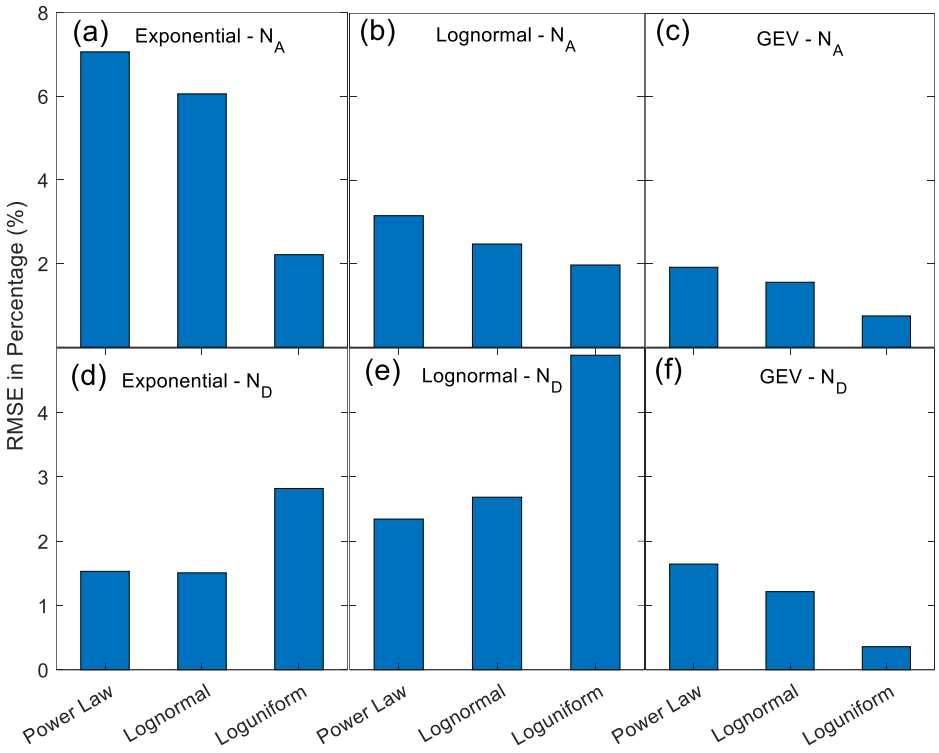


Fig. 5.7 RMSE for each fitted curve in Fig. 5.6 and the RMSE is within a few percent.

The solid lines in Fig. 5.8 were the kinetics fitted with data between 10^{-4} and 10 sec and the data beyond 10 sec were not used for the fitting. These fitted kinetics were then extrapolated from 10 to 6×10^{4} sec, as represented by the dashed lines. Although the differences between the fitted solid lines appear small, they become substantial for the extrapolated dashed lines as time increases. The sum of squared errors is summarized in Fig. 5.9(a) and Fig. 5.9(b) and discussed next.

For the Exponential δVth distribution, the Log-normal kinetics gives the lowest error for N_A in Fig. 5.8(a) and Fig. 5.9(a), but N_D agrees better with the power law, as shown in Fig. 5.8(d) and Fig. 5.9(b). It should be noted, however, that even though the power law gives the lowest error for N_D with Exponential δVth, i.e., the first blue bar on the left in Fig. 5.9(b), this blue bar is higher than the errors of Lognormal and GEV δVth.

For N_A with Log-normal δVth distribution, Fig. 5.9(a) shows that power law has the highest error and errors are similar for Log-normal and Log-uniform kinetics. For N_D, Fig. 5.8(e) and Fig. 5.9(b) show that Log-normal kinetics has the lowest error. For GEV δVth distribution. Fig. 5.8(c), Fig. 5.8(f), Fig. 5.9(a) and Fig. 5.9(b) show that Log-uniform has the lowest error for both N_A and N_D.

The Log-normal kinetics should lead to an eventual saturation of N_A and N_D as time increases [30]. It can, however, fit the test data within the measurement window without approaching saturation. Based on the result in Fig. 5.8, it is not clear enough to rule out the log-normal kinetics.

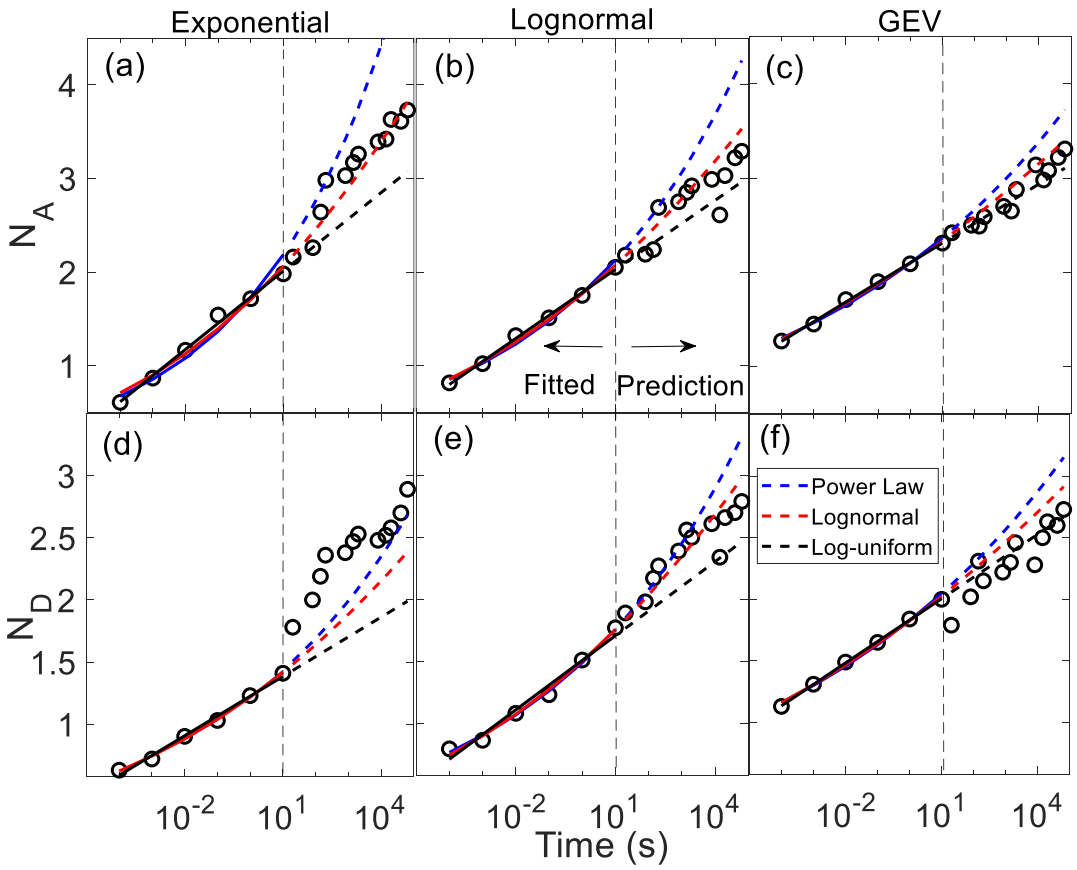


Fig. 5.8 Predicting the average number of acceptor-like traps, N_A, in the top row (a, b, c) and donor-like traps, N_D, in the bottom row (d, e, f). Symbols are extracted by fitting with the δVth distribution of Exponential in the left column (a, d), Log-normal in the middle column (b, e), and GEV in the right column (c, f). The solid lines were fitted with symbols between 10^{-4} and 10 sec for different kinetics. The dashed lines were extrapolated to 6×10^{4} sec. The symbols beyond 10 sec were not used for fitting.

Based on Fig. 5.8 and Fig. 5.9, the power law gives poor prediction of RTN overall. The RTN and ageing follows different kinetics, therefore. RTN interacts with device ageing. On one hand, the traps responsible for the RTN can also contribute to the charge build-up during ageing tests, such as bias temperature instability (BTI) [118-119]. On the other hand, it has been reported that, after stress, some RTN signals can disappear for some time and then reappear [118]. One may speculate that stress releases some hydrogenous species, which interact with the RTN trap through temporary bonding, deactivate the RTN, and then migrate away. The detailed physical process is beyond the scope of this work [109], [118].

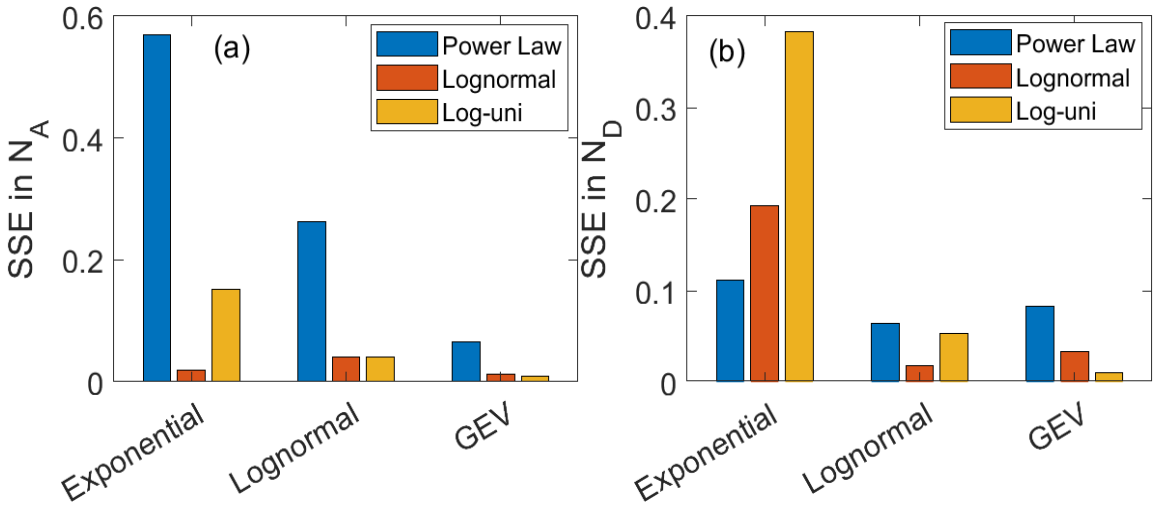


Fig. 5.9 The sum of squared errors for the prediction in Fig. 5.7 for acceptor-like (a) and donor-like (b) average number of traps per device. The lowest errors were obtained for Log-uniform kinetics with GEV δVth distribution.

The percentage of ECTs against the active traps available for a given time window is estimated. The average number of active acceptor-like (N_{EA}) and donor-like (N_{ED}) traps per device can be estimated from the Upper- and Lower-Envelopes in Fig. 5.3 by dividing their average δVth per trap, i.e., μ, given in Fig. 5.6(c) and Fig. 5.6(d), respectively. Fig. 5.10 gives the ratio of ECT, N_A and N_D extracted for the GEV distribution in Fig. 5.8(c) and Fig. 5.8(f) against N_{EA} and N_{ED}, respectively. In a short time, over 50% of traps are effectively charged. This ratio decreases for a longer time window and settles around 1/3.

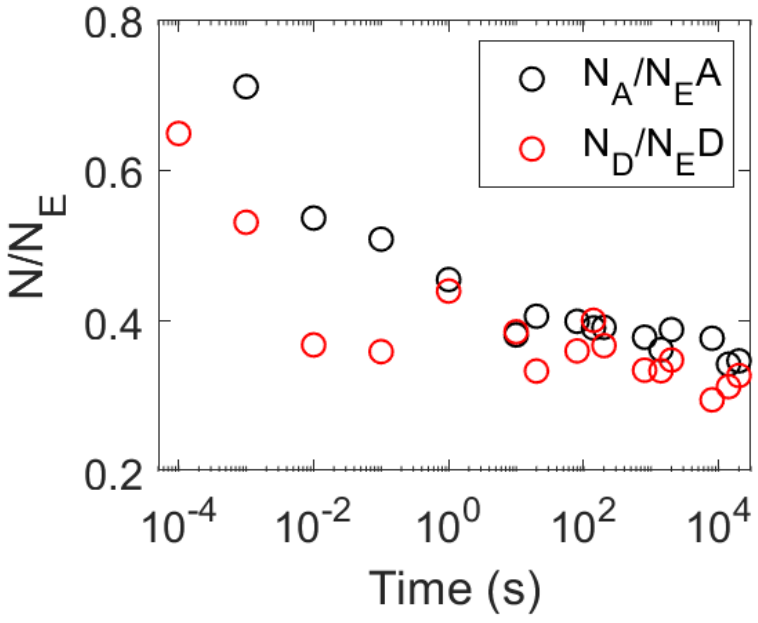


Fig. 5.10 The ratio of Effective Charged Traps, N_A and N_D, against the active traps available estimated from the Upper envelope, N_{EA}, and Lower Envelope, N_{ED}, respectively.

5.6 The CDF Prediction of Long-term RTN

As discussed in the introduction, the probability distribution function is what a designer needs for assessing the impact of RTN within a given time window. The question is how well one can predict the long-term CDF of RTN, based on the parameters extracted over a short time. For the first time, we attempt to predict the CDF of RTN at 6×10^4 sec by the model extracted from the data between 10^{-4} and 10 sec.

For each δVth distribution, the number of acceptor-like and donor-like traps at 6×10^4 sec is predicted by the kinetics of the lowest errors, as given in Table 5.2. The statistical parameters of δVth distribution and thermal noise are assumed to be independent of time, and their values given in Table 5.1 were used. With these parameters, the probability that one device will have a given ΔVth can be calculated directly, making the RTN simulation more efficient than the Monte Carlo simulation.

Fig. 5.11(a) and Fig. 5.11(b) compare the measured and predicted CDF at 6×10^4 sec for the Exponential, Log-normal and GEV δVth distributions. For the Exponential, the kinetics used were Log-normal for acceptor-like traps and power law for donor-like traps. For the Log-normal δVth distribution, Log-normal kinetics were used for both acceptor-like and donor-like traps. For the GEV, Log-uniform kinetics were used for both acceptor-like and donor-like traps. The two distributions selected for ΔVth (Vg = 0.5V and Vd = 0.1V) are,

1. Impact per trap: GEV Distribution
2. Long-Term RTN Kinetic: Log-uniform Distribution.

Fig. 5.11(c) shows that the errors reduce in the order of Exponential, Log-normal, and GEV. A reasonable agreement is obtained with test data, as shown in Fig. 5.10(a) and Fig. 5.10(b). This verifies the integral methodology proposed and its ability to predict the RTN at 6×10^4 sec based on data measured between 10^{-4} and 10 sec, a factor of 6×10^3 ahead. If one uses the data between 10^{-4} and 6×10^4 sec to make the prediction, it is reasonable to expect that one can predict a factor of 6×10^3 beyond this time. This will take the time to 3.6×10^8 sec, which is more than 10 years. It should be pointed out that, in addition to RTN, ageing also shifts device parameters, which must be included to predict the overall shift in the long term [118-119].

Table 5.2 The kinetics fitted with data between 10-4 and 10 sec.

PDF	ECT	Kinetics
Exponential	N_A	$\frac{1}{2}erfc\left(-\frac{\ln(t)-21.5}{19.1\sqrt{2}}\right)$
	N_D	$1.2152 * t^{0.0716}$
Lognormal	N_A	$\frac{1}{2}erfc\left(-\frac{\ln(t)-24.8}{22.5\sqrt{2}}\right)$
	N_D	$\frac{1}{2}erfc\left(-\frac{\ln(t)-29.04}{24.2\sqrt{2}}\right)$
GEV	N_A	$0.091 * \ln(t) + 2.1$
	N_D	$0.076 * \ln(t) + 1.83$

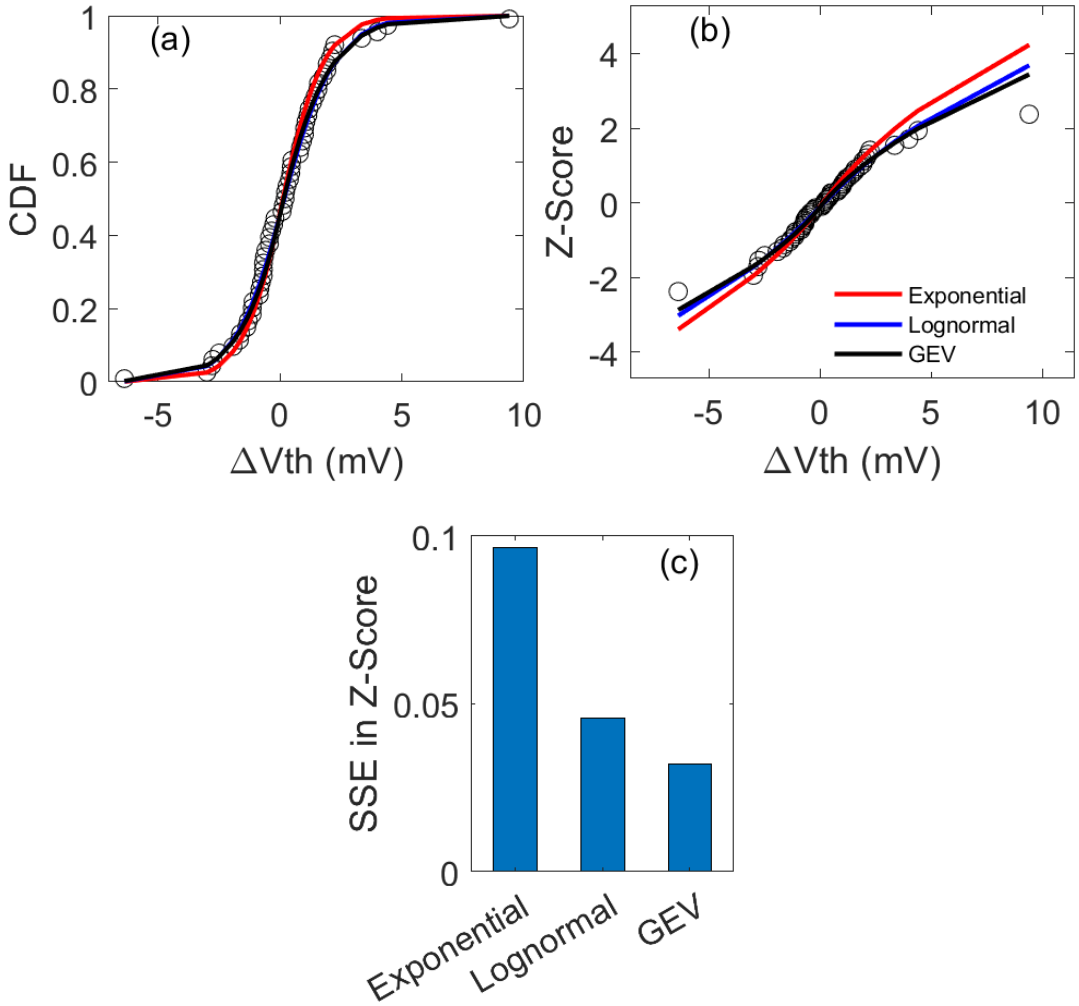


Fig. 5.11 Predicting the CDF of RTN at 6×10^4 sec based on the models extracted from the data between 10^{-4} and 10 sec. The symbols were measured data and the lines were the CDF predicted by using the N_A and N_D predicted in Fig. 5.8 and the average μ in Figs. 5.6(c) and 5.6(d). The CDF is plotted linearly in (a) and in Z-score in (b). The lines in (a) and (b) are the CDF fitted with different δVth distributions. (c) shows that the lowest error was obtained with GEV δVth distribution and Log-uniform kinetics.

5.7 Conclusion

In conclusion, the capability to predict the long-term RTN by the models developed in early works has not been verified, while this work both proposes an integral methodology and verifies its capability to make this prediction. Instead of characterizing the contribution of individual traps to RTN, the impact of traps in a device was measured collectively, with no need to select devices. Through integrating RTN measured on multiple devices into one dataset and using the concept of effective charged traps (ECT), the statistical distribution of device parameters at a given time is modelled, removing the formidable burden of characterizing the time constants of individual traps. This transforms the distribution of time constants to the kinetics of ECT, making the prediction of long term RTN similar to predicting ageing.

Furthermore, the accuracy of RTN amplitude distribution per trap proposed by early works was assessed, including Exponential, Log-normal, and Generalized Extreme Value (GEV). The three kinetics examined are power law, Log-normal, and Log-uniform. The power law gives poor prediction and RTN follows different kinetics from ageing. Based on the results, the lowest error was obtained with GEV δVth distribution and Log-uniform kinetic. Fluctuations in both positive and negative directions are modelled through using acceptor-like and donor-like traps. The work shows that this integral methodology can predict RTN by a factor of 6×10^3 ahead, opening the way for predicting RTN to 10 years based on measurements in a time window of one day.

CHAPTER 6: AC RTN INVESTIGATION

This chapter investigates RTN under AC biases. The first section gives the methodology and measurement method on AC RTN measurement and data processing. The second covers the results and discussion of AC RTN based on integral methodology. Analysis of AC RTN on the time constant of single trap against frequency is also studied. Lastly, the CDF prediction of long-term AC RTN has been applied and verified.

6.1 Introduction

As discussed in Chapter 2, random telegraph noise (RTN) in MOSFETs is caused by capturing a charge carrier from the conduction channel and then giving it back [1-4], [6], [12-17]. The need for low power consumption is driving operating voltage towards threshold, Vth, where a single trapped charge can have a large impact has led to an increment of attention on RTN [1-2], [19], [64], [120].

Many efforts were made on both characterizing and modelling RTN [1-4], [6], [12-17]. On characterization, many early works were carried out under DC conditions, where the gate and drain biases were fixed at a constant level [1], [21], [29], [31], although digital circuits typically operate under AC conditions. It has been reported that AC RTN is considerably different from DC RTN [3], [49], [121-122]. On modelling, works were carried out in both time [2-3], [21], [27], [30], [121-122] and frequency [107] domains. In the time domain, Monte Carlo simulation has been carried out for both DC and AC RTN by assuming that RTN transitions are memoryless random Markov process [19], [49], [113], [121-122]. For AC RTN, τ_C and τ_E under both high ('H') gate bias Vg=Vdd and low ('L') Vg=0 is needed, which hereafter are represented by τ_{CH}, τ_{EH}, τ_{CL}, and τ_{EL}.

It is widely accepted that the number of traps per device follows a Poisson distribution [102]. It has been proposed that trap amplitude can follow Exponential [3], [19], [102], Log-normal [1], [3], [31], or Generalized Extreme Value (GEV) distribution [30]. The time constants have been assumed to follow either Log-normal [29], [61] or Log-uniform [1], [21], [114] distributions. When compared with the amplitude distribution, there are fewer data available to underpin the τ_C and τ_E distribution, as they are difficult to collect in large numbers, even for DC RTN. For AC RTN, there are more challenges, as the transistor is switched off under Vg=0, so that τ_{CL} and τ_{EL} cannot be measured directly.

Several methods were used to overcome the challenge and to obtain τ_{CL} and τ_{EL}. One of these is to assume that τ_{CL}/τ_{CH} and τ_{EH}/τ_{EL} constants, so that τ_{CL} and τ_{EL} can be estimated from the measured τ_{CH} and τ_{EH}, but the simulation results based on this assumption did not agree well with test data [29]. Another method uses compact models and τ_C and τ_E are assumed to be exponentially related to Ef-Et, where Ef is the Fermi-level at the dielectric/substrate interface and Et is the trap energy level [1], [49], [64], [112], [120]. Ef-Et under Vg=0 can be calculated, and it can then be used to evaluate τ_{CL} and τ_{EL}. The problem is that the time constants obtained in this way are not verified by test data, so that the accuracy of these compact models for AC RTN simulation is typically unknown.

To provide experimental data for AC RTN, the measured data under Vg=Vdd were joined together by removing the time of Vg=0 [121]. On one hand, it has been reported that the τ_C measured in this way changes little from its DC value. As τ_{CL} is typically much larger than τ_{CH}, the duration of Vg=0 contributes little to trap capture and its removal has little effect on capture. On the other hand, τ_{EL} is typically smaller than τ_{EH}, so that the time under Vg=0 can reduce the measured τ_E. By measuring both τ_C and τ_E under Vg=Vdd and removing the Vg=0 duration, the AC RTN of individual traps has been modelled successfully [3], [49], [121-122]. At present, however, there is a lack of reliable statistical distribution of τ_C and τ_E measured in this way. As a result, the statistical accuracy of Monte Carlo modelling for AC RTN remains to be verified for multiple traps and devices in real circuits, especially in term of its capability to predict the AC RTN beyond the test time window.

The objectives of this Chapter are threefold:

- to provide statistical AC RTN data at device level, rather than a single trap;
- to develop a model for AC RTN, that can not only fit test data, but also predict the long-term AC RTN;
- to analyse test data and simulation results, leading to an improved understanding of AC RTN.

Unlike the early works that focus on individual traps [8], [80-81], in this work, AC RTN data with a time window up to 3×10^4 sec will be collected at a device level that can have multiple traps. This can be used to verify the capability and accuracy of the long-term prediction by a model extracted from short-time test data.

In Chapter 5, an integral methodology is proposed for long-term RTN prediction and used to model DC RTN. The applicability of this methodology to AC RTN will be tested in this Chapter. Moreover, the dependence of AC RTN on frequency and time window will be analysed. Lastly, both acceptor-like and donor-like traps are modelled and their differences in carrier tunnelling will be studied.

6.2 Devices and Experiments

6.2.1 Devices used in this work

nMOSFETs fabricated by a 28 nm CMOS process were used in this work. The channel width and length are 90 nm and 27 nm, respectively. The gate dielectric consists of a stack with a Hf-high-k layer and an interfacial SiON layer with an equivalent oxide thickness of 1.2 nm. The devices are arranged in arrays and each of them is accessed through address codes. The average threshold voltage, Vth, is 0.45 V.

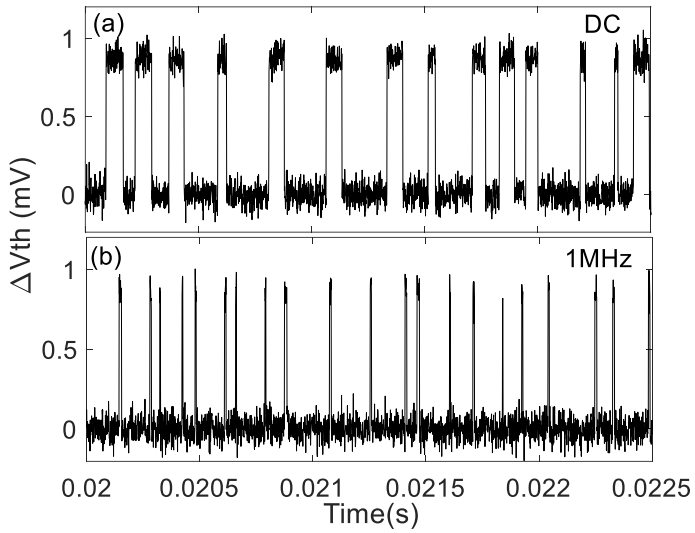


Fig. 6.1 An example of measured data for a device with one acceptor-like trap. Vg is under DC (a) and 1 MHz (b). The duration of Vg=0 was removed in (b).

6.2.2 Experiments

The tests were carried out with Vg alternating between 0.5 V and 0 V under a constant drain bias of 0.1 V. The top Vg is chosen to be close to Vth to make the test relevant to low-power operation, where RTN can be significant [21], [30]. The frequencies of Vg are in a range of 100 Hz to 1 MHz, with a duty cycle of 50%. Temperature is 125ºC. To enable statistical analysis, 402 devices were tested for a relatively short time window of 7.8 sec. This window is the total time after the removal of the duration when Vg = 0. To verify that the AC RTN model

extracted from data in this short window can be used to predict long-term RTN, tests were also carried out with a time window up to 3×10^4 sec for 60 devices.

A pulse (3 μs) I_D-Vg was measured on a fresh device before RTN measurement. During RTN tests, the drain current, I_D, was monitored continuously at a sampling rate of 1 Mpoint/sec [123]. The reference I_D, Iref, was obtained from the average of the first 10 points and the threshold voltage shift was evaluated from ΔVth=(Iref- I_D)/ g_m , where g_m is the transconductance obtained from the pulse I_D-Vg for each device at Vg=0.5 V.

6.3 Results and Discussions

6.3.1 AC RTN data at device level

Fig. 6.1(a) and Fig. 6.1(b) compare DC and AC RTN when there is only one trap in a device. Following earlier works [49], [121-122], the duration of Vg=0 was removed in Fig. 6.1(b). The term 'time' or 'time window' is applied hereafter for the cumulative time under Vg=0.5 V. For a duty cycle of 50%, the Vg at high level shares the exact same time with Vg at low level. Hence, the actual AC RTN test time doubles the 'time window'.

Under AC, the trap in Fig. 6.1 spends more time at the low level of ΔVth and there are more transitions between the two levels, in agreement with early work [121]. The result supports that this trap is acceptor-like, capturing an electron and inducing a positive ΔVth. Because emission under Vg = 0 is more efficient, the trap can only hold the electron for a short time at the high level of ΔVth in Fig. 6.1(b).

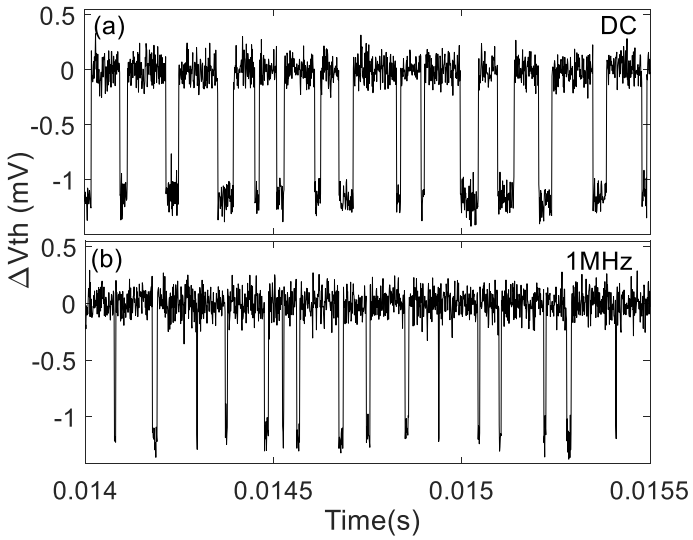


Fig. 6.2 An example of measured data for a device with one donor-like trap. Vg is under DC (a) and 1 MHz (b). The duration of Vg=0 was removed in (b).

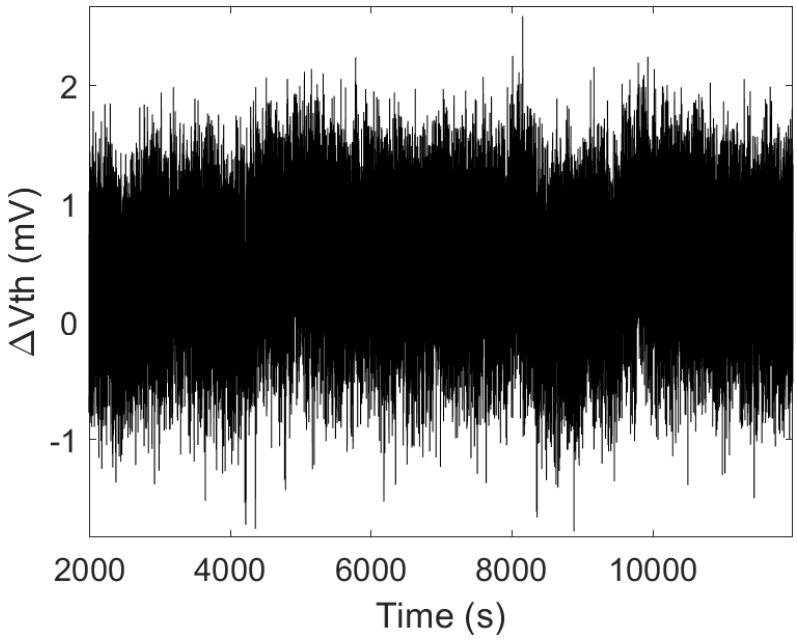


Fig. 6.3 An example of complex RTN over a long time window.

The positive ΔVth in Fig. 6.1(a) and Fig. 6.1(b) corresponds to a reduction of drain current, I_D. An increase of I_D was also often observed, which led to a negative ΔVth, as shown in Fig. 6.2(a) and Fig. 6.2(b). This negative ΔVth can come from either charging a donor-like trap positively or emitting an electron from a prefilled acceptor-like trap. Following the discussion in Chapter 5, the term 'donor-like trap' is used hereafter to represent ΔVth < 0. For AC RTN, Fig. 6.2(b) confirms that the emission of positive charge, i.e., capturing an electron and becoming neutral, is also more efficient under Vg = 0.

The devices in Fig. 6.1 and Fig. 6.2 have only one active trap within the measurement window. The trap time constant in this case can be readily extracted and used for Monte Carlo simulation. Many devices, however, do not have such a clear RTN signal. As the time window increases, the number of active traps in a device increases, resulting in complex fluctuation: an example is given in Fig. 6.3. This makes it difficult to extract the time constants of individual traps in this case. An integral methodology has been proposed in Chapters 4 and 5 to overcome this challenge and will be briefly described below.

6.3.2 Integral Methodology

This integral method uses the ΔVth measured at device level as inputs and does not require experimental separation of ΔVth into the contributions of individual traps. As a result, it removes the requirement of one device having only one trap and is applicable to devices with any number of traps. At a given time, ΔVth measured from multiple devices is grouped together to form a

dataset and their cumulative distribution function (CDF) is given in Fig. 6.4. The concept of 'Effective Charged Traps (ECT)" is introduced, which assumes that this CD comes from a set of traps that are always charged. In this way, the burden of finding the statistical distribution of trap time constants is removed, greatly simplifying the testing and analysis.

The next step is to extract the number of ECTs that can reproduce a given CD of ΔVth. By assuming the number of traps per device follows the Poisson distribution and selecting an RTN amplitude distribution, such as Exponential, Log-normal, or GEV, the average number of ECTs per device can be extracted by the maximum likelihood estimation for both acceptor-like (N_A) and donor-like (N_D) traps. The detail of the integral methodology is given in Chapter 4.

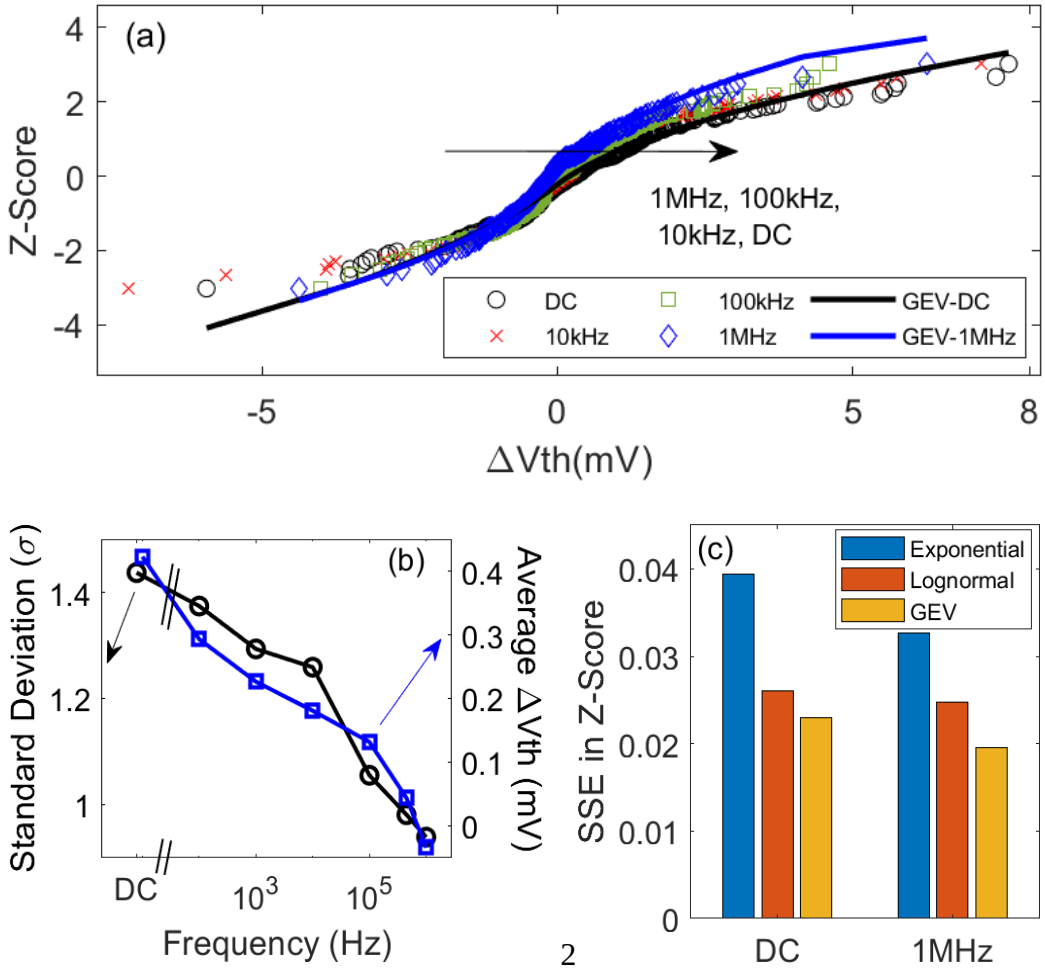


Fig. 6.4 (a) The Cumulative Distribution (CD) of test data (symbols). The lines are fitted function with GEV for DC (black) and 1 MHz (blue). (b) Dependence of standard deviation (σ) and mean ΔVth on frequency. (c) A comparison of the sum of squared error (SSE) per device of the CDF extracted for DC and 1 MHz RTN. The RTN amplitude distribution per trap is assumed to follow Exponential, Log-normal, and GEV, in turn. The time window is 7.8 s.

The process described above for extracting N_A and N_D is repeated for different time windows.

It is widely accepted that a longer time window covers slower traps and increases |ΔVth|, which in turn results in larger N_A and N_D. The N_A and N_D versus time obtained in this way can be used to extract their kinetics. For DC RTN, it has been shown in Chapter 5 that these kinetics can be used to predict N_A and N_D at longer times. Once N_A and N_D are known, the probability that a device will have a specific ΔVth can be calculated.

Table 6.1 The pdf formula and their extracted parameter values. δVth is the threshold voltage shift per trap.

	PDF of δVth	Acceptor	Donor
Exponential	$\frac{1}{\eta} e^{-\frac{\delta Vth}{\eta}}$	$\eta = 0.54$	$\eta = 0.49$
Lognormal	$\frac{1}{\delta Vth \theta \sqrt{2\pi}} e^{\left(-\frac{(\ln(\delta Vth)-\epsilon)^2}{2\theta^2}\right)}$	$\epsilon = -0.37$ $\theta = 0.14$	$\epsilon = -0.64$ $\theta = 0.16$
GEV	$\frac{1}{\beta}(k)^{\xi+1} e^{-k}$ $k = \left(1 + \xi\left(\frac{\delta Vth - \alpha}{\beta}\right)\right)^{-\frac{1}{\xi}}$	$\xi = 0.32$ $\alpha = 0.41$ $\beta = 0.36$	$\xi = 0.36$ $\alpha = 0.29$ $\beta = 0.23$
Thermal	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\Delta Vth}{\sigma}\right)^2}$	$Exponential, \sigma = 0.09$ $Lognormal, \sigma = 0.13$ $GEV, \sigma = 0.11$	

6.3.3 Applicability of Integral Methodology to AC RTN

ΔVth under different frequencies is compared in Fig. 6.4(a) and Fig. 6.4(b). The ΔVth, its mean value, and standard deviation is smaller for AC RTN compared to DC RTN, because of the enhanced emission during Vg=0. As a result, the RTN under AC operation cannot be modelled from DC RTN data and separate AC RTN measurements must be carried out. Since ΔVth can be either positive or negative, its mean value (<1 mV) is much smaller than the standard deviation.

The lines in Fig. 6.4(a) are the fitted cumulative distribution function (CDF) by the integral method. Their corresponding probability distribution function (PDF) and the extracted model parameters are given in Table 6.1. Fig. 6.4(c) shows that the error for AC is not larger than that of DC. Hence, it is concluded that the integral method is equally applicable for AC RTN. This is understandable: AC mainly impacts emission, but the integral methodology assumes CDF originating from ECTs that are always charged and a detailed emission process is not needed here.

The results in Fig. 6.4 were obtained for a time window of 7.8 s. The same procedure was applied at other time points between 10^{-4} s and 7.8 sec and the extracted N_A and N_D are given in Fig. 6.5 for Exponential, Log-normal, and GEV RTN amplitude distributions. As expected, both N_A and N_D increased with increasing time windows, but decreased for higher frequencies due to reduced trap occupancy. As shown in Fig. 6.5(a-f), the N_A and N_D decrease accordingly with increasing frequency in the order of DC, 10kHz, 100kHz and 1MHz.

Fig. 6.6 gives the extracted average threshold voltage shift per trap, μ. Unlike the N_A and N_D in Fig. 6.5, μ is insensitive to frequency. This can be explained as the same traps are responsible for AC and DC RTN. Although AC enhances emission, it does not change the RTN amplitude, as shown in Fig. 6.1 and Fig. 6.2.

The N_A and N_D versus time in Fig. 6.5 can be used to extract their kinetics. Three different kinetics were tested: power law, scaled Log-normal CDF, and Log-uniform. Fig. 6.5 shows that all three can fit the data reasonably well within the short time window of 7.8 s. One powerful feature of the integral methodology is that these kinetics can be extrapolated to a longer time window to predict the long-term DC RTN, making RTN prediction similar to predicting ageing induced by bias temperature instabilities [109], [119] and hot carriers [124]. The prediction capability for AC RTN will be tested and discussed next.

Fig. 6.7 gives the N_A and N_D for AC RTN measured in a time window of up to 3×10^4 sec. The three kinetics extracted from the data within 7.8 s are extrapolated and compared with the test data. The differences between the predictions by the three kinetics can be considerable and some kinetics clearly agree better with the test data based on Fig. 6.7. To make a quantitative comparison, Fig. 6.8 gives the errors between predictions and test data. When RTN amplitude is assumed to follow Exponential or Log-normal distribution, the scaled Log-normal CDF kinetics give lower errors. The lowest error, i.e., the best fit, however, is obtained from GEV

with Log-uniform kinetics. The two distributions selected for ΔVth (Vg (AC) = 0.5V and Vd = 0.1V) are,

1. Impact per trap: GEV Distribution

2. Long-Term RTN Kinetic: Log-uniform Distribution.

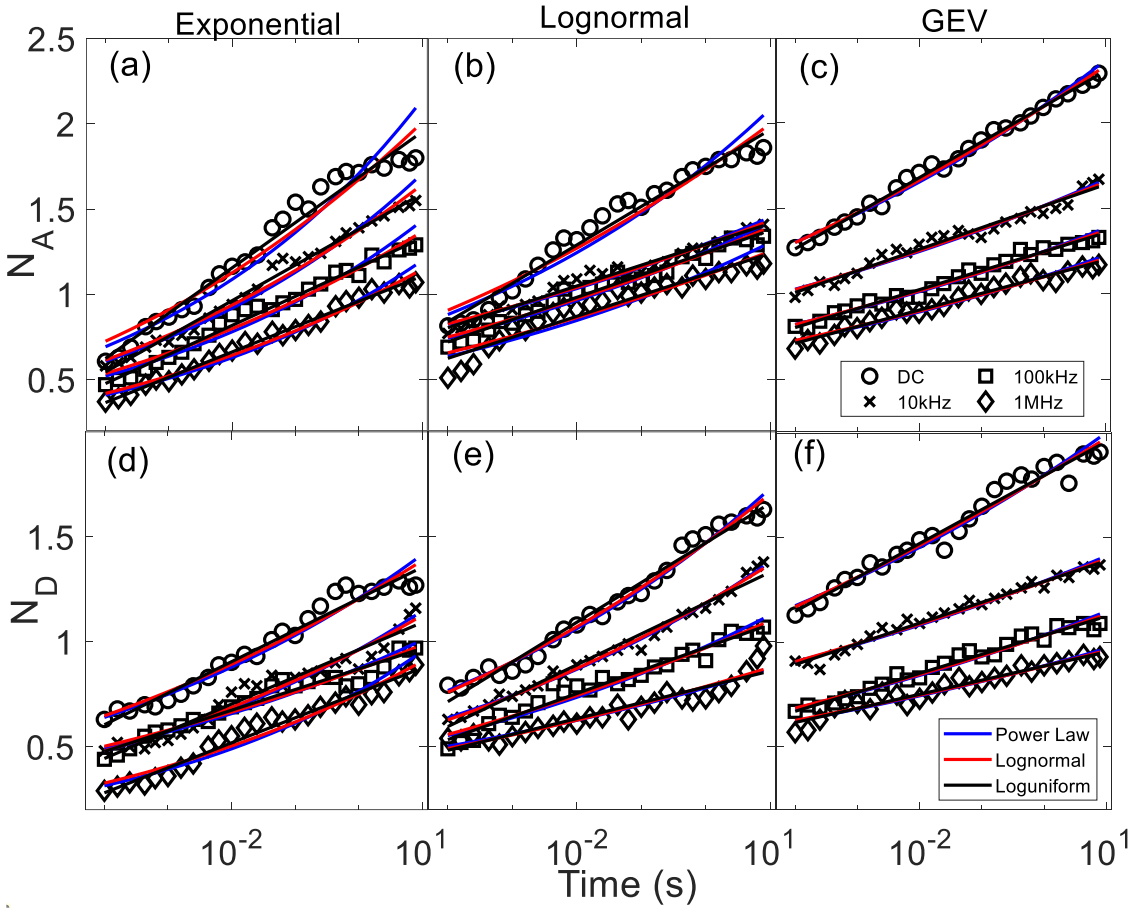


Fig. 6.5 Dependence of average number of ECTs per device on time window (symbols) at different frequencies. N_A in the top row (a,b,c) is for acceptor-like traps with Exponential, Log-normal, and GEV distributions respectively. N_D in the bottom row (d,e,f) is for donor-like traps. The lines are fitted kinetics with Power law (blue), scaled Log-normal CDF (red), and Log-uniform (black).

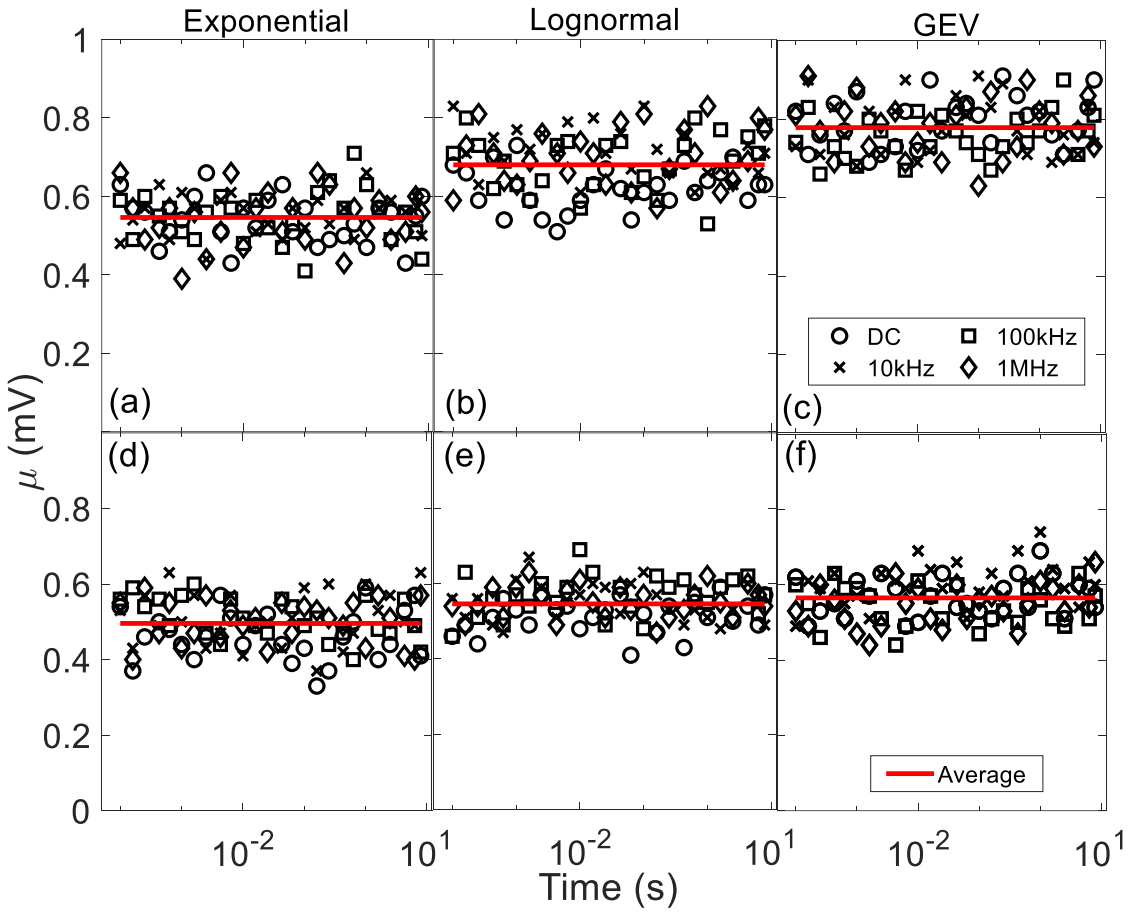


Fig. 6.6 The extracted average δVth per trap, μ, for DC and AC RTN at different frequencies. Solid lines are the mean values of all data. (a, b, c) are acceptor-like traps and (d, e, f) are donor-like traps.

We now investigate how well the N_A and N_D predicted in Fig. 6.7 can be used to evaluate the CDF at an AC time window of 3×10^4 sec. Fig. 6.9(a) and Fig. 6.9(b) compare the CDF calculated from the predicted N_A and N_D with the measured value and find good agreement. Fig. 6.9(c) shows that the prediction error for AC RTN is similar to that for DC RTN. Hence, the integral methodology can be used to extract a model from 7.8 s tests that is able to predict AC RTN at 3×10^4 sec, over three orders of magnitude ahead in time.

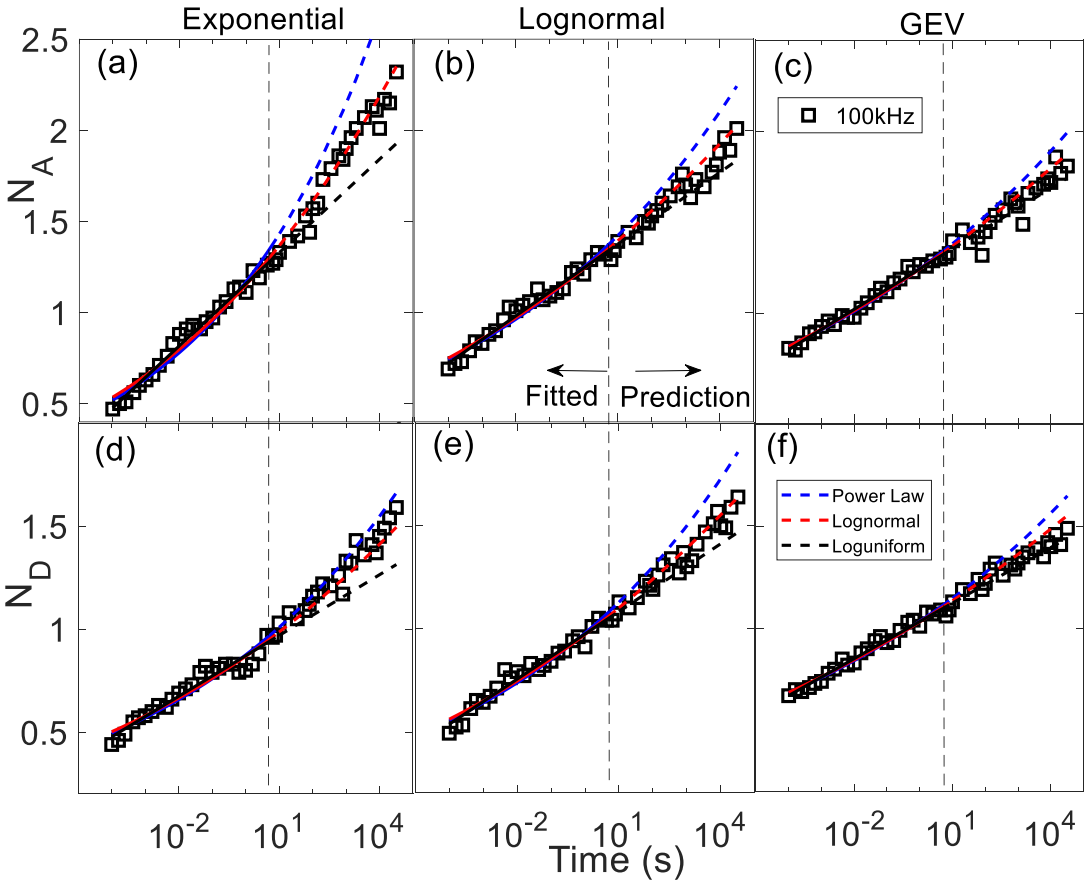


Fig. 6.7 Prediction of the average number of acceptor-like traps, N_A, in the top row (a, b, c) and donor-like traps, N_D, in the bottom row (d, e, f) for AC RTN at 100 kHz. Symbols are extracted by fitting the δVth distributions (Exponential, Lognormal, and GEV). The solid lines within 7.8 s are the fitted kinetics with Power law (blue), Log-normal (red), and Log-uniform (black). The dashed lines beyond 7.8 s are extrapolated from the solid lines. The symbols beyond 7.8 s were not used for fitting.

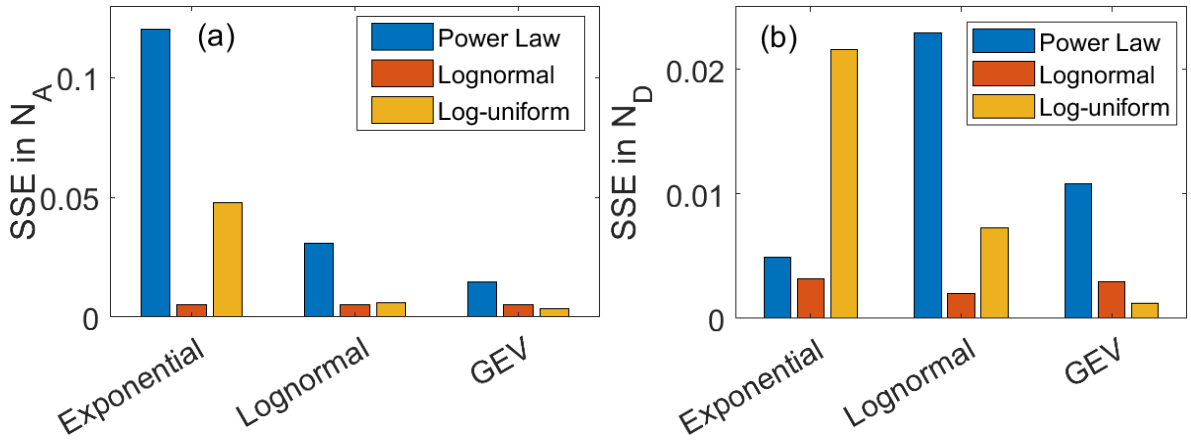


Fig. 6.8 The sum of squared errors per device of the prediction in Fig. 6.7 for acceptor-like (a) and donor-like (b) average number of traps per device. The lowest errors were obtained for Log-uniform kinetics with GEV δVth distribution.

A comparison of Fig. 6.4(a) with Fig. 6.9(a) shows that their ΔVth has similar range. Fig. 6.4(a) and Fig. 6.9(a) were obtained for time windows of 7.8 s and 3×10^4 s, respectively. One would

expect Fig. 6.9(a) to have a larger ΔVth range, as a larger time window allows slower traps to be included. This apparent discrepancy originates from two differences: (i) The number of devices used is 402 for the 7.8 sec test in Fig. 6.4(a) and only 60 for the $3x10^4$ sec test in Fig. 6.9(a); (ii) All data were saved for the 7.8 sec test, to give a total 7.8 M data points, while only 200 data points were saved per second to give a total 6 M data points for the $3x10^4$ sec test. This is further explained below.

A larger number of samples has more chances of capturing rare events and, in turn, increases the statistical range. To demonstrate the impact of sample number on the statistical range, the graph of measured data against time window is given in Fig. 6.10(a). The two blue lines represent the envelope of ΔVth where multiple traps were simultaneously charged. The envelope increases with the time window, as a larger window allows slower traps to be charged. However, there are only a few points that hit the envelope, as it is rare for multiple traps to be charged simultaneously. Within 7.8 sec, both the RTN data and envelope were sampled and saved at 1 MSample/sec. After 7.8 sec, the data was still measured at 1 MPoint/s, but the data was saved at only 200 Samples/sec to reduce the size of the dataset. The lower saving rate reduces the probability of catching a multiple trapping event; hence, the data range immediately after 7.8 s appears reduced, although the envelope is increasing. Fig. 6.10(b) shows the comparison between 400 points at 7.8 sec and 60 points at $3x10^4$ sec and, based on this result, they appear similar. This similarity, however, is an artefact, as the high number of samples at 7.8 s compensates the effect of the smaller time window.

6.3.4 Analysis of AC RTN

Fig. 6.11 shows N_A(AC)/N_A(DC) and N_D(AC)/N_D(DC) against the time window. The main feature is that the ratios initially decrease relatively fast with time and then more slowly for a longer time window. For a given trap, its emission time follows [1], [49], [64], [112], [120],

$$\tau_E \propto \exp\left(\frac{Ef - Et}{kT}\right) \tag{6.1}$$

At a given distance from the interface, a trap with its energy level Et further below Ef will have larger τ_E. When Vg is switched from Vdd to zero for AC RTN, it shifts (Ef-Et) by the same amount for traps of different Et. This reflects that τ_{EL}/τ_{EH} should be the same for fast and slow traps, so that one may expect N_A(AC)/N_A(DC) to be independent of the time window, against

the observed reduction in Fig. 6.11. To understand the reduction in Fig. 6.11, the impact of frequency on τ_E is examined next.

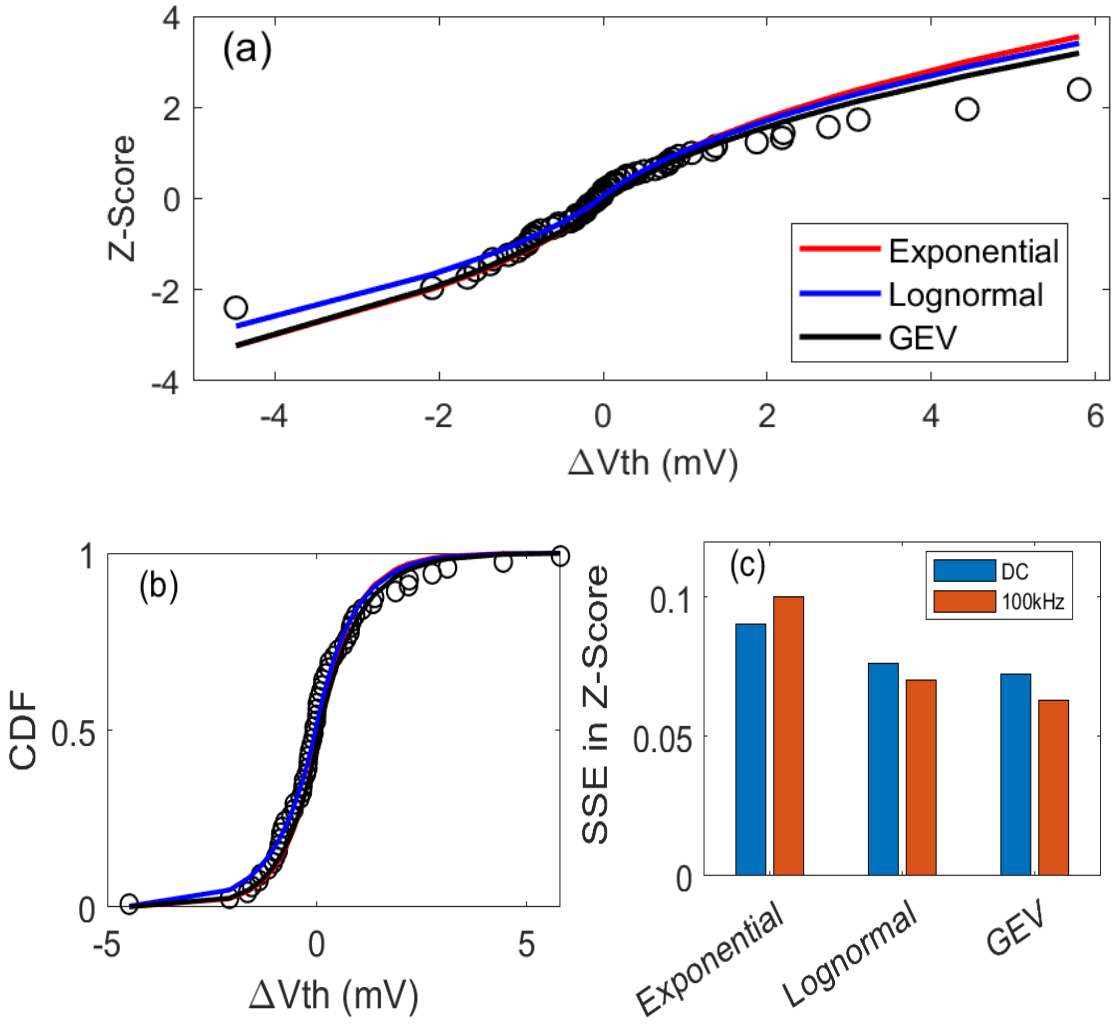


Fig. 6.9 AC (100 kHz) RTN prediction for the CDF of ΔVth at the time window of 3×10^4 s by the model extracted from test data within 7.8 s. Symbols are the measured data and the lines are the calculated CDF by using the predicted N_A and N_D in Fig. 6.7 and the average μ in Fig. 6.6, based on different δVth distributions. The CDF is plotted in Z-score in (a) and linearly in (b). (c) compares the errors of DC and AC predictions and shows that the smallest error was obtained for Log-uniform kinetic with GEV δVth distribution.

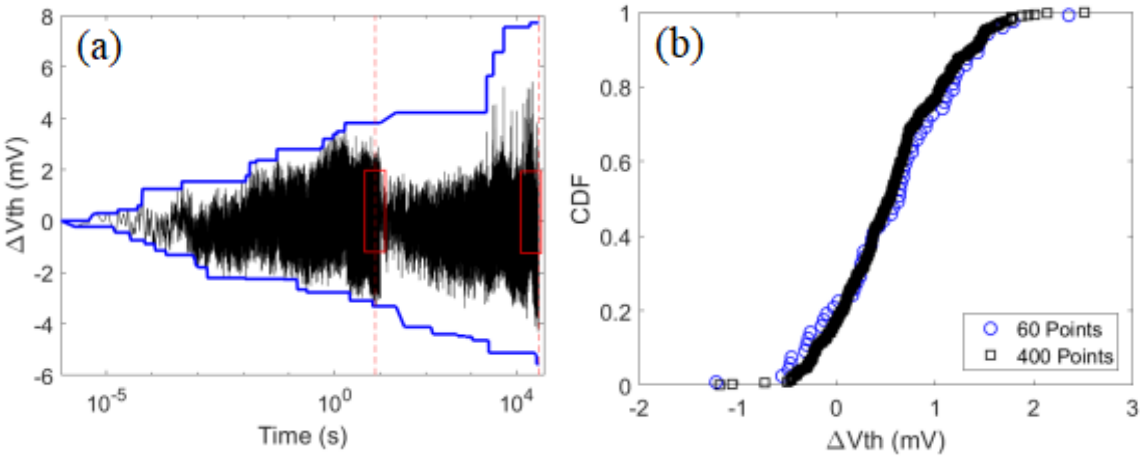


Fig. 6.10 (a) The impact of data saving rate on the range. At 7.8 sec, the saving rate is reduced from 1 MSample/sec to 200 Sample/sec. (b) A comparison of the CDF of 400 data points taken at 7.8 sec with that of 60 points taken at $3x10^4$.

When there is only one active trap in a device, such as in Fig. 6.1 and Fig. 6.2, its τ_C and τ_E can be readily extracted and one example each for acceptor-like and donor-like traps is given in Fig. 6.12(a) and Fig. 6.12(b), respectively. In agreement with some of these early works [3], [49], [121-122], τ_C is independent of frequency, f, because $\tau_{CL} >> \tau_{CH}$ and removing the Vg = 0 durations has little impact on capture.

On the other hand, it has been claimed that τ_C can be frequency dependent where the capture probability decreases, and capture time increases with increasing frequency [125-126]. The marginal increment of τ_C can be more than one order of magnitude depending on the selection of defect to be investigated [125-126]. However, based on Fig. 6.12, it is frequency independent up to 1Mhz. Results beyond 1Mhz is not available to further support it is frequency independent due to the limitation of instrument and requires further works.

Although τ_{EH} and τ_{EL} are determined by (Ef-Et) and should be independent of frequency, the τ_E measured at Vdd in Fig. 6.1(b) and Fig. 6.2(b) is neither τ_{EH} nor τ_{EL} and Fig. 6.12 shows that it can depend on frequency, in agreement with early work [3], [49], [121-122]. The dependence of τ_E on frequency can be divided into three regions. In the low-frequency Region 1 (R1) in Fig. 6.12, τ_E is insensitive to frequency. In Region 2 (R2), τ_E reduces for higher frequency. Finally, in high-frequency Region 3 (R3), τ_E becomes insensitive to frequency again. A physical interpretation of this dependence is given below.

In the inset of Fig. 6.12(a), t_H=1/(2f) is half of the AC period under Vg=0.5 V. The number of RTN transitions during t_H, n, is

$$n = \frac{2t_H}{\tau_{CH} + \tau_{EH}} \tag{6.2}$$

In R1, f is low enough to make $t_H >> (\tau_{CH}+\tau_{EH})$. There are many RTN transitions (both capture and emission events) within t_H, i.e., n>>1. During Vg=0, $t_L >> \tau_{EL}$, but when t_L is removed by joining the two dots in Fig. 6.12(a) together, the number of emissions will be increased by a maximum of 1 to (n+1)≈n where n is hardly affected, so that the measured $\tau_E \approx \tau_{EH}$, which is hardly affected by Vg = 0 duration and independent of frequency.

As frequency increases and t_H reduces in R2, n, reduces and approaches 1. If Vg = 0 duration, $t_L >> \tau_{EL}$, trapped charge will be emitted during t_L and the number of emissions during one AC period is (n+1)>n. The emission during t_L reduces the measured τ_E. When frequency is further increased and induced n<<1, (n+1)≈1 by removing t_L, the measured τ_E now is controlled and capped by frequency and is not equal to τ_{EH}. An increase in frequency reduces the duration of Vg = 0, t_L and, in turn, τ_E.

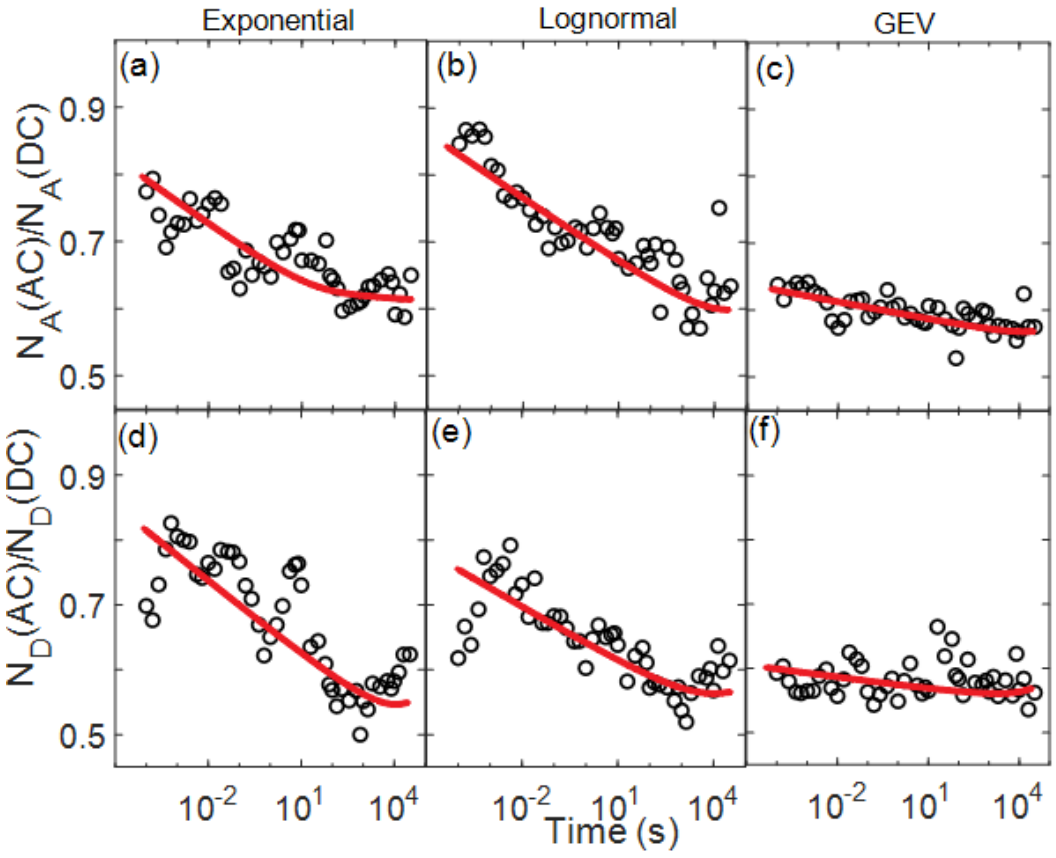


Fig. 6.11 AC (100 kHz) RTN against DC RTN ratio from 10^{-4} to $3x10^4$s for acceptor-like traps, N_A, in the top row (a, b, c) and donor-like traps, N_D, in the bottom row (d, e, f). The red lines are eye guides, showing that the reduction trend slows down at longer times.

Finally, when the frequency is high enough to make the time period of Vg = 0 short enough and fall in $t_L << \tau_{EL} < \tau_{EH}$ while n ≈ 0, it enters R3, where emission happens during both t_H and t_L to give

$$\tau_E = \frac{\tau_{EH}\tau_{EL}}{\tau_{EH} + \tau_{EL}} \tag{6.3}$$

τ_E now is independent of frequency again. If $\tau_{EL} << \tau_{EH}$, the trapped charged will be emitted at τ_{EL} instead of τ_{EH} hence, the measured $\tau_E \approx \tau_{EL}$, i.e., although the measurement is at Vg=Vdd, the measured τ_E is actually the emission time at Vg=0. In short, the measured τ_E is roughly equals to τ_{EH} in R1, and τ_E is roughly equals to τ_{EL} in R3 while τ_E will be controlled by frequency in R2.

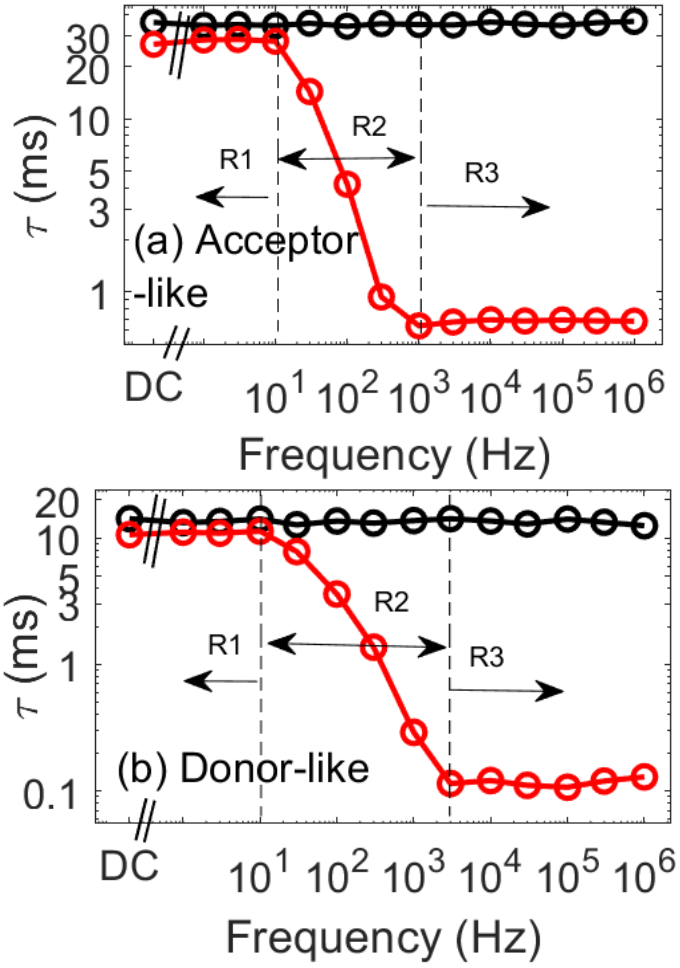


Fig. 6.12 Time constants, τ, versus frequency for a single acceptor-like trap (a) and donor-like trap (b). The red and black symbols are emission (τ_e) and capture (τ_c) time, respectively. R1, R2, and R3 are the three regions for the emission time.

Fig. 6.12 can be used to explain the results in Fig. 6.11. When the time window is small, active traps are fast and they can be either in R1 or close to the top of R2, where AC reduces τ_E

modestly, resulting in a relatively large N_A(AC)/N_A(DC) in Fig. 6.11. As time increases, slower traps become active and they are in R2, where AC causes an increasingly larger reduction of τ_E, leading to the reduction of N_A(AC)/N_A(DC). As time increases further, the new active traps are slow enough to be in R3, where the impact of AC on τ_E stabilizes, so that N_A(AC)/N_A(DC) only decreases slowly with time here.

The test facilities used in this work only allow reliable measurement to be made up to 1 MHz. There are traps that are too fast to make $t_L \ll \tau_{EL}$ even at 1 MHz, so that the results do not enter R3. Examples of this case are given in Fig. 6.13(a) and Fig. 6.13(b) for acceptor-like and donor-like traps, respectively. This explains the continuous reduction of N_A and N_D with frequency in Fig. 6.14.

AC RTN can be used to study the difference in tunnelling process between acceptor-like and donor-like traps. For acceptor-like traps, emission will be more efficient under Vg=0, if electrons are emitted to the substrate, as illustrated in Fig. 6.15(a). Thus, the enhanced emission of AC RTN supports that tunnelling occurs between traps and substrate.

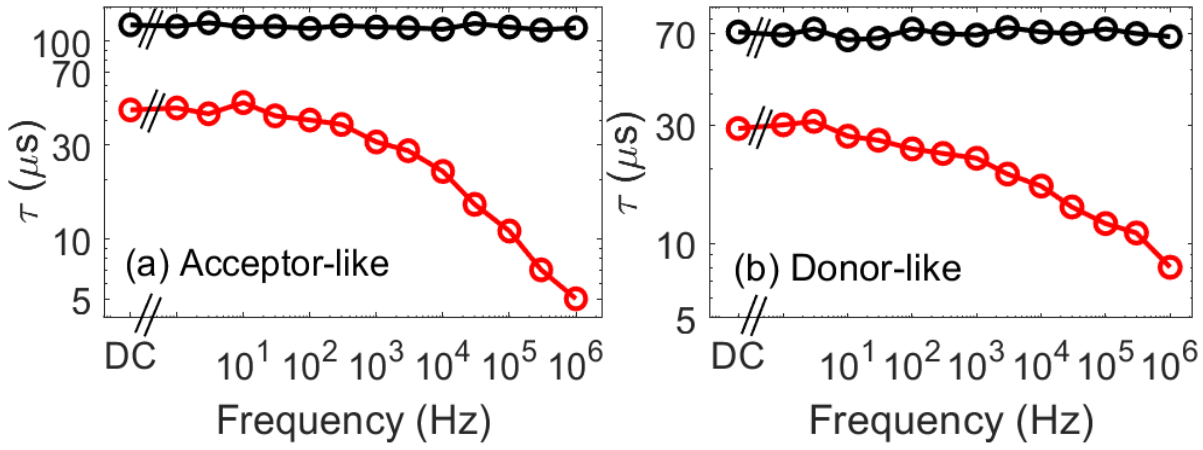


Fig. 6.13 Time constant, τ, versus frequencies for single acceptor-like trap (a) and donor-like trap (b). In contrast with Fig. 6.12, the emission time does not enter into region R3 even at 1 MHz here.

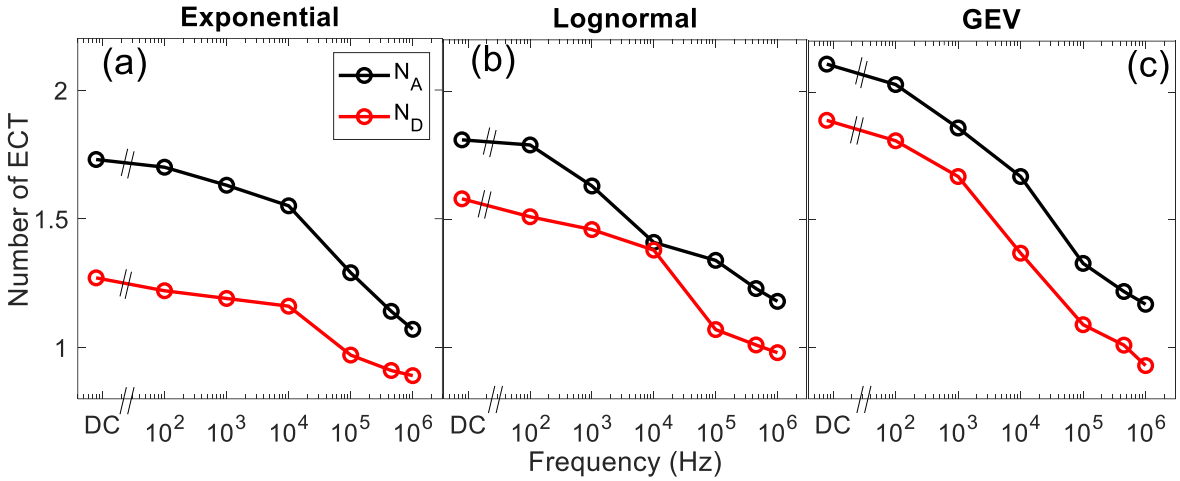


Fig. 6.14 Dependence of the extracted N_A (black) and N_D (red) at 7.8 s on frequency. (a,b,c) uses Exponential, Log-normal, and GEV δVth distribution, respectively.

For donor-like traps, AC also enhances the emission, as shown in Fig. 6.2(b), Fig. 6.12(b) and Fig. 6.13(b). To neutralize the positive charge or to restore the pre-trapped electron, an electron must tunnel to the trap. If this electron comes from the substrate, it should be less efficient under Vg = 0 than under Vg = +0.5 V, which disagrees with enhanced emission under AC. On the other hand, electron tunnelling to the trap is more efficient under Vg = 0, if it is from the gate, as illustrated in Fig. 6.15(b). As a result, acceptor-like and donor-like traps should be dominated by tunnelling from the substrate and the gate, respectively. For a given time window, one may also expect that active donor-like traps are closer to the gate. This is supported by the smaller average impact per donor-like trap in Fig. 6.6. On the relative tunnelling from the gate against that from the substrate, Fig. 6.15(c) and Fig. 6.15(d) shows N_D/N_A is around 0.8 and insensitive to either time or frequency.

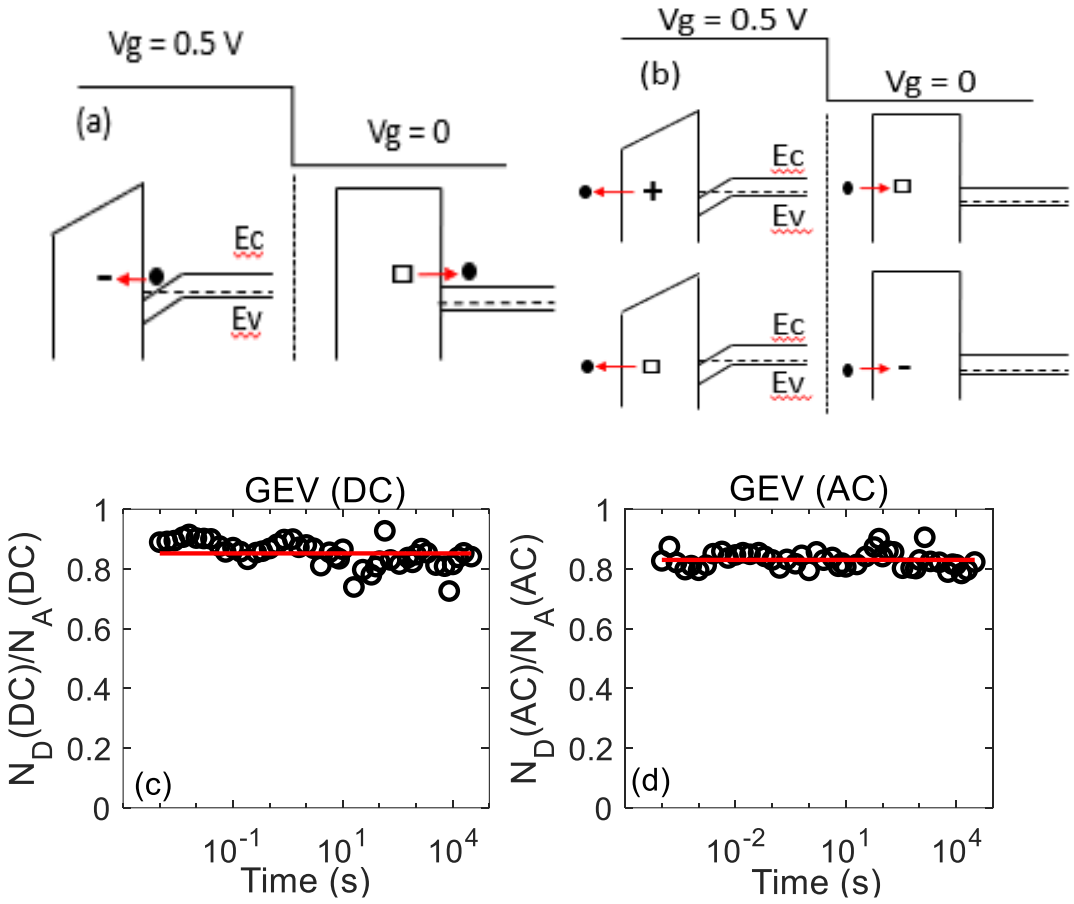


Fig. 6.15 A schematic illustration of carrier tunnelling for acceptor-like trap (a) and donor-like trap (b) to explain AC-enhanced emission. ND/NA is independent of time for both DC (c) and AC (d) RTN. ●, -, +, and □ represent electron, trapped negative, trapped positive charges, and neutral traps, respectively.

6.4 Conclusion

In summary, unlike early works that focused on individual traps, this work reports statistical experimental data for AC RTN at device level with a time window of up to 3×10^4 sec. The ΔVth values measured from multiple devices were grouped together to form a dataset and the integral methodology was found to be equally applicable to AC RTN as to DC RTN. At a given time, the CDF of ΔVth is used to extract the number of effective charged acceptor-like and donor-like traps through the maximum likelihood estimation. The N_A and N_D obtained from data within a time window of 7.8 s are used to extract their kinetics.

For the first time, it is shown that these kinetics can predict the N_A and N_D at 3×10^4 sec, a factor of 3,846 ahead in time. The CDF evaluated from the predicted N_A and N_D agrees well with test data. If one assumes that N_A and N_D kinetics extracted from a test time of one day can predict the AC RTN by the same factor of 3,846 ahead, it will cover a time range of more than 10 years.

The dependence of AC RTN on time window and frequency is analysed and RTN still reduces with frequency at 1 MHz, indicating there are traps with an emission time under Vg = 0 less than 1 μs. The AC RTN data support that there are substantial tunnelling activities between traps and both gate and substrate.

CHAPTER 7: CHARACTERIZING AND MODELING RTN UNDER REAL CIRCUIT BIAS CONDITIONS

This chapter studies the key parameters needed to model RTN for real circuit operation. In the discussion, both driving current and threshold voltage, as well as transistors under linear and saturation region are assessed by using the integral methodology. Follows by the accuracy of ΔVth extraction by g_m is investigated and the number of ECTs under linear and saturation regions is discussed. Lastly, the prediction of long-term RTN in linear driving current is assessed.

7.1 Introduction

To optimize circuit design, one needs to assess the challenges, a major challenge for designing modern CMOS circuits being device-to-device variation (DDV) [2-4], [19], [29], [31], [49]. There are two types of DDV: time-invariant DDV, such as line-edge roughness [127], and time-dependent variation (TDV), such as age-induced DDV [11]. As transistor sizes downscale, random telegraph noise (RTN) has attracted much attention and become one of the main sources of TDV [2-4], [29], [31], [49]. A single trap in gate dielectric can cause substantial fluctuation of both driving current, ΔI_D, and threshold voltage, ΔVth, by capturing/emitting a charge carrier from/to the conduction channel [2-4], [19], [29], [31], [49]. Moreover, the number of traps in a device follows a Poisson distribution and RTN induces substantial stochastic TDV [2-4], [29], [31], [49].

Despite the efforts made by early works on RTN [2-4], [29], [31], [49], accurately modelling and predicting RTN for circuit optimization remains a challenge for modern CMOS technologies and there are knowledge gaps that prevent this optimization being achieved. One such gap arises because early works focused on $\Delta I_D/I_D$ under linear operating condition, i.e. (Vg-Vth)>Vd, while real circuits can operate under not only linear, but also saturation modes. For example, Fig. 7.1 shows that the transistors M1 and M5 in a standard SRAM cell can operate in linear and saturation mode, respectively, during a read operation. The limited early works on RTN under saturation [128-131] mainly investigated the impact of drain bias on individual traps and the key information, such as the CDF per trap for $\Delta I_{D,SAT}/I_{D,SAT}$ and $\Delta Vth_{,SAT}$, is still missing. Without it, the simulation of RTN for real circuit operation cannot be completed.

In addition to driving current, circuit simulation also requires threshold voltage. Under linear mode, one common practice of early works [31], [41], [132-133] is to evaluate the RTN in $V_{th,LIN}$ by $\Delta V_{th,LIN} = \Delta I_{D,LIN}/g_m$, where $\Delta I_{D,LIN}$ is the fluctuation of linear I_D and g_m the transconductance. When both $\Delta I_{D,LIN}$ and g_m were measured at $V_g \approx V_{th}$, $\Delta V_{th,LIN}=\Delta I_{D,LIN}/g_m$ was a good estimation of the real ΔV_{th} [4]. Most designed circuits and early works, however, use an operating voltage of $V_g>V_{th}$. Under $V_g>V_{th}$, it has been reported that $\Delta V_{th,LIN} = \Delta I_{D,LIN}/g_m$ substantially overestimates the real $\Delta V_{th,LIN}$ directly measured from the pulse I_D-V_g at $V_g \approx V_{th}$ [4].

The impact of a trapped charge on a device depends on the local charge carrier density beneath it [2-4], [49]. Under $V_g \approx V_{th}$, current flows through narrow paths and traps located away from these paths have little impact [4], [30], as shown in Fig. 7.2. In contrast, under $V_g>V_{th}$, current distribution is more uniform, and more traps can be charged and make a substantial contribution to RTN. This explains why $\Delta V_{th,LIN} = \Delta I_{D,LIN}/g_m$ measured under $V_g > V_{th}$ is larger than the $\Delta V_{th,LIN}$ measured at $V_g \approx V_{th}$. At present, there are insufficient data to establish the CDF of real $\Delta V_{th,LIN}$ and there are even fewer data on the directly measured $\Delta V_{th,SAT}$. As the current flow pattern under $V_g \approx V_{th}$ is different from that under $V_g > V_{th}$, one open question is whether the RTN in driving current and Vth follows the same CDF.

Hence, the objectives of this Chapter are:

- To provide experimental data for RTN under saturation condition;
- To directly measure ΔV_{th} at $V_h \approx V_{th}$ and obtain its statistical data;
- To extract the CDF per trap of the four parameters needed for circuit simulation: $\Delta I_{D,LIN}/I_{D,LIN}$, $\Delta I_{D,SAT}/I_{D,SAT}$, $\Delta V_{th,LIN}$, and $\Delta V_{th,SAT}$.

In addition, based on the RTN measured at $V_g \approx V_{th}$, an integral methodology for modelling RTN has been proposed in Chapters 4, 5 and 6 and this can be used to predict the long-term RTN. The predicative capability of this integral method for $\Delta I_{D,LIN}/I_{D,LIN}$ measured under $V_g > V_{th}$ will be tested in this Chapter.

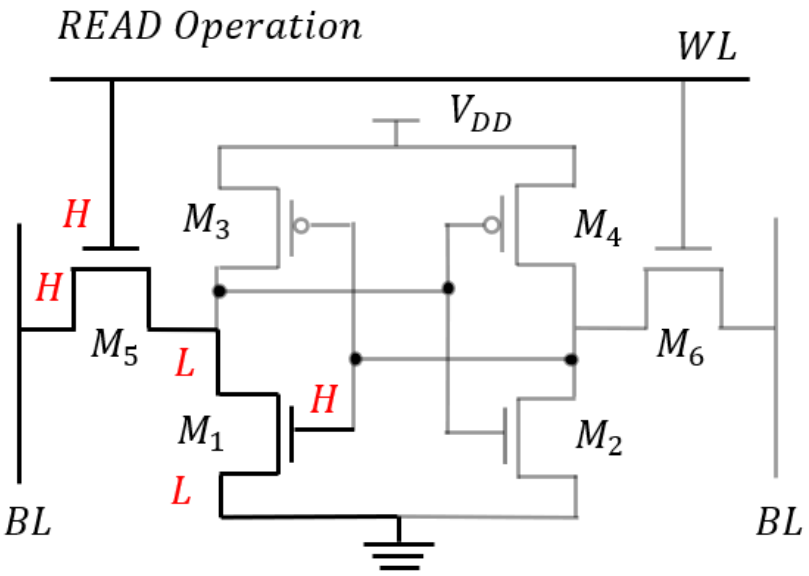


Fig. 7.1 Schematic 6-transistor SRAM cell under Read Operation. M1 and M5 operates under linear and saturation conditions, respectively.

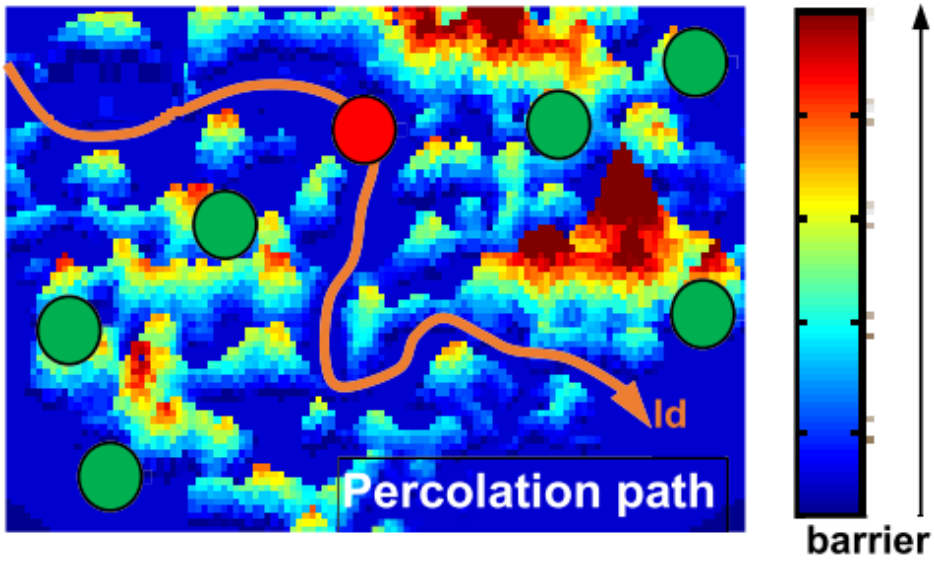


Fig. 7.2 A schematic illustration of the impact of traps (circles) on current path near threshold condition [30].

7.2 Devices and Experiments

7.2.1 Devices

The nMOSFETs used in this Chapter were fabricated by a commercial 28 nm CMOS process. The channel length and width are 27 and 90 nm, respectively. The gate stack consists of a metal layer, a Hf-high-k layer, and an interfacial SiON layer. The equivalent oxide thickness is 1.2 nm. The average threshold voltage, V_{TH}, is 0.45 V.

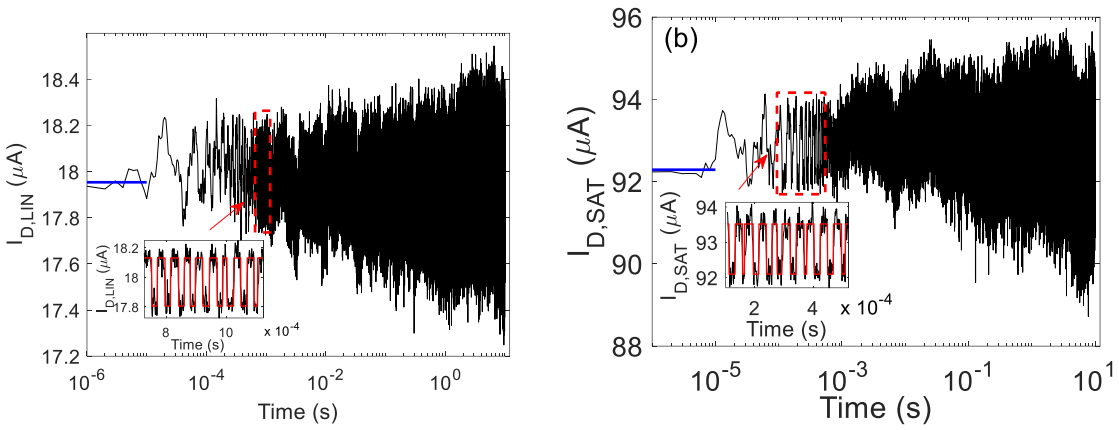


Fig. 7.3 The raw ID recorded against log(time) under linear (a) and saturation (b) modes, respectively. The solid blue lines represent the $I_{D,REF}$. The insets show the two-level RTN within a relatively short time window. The red lines in the insets were fitted RTN by the hidden Markov model (HMM).

7.2.2 Experiments

The tests were carried out under either linear or saturation conditions. Linear condition has a constant gate bias, Vg, of 0.9 V and a constant drain bias, Vd, of 0.1 V. The saturation condition has Vg = Vd = 0.9 V. Under these biases, the hot carrier and PBTI ageing are insignificant when compared with RTN fluctuation.

All tests were carried out under 125°C. This is the temperature of thermal chuck, and the self-heating effect is not included, since the test devices are bulk MOSFETs, rather than FinFETs. To assess RTN at different temperatures, more tests are needed, which is out of the scope of this work.

The test was repeated on 402 devices with a time window of 10 sec for statistical analysis. To verify that the RTN model extracted from data in this short window can be used to predict long-term RTN, tests with a time window up to 6×10^4 sec were also carried out for 60 devices.

During RTN tests, drain current, I_D, was monitored continuously at a sampling rate of 1 MSa/sec [30] and an example of the raw I_D is given in Fig. 7.3(a) and Fig. 7.3(b) for linear and saturation modes, respectively. The reference I_D, Iref, was obtained from the average of the first 10 points, as represented by the blue lines; it was used for evaluating the relative shift of I_D, $\Delta I_D/I_D=(Iref-I_D)/I_D$ so that a positive $\Delta I_D/I_D$ represents a reduction of I_D. The insets in Fig. 7.3(a) and Fig. 7.3(b) show that I_D can follow the two-level RTN within a relatively short time window; as time increases, slower traps start contributing, leading to increased fluctuation.

The test was periodically interrupted by ramping Vg down from 0.9 V to 0.3 V to obtain the pulse $I_D - Vg$ at that time point. The Vd used for the pulse I_D-Vg is 0.1 or 0.9 V as a constant for the linear or saturation tests, respectively. The Vg waveform for measuring ΔVth is given in Fig. 7.4(a). The reference Vth, $Vth_{,REF}$, is first extracted by the maximum g_m method from a reference pulse (3 μs) I_D-Vg, taken before starting the RTN test for each device and an example is given in Fig. 7.4(b). The current at Vg = $Vth_{,REF}$, I_{DTH}, is recorded. The recorded I_D-Vg pulse at each time point can be used to compare with the I_{DTH}. The instabilities in Vth were extracted from the Vg shift at I_D=I_{DTH}, i.e., ΔVth= Vg (I_D =I_{DTH})-$Vth_{,REF}$, as shown by the inset of Fig. 7.4(b).

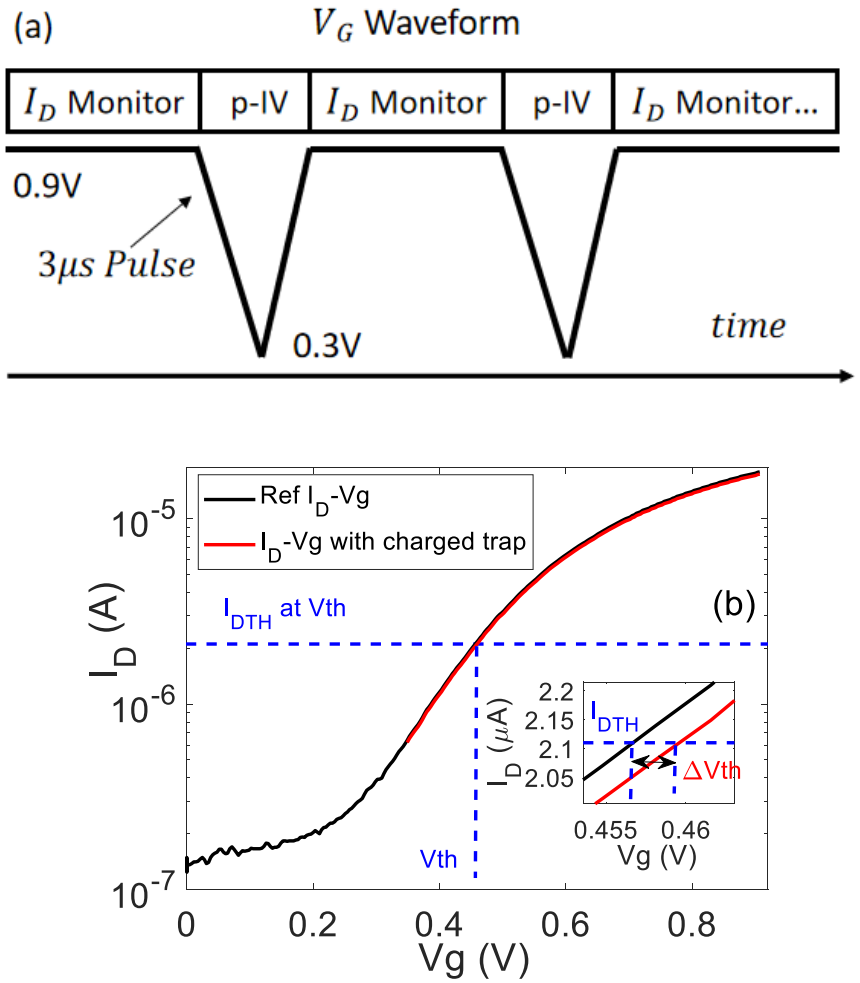


Fig. 7.4 (a) The Vg waveform for measuring ΔVth. (b) The ΔVth was extracted from the shift of Vg at I_D=I_{DTH} when comparing the pulse I_D-Vg during RTN test with the reference I_D-Vg measured before starting the RTN test.

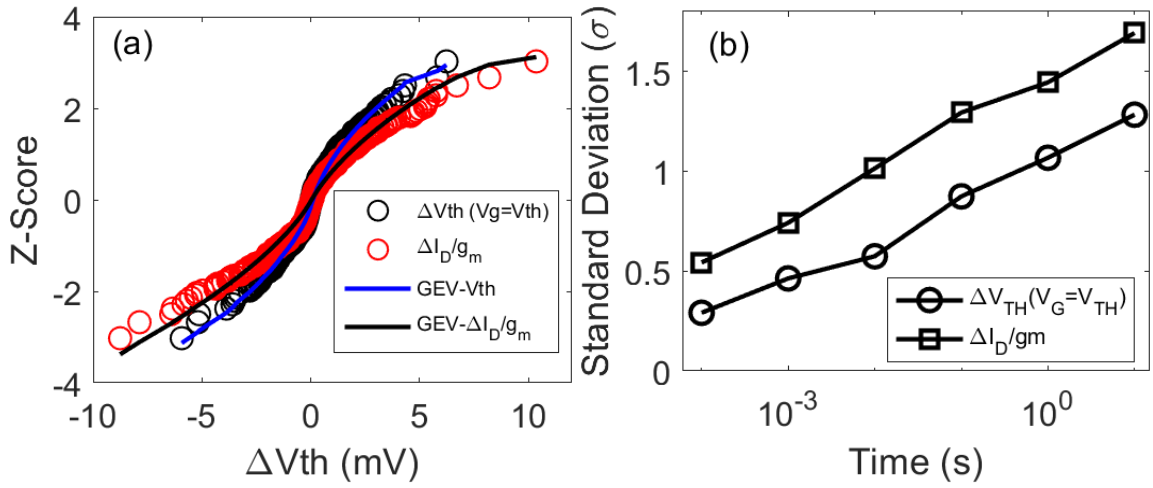


Fig. 7.5 A comparison between ΔVth measured at Vg=Vth and $\Delta I_D/g_m$. (a) The CDF of ΔVth and $\Delta I_D/g_m$ at 10 sec. (b) Standard deviation at different time.

7.3 Results and Discussions

7.3.1 Linear Mode: CDF of $\Delta Vth_{,LIN}$ and $\Delta I_{D,LIN}/I_{D,LIN}$

Integral methodology: The details of integral methodology were given in Chapters 4, 5 and 6, and a brief description of its principle is included here for self-completeness. Early works [3], [19], [24], [27], [29], [31], [61], [64], measured the impact of individual traps separately and collected a number of traps to build the CDF per trap. In contrast, the integral method measures the total RTN in a device, which can come from the cumulative contribution of multiple traps. One advantage of the integral method is that it does not require separating the measured fluctuation, such as that in Fig. 7.3(a) and Fig. 7.3(b), into the contributions of individual traps experimentally. This separation is carried out numerically based on the maximum likelihood estimation as applied in Chapters 5 and 6. At a given time, each device will give one data point in Fig. 7.5(a) and 402 devices form the dataset for the CDF.

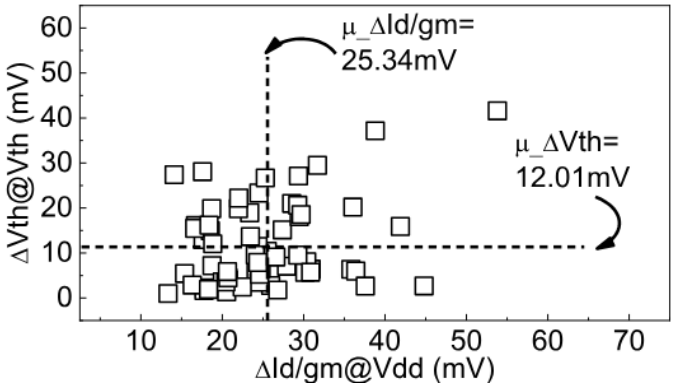


Fig. 7.6 Poor correlation between ΔId/gm at Vdd and ΔVth at Vg = Vth. Each point was taken from a different device. The dotted lines mark the mean values [4].

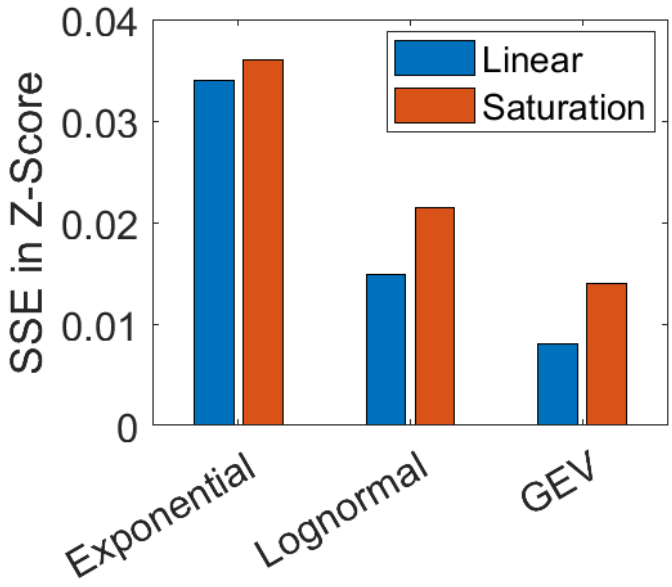


Fig. 7.7 The sum of squared errors (SSE) per device of the fitted CDFs on the ΔVth(Vg=Vth). Blue bars are for $\Delta Vth_{,LIN}$ in Fig. 7.5(a). Red bars are for ΔVth_{SAT} in Fig. 7.12(b).

It should be pointed out that the CDF in Fig. 7.5 is for $\Delta Vth_{,LIN}$ per device, which is different from the CDF per trap, as one device can have multiple traps. By assuming that the number of traps per device follows the Poisson distribution and the CDF per trap follows one of three distributions - Exponential, Log-normal, or General Extreme Value (GEV) - the parameters in these distributions can be extracted through the Maximum Likelihood method [39]. In this way, the integral method decomposes $\Delta Vth_{,LIN}$ into the contribution of individual traps numerically and statistically, rather than experimentally.

A comparison between $\Delta Vth_{,LIN}$ and $\Delta I_{D,LIN}/g_m$: Fig. 7.5(a) and Fig. 7.5(b) compare the CDF and standard deviations, respectively, of these two. In agreement with early work [4], $\Delta I_{D,LIN}/g_m$ has substantially larger deviation and longer distribution tails. It has been found out that the difference between $\Delta I_{D,LIN}/g_m$ and $\Delta Vth_{,LIN}$ can be more than half, as shown in Fig. 7.6 [4]. Consequently, using $\Delta I_{D,LIN}/g_m$ will overestimate the RTN in threshold voltage due to the current is more uniformly distributed and led to more traps can be charged under Vg > Vth. $\Delta Vth_{,LIN}$ should be directly measured at Vg = Vth, as shown in Fig. 7.4(b). Hereafter, the directly measured $\Delta Vth_{,LIN}$ is applied in this work.

Selection of CDF per trap for $\Delta Vth_{,LIN}$: By assuming the impact per trap following a specific CDF - Exponential, Log-normal, or GEV - the parameters in these CDFs can be extracted through the maximum likelihood estimation as applied in Chapters 5 and 6. To find which CDF

best matches the test data, the sum of squared error is calculated (SSE) per device between the extracted CDF and the test data. The blue bars in Fig. 7.7 show that GEV has the least SSE.

One possible explanation for GEV showing the lowest errors is that it has three fitting parameters, while Log-normal has two and Exponential has one. To further support the selection of GEV, we examine the dependence of SSE on the number of devices, N. If the data truly follows a CDF, SSE should reduce with increasing N. As N approaches infinite, SSE per data/device should approach zero [30].

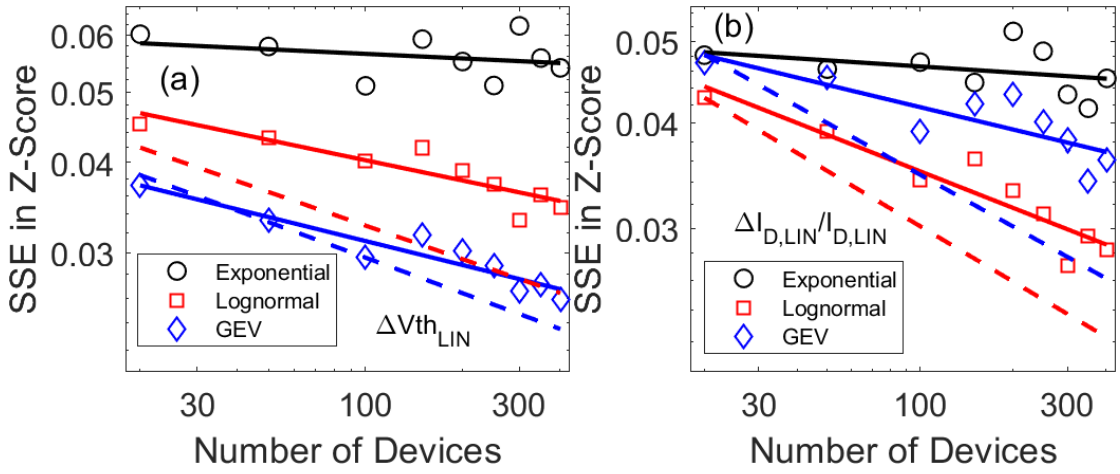


Fig. 7.8 Dependence of SSE per device on the number of devices used to extract the CDFs. The solid lines are fitted with test data. The dashed lines are results of hypothetical devices generated from theoretical Lognormal or GEV CDFs. (a) $\Delta Vth_{,LIN}$: GEV (blue) agrees best with the theoretical line. (b) $\Delta I_{D,LIN}/I_{D,LIN}$: Log-normal (red) agrees best with theoretical line.

For the red dashed line in Fig. 7.8(a), a theoretical Log-normal CDF was applied to generate different numbers of hypothetical devices. These devices were then fitted with Log-normal CDF and the SSE per device was calculated from the difference between the fitted and the theoretical Log-normal CDFs. As expected, the SSE indeed reduced for a higher number of devices. Repeating the same procedure for GEV gives the blue dashed line. When comparing the theoretical dashed lines with the corresponding solid lines of test data, GEV (blue) is the closest, supporting its selection for $\Delta Vth_{,LIN}$.

Selection of CDF per trap for $\Delta I_{D,LIN}/I_{D,LIN}$: The CDF of measured $\Delta I_{D,LIN}/I_{D,LIN}$ at Vg = 0.9 and Vd = 0.1 V at 10 sec is given in Fig. 7.9(a). By repeating the same procedure as that for $\Delta Vth_{,LIN}$, the SSE of extracted CDF per device is given in Fig. 7.9(b) as the blue bars. The Log-normal has the lowest error here, although it has less fitting parameters than GEV. Fig. 7.8(b) also shows that differences in the SSE dependence on device number between the

theoretical and test data are smallest for Log-normal (the red lines). Hence, Log-normal should be selected for $\Delta I_{D,LIN}/I_{D,LIN}$.

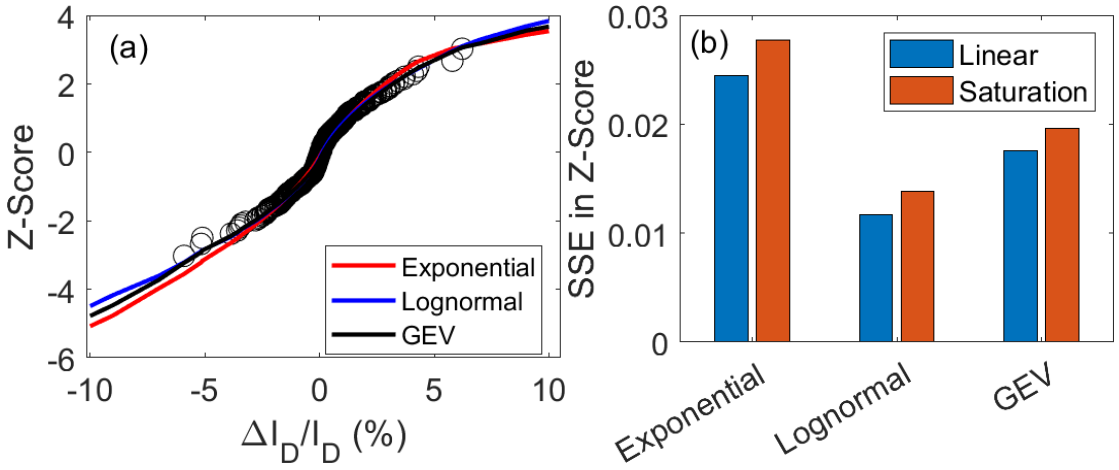


Fig. 7.9 $\Delta I_{D,LIN}/I_{D,LIN}$ fitted with three different CDF per trap: Exponential, Log-normal, and GEV. (b) The errors. Blue bars are errors for $\Delta I_{D,LIN}/I_{D,LIN}$ in (a) and red bars are errors for $\Delta I_{D,SAT}/I_{D,SAT}$ in Fig. 7.12(a). In both cases, Log-normal gives the lowest errors.

The results show that different types of CDF per trap should be used for $\Delta I_{D,LIN}/I_{D,LIN}$ and $\Delta Vth_{,LIN}$: Log-normal for $\Delta I_{D,LIN}/I_{D,LIN}$ and GEV for $\Delta Vth_{,LIN}$. To explain this difference, it is noted that GEV generally has a wider statistical spread than Log-normal, when fitting the same set of data [30]. As mentioned earlier, the current follows a narrow percolation path under Vg=Vth, but is more uniformly distributed under Vg>Vth. The trapped charges above the narrow current path will have profound impacts on the device under Vg=Vth, generating outliers and a wide statistical spread, as shown in Fig. 7.2. GEV describes these outliers better.

Impact of Vg on RTN: This work measures RTN at Vg=0.9 V, since this is the operating voltage specified for this CMOS technology by the foundry. The desire to minimize power consumption may push Vg toward Vth = 0.45 V in future, so that it is of interest to compare the RTN at Vg = 0.9 V with that at Vg = 0.5 V. Fig. 7.10 shows that RTN has a higher standard deviation, σ, at Vg = 0.9 V. When compared with I_D at Vg = 0.5 V, I_D at Vg = 0.9 V is higher and more uniformly distributed. This allows more traps to be charged, resulting in a higher RTN.

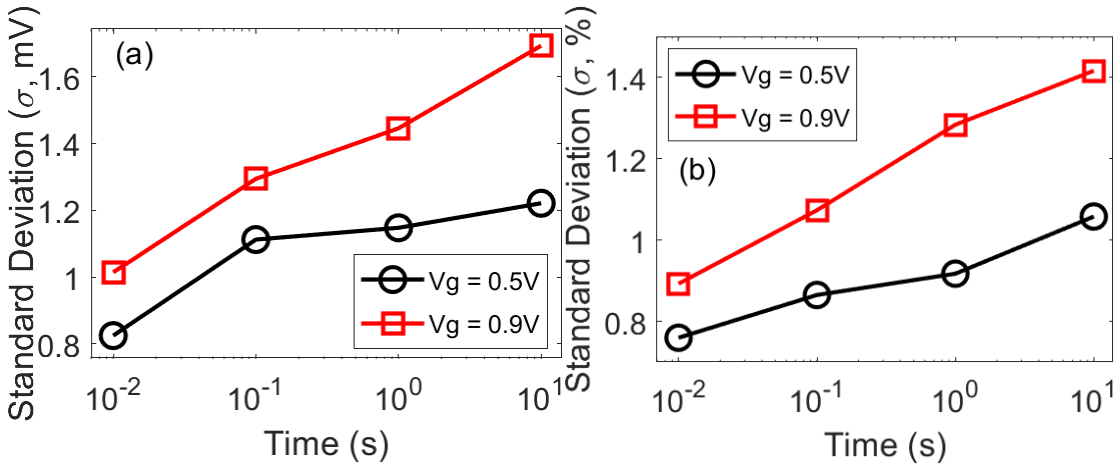


Fig. 7.10 A comparison of the RTN standard deviation under Vg = 0.9 V with that under Vg = 0.5 V. (a) ΔVth and (b) $\Delta I_{D,LIN}/I_{D,LIN}$.

7.3.2 Saturation Mode: CDF of $\Delta I_{D,SAT}/I_{D,SAT}$ and $\Delta Vth_{,SAT}$

The CDF per trap for $\Delta I_{D,SAT}/I_{D,SAT}$: The measured $I_{D,SAT}$ and $\Delta I_{D,SAT}$ on 402 devices are given in Fig. 7.11(a) and Fig. 7.11(b), respectively. When compared with $I_{D,LIN}$, the mean $I_{D,SAT}$ is a factor of 4.8 higher. The $\Delta I_{D,SAT}$ in Fig. 7.11(b) has a wider statistical spread than the $\Delta I_{D,LIN}$. This is mainly caused by the higher transconductance in saturation, so that the same shift in overdrive voltage (Vg-Vth) results in a larger shift in I_D.

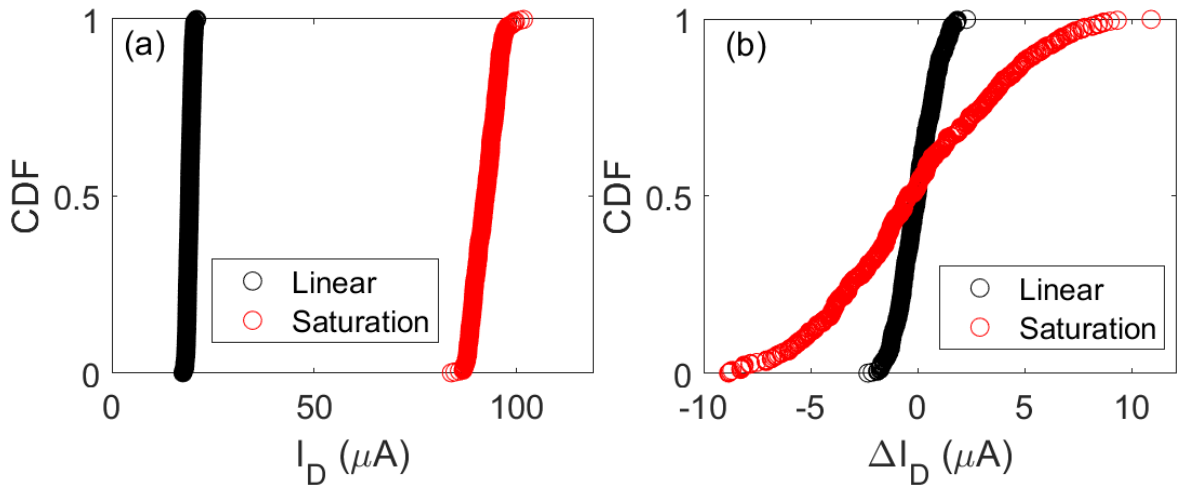


Fig. 7.11 A Comparison between linear (black) and saturation (red) measurements on 402 devices: (a) I_D and (b) ΔI_D.

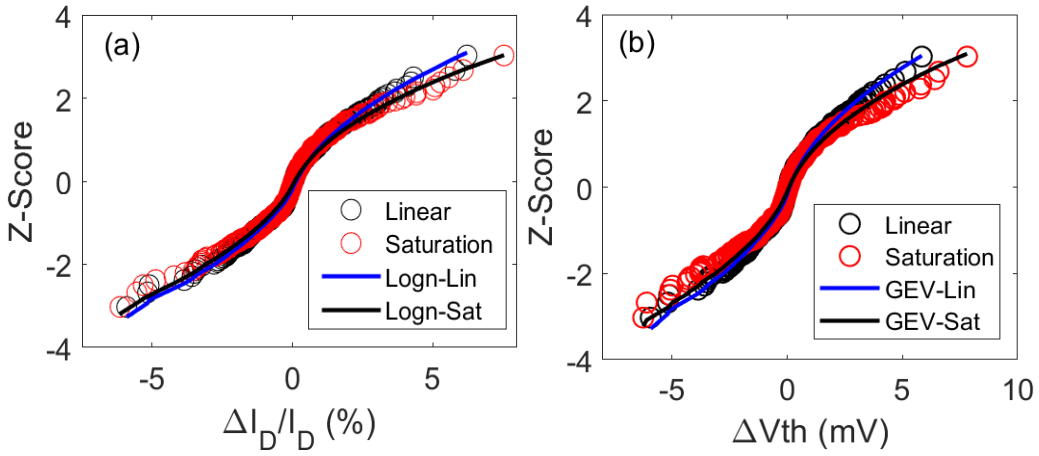


Fig. 7.12 A Comparison between linear and saturation measurements: (a) $\Delta I_D/I_D$ and (b) ΔVth. The lines are fitted results.

Fig. 7.12(a) compares $\Delta I_{D,SAT}/I_{D,SAT}$ with $\Delta I_{D,LIN}/I_{D,LIN}$. The normalization brings these two much closer, but $\Delta I_{D,SAT}/I_{D,SAT}$ is clearly higher and has a wider statistical spread than $\Delta I_{D,LIN}/I_{D,LIN}$. The integral methodology is again used to extract the CDF per trap for $\Delta I_{D,SAT}/I_{D,SAT}$. As reported in Section 7.3.1 , GEV best describes the RTN with a narrow percolation current path, while Log-normal CDF better describes the RTN under a relatively uniform current flow. Under saturation, current flow should be relatively uniform near the source, but can be less uniform near the pinch-off point, where it is close to threshold condition. Fig. 7.9(b), however, shows that Log-normal CDF has the lowest error for $\Delta I_{D,SAT}/I_{D,SAT}$. This indicates that the inversion channel region has a larger impact on $\Delta I_{D,SAT}/I_{D,SAT}$ distribution.

To further compare $\Delta I_{D,SAT}/I_{D,SAT}$ with $\Delta I_{D,LIN}/I_{D,LIN}$ quantitatively, the statistical properties are extracted at different time windows. Fig. 7.13 shows that the number of average effective charged acceptor-like (N_A) and donor-like (N_D) traps is clearly larger for $\Delta I_{D,SAT}/I_{D,SAT}$. The procedure and formula used to extract the N_A and N_D for $\Delta I_D/I_D$ are the same as those used for ΔVth in Chapter 5 and 6, but the unit for average impact per trap, μ, is % for $\Delta I_D/I_D$ and mV for ΔVth. Fig. 7.14 shows that μ is insensitive to time and in a range of 0.48 ~ 0.75% for N_A, and 0.46 ~ 0.6% for N_D, depending on the CDF per trap used.

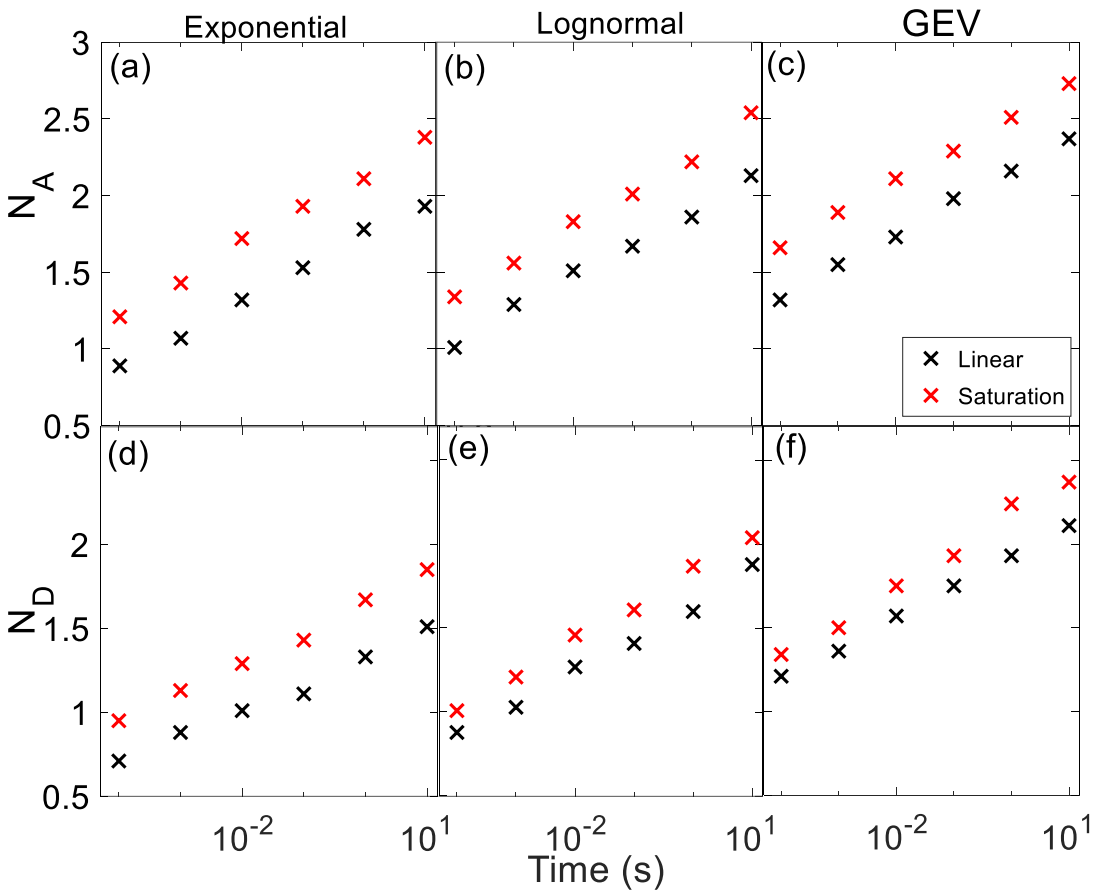


Fig. 7.13 $\Delta I_D/I_D$ at different time windows: The extracted average number of acceptor-like (a)-(c) and donor-like (d)-(f) traps per device. The red and black symbols are for $\Delta I_{D,SAT}/I_{D,SAT}$ and $\Delta I_{D,LIN}/I_{D,LIN}$, respectively.

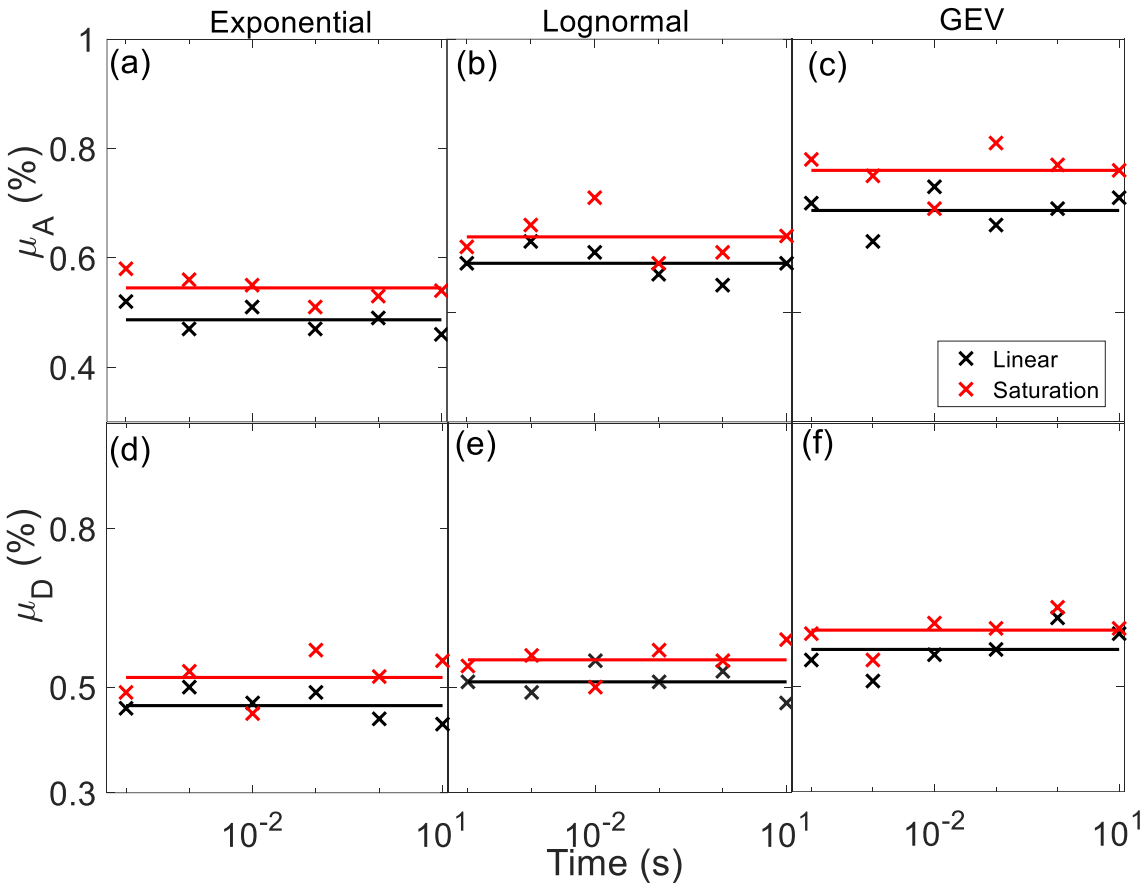


Fig. 7.14 $\Delta I_D/I_D$ at different time windows: The extracted average impact per acceptor-like, μ_A, (a)-(c) and donor-like, μ_D, (d)-(f) trap. The red and black symbols are for $\Delta I_{D,SAT}/I_{D,SAT}$ and $\Delta I_{D,LIN}/I_{D,LIN}$, respectively.

Based on the extracted Log-normal CDF, the projected $\Delta I_{D,SAT}/I_{D,SAT}$ and $\Delta I_{D,LIN}/I_{D,LIN}$ at different multiples of standard deviation, σ, are given in Fig. 7.15(a). At 6σ, $\Delta I_{D,SAT}/I_{D,SAT}$ = 35.7% and $\Delta I_{D,LIN}/I_{D,LIN}$ = 27.5%. Typically, $\Delta I_D/I_D$ = 10% has been used to define ageing-induced device lifetime [134]. At 10%, Fig. 7.16(a) shows that the failure rates for $\Delta I_{D,SAT}/I_{D,SAT}$ and $\Delta I_{D,LIN}/I_{D,LIN}$ are 90 and 30 parts-per-million (ppm), respectively. At 25%, they become 0.17 for $\Delta I_{D,SAT}/I_{D,SAT}$ and 0.008 for $\Delta I_{D,LIN}/I_{D,LIN}$. These differences confirm the need to model $\Delta I_{D,SAT}/I_{D,SAT}$ and $\Delta I_{D,LIN}/I_{D,LIN}$ separately.

Selection of CDF per trap for $\Delta Vth_{,SAT}$: Similar to $\Delta Vth_{,LIN}$, $\Delta Vth_{,SAT}$ was measured from the shift of pulse I_D-Vg, but under Vd = 0.9 V. Their distributions are compared in Fig. 7.12(b). Like $\Delta I_D/I_D$, $\Delta Vth_{,SAT}$ is higher and has higher statistical spread than $\Delta Vth_{,LIN}$. As both of them were measured under the condition where current follows narrow percolation paths, it is not surprising to find that GEV also best describes the CDF per trap for $\Delta Vth_{,SAT}$.

The larger RTN in saturation mode should be explained. The occupancy of a trap in the gate dielectric should increase with

i. a higher number of mobile charge carriers beneath it in the substrate, and
ii. the higher energy of these charge carriers.

When an electron in silicon has higher energy, the potential barrier for its tunnelling will be lower. As a result, hot carriers are more efficient to fill a trap and increase the trap occupancy.

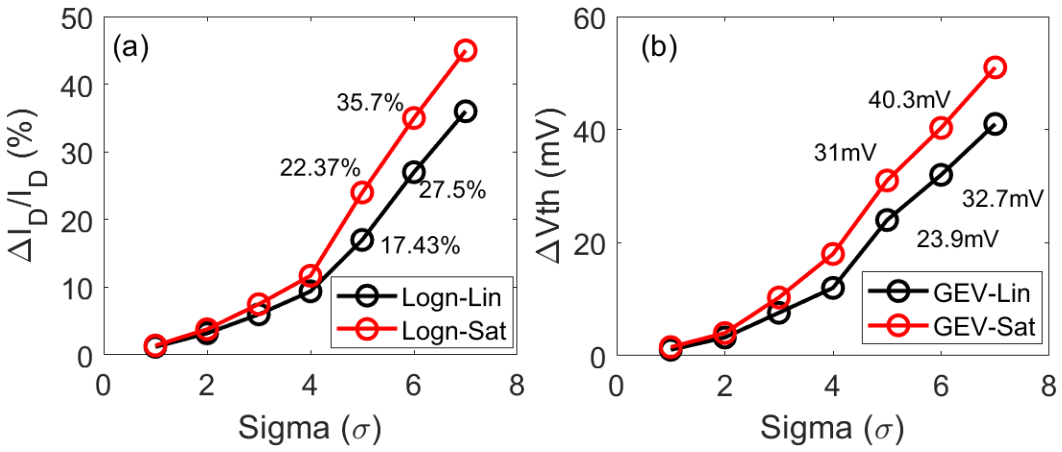


Fig. 7.15 A comparison of linear and saturation $\Delta I_D/I_D$ (a) and ΔVth (b) at different multiples of standard deviation, evaluated from Log-normal for (a) and GEV for (b).

Under the same Vg = 0.9 V, raising Vd from 0.1 to 0.9 V increases the non-uniform electrostatic field distribution. This reduces the electrical field between gate and the substrate channel away from the source, leading to a reduction of the charge carrier density towards the drain, which in turn reduces the trap occupancy. Hence, the reduced field between the gate and the channel does not support a high $\Delta Vth_{,SAT}$. On the other hand, the lateral field increases with rising Vd, which makes carriers 'hot' and increases trap filling. The experimental results show that the rise in carrier energy overcompensates the reduction in carrier density, resulting in higher $\Delta Vth_{,SAT}$.

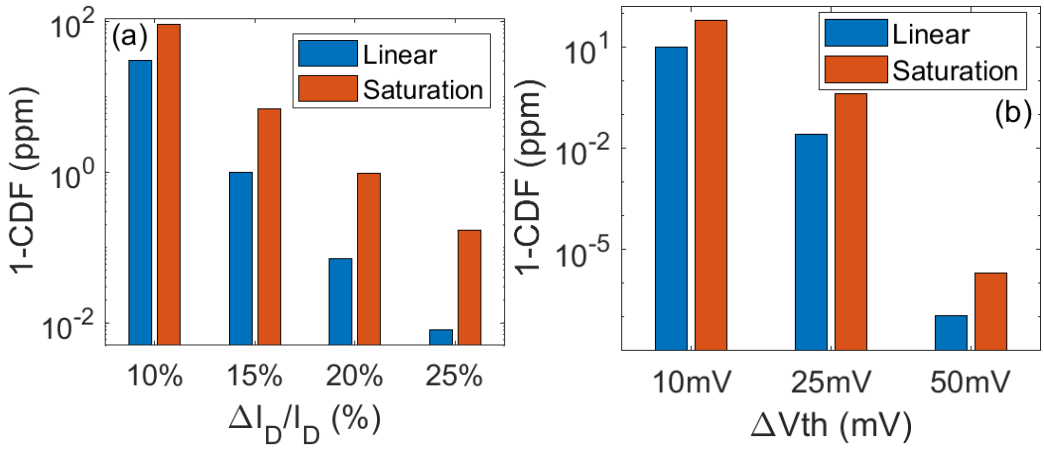


Fig. 7.16 The failure rates against failure criteria evaluated by log-normal CDF for $\Delta I_D/I_D$ (a) and GEV for ΔVth (b)

Table 7.1 The CDF formula and their average parameter values extracted between 10^{-4} and 10 sec. The 'δ' is used in the formula to emphasize that it is the impact per trap, rather than per device.

	CDF	Acceptor	Donor
Lognormal $\frac{\Delta I_{D,LIN}}{I_{D,LIN}}$	$\frac{1}{2}+\frac{1}{2}\mathrm{erf}\left(\frac{\ln(\delta I_{D,LIN}/I_{D,LIN})-\epsilon}{\sqrt{2}\theta}\right)$	$\epsilon = -0.53$ $\theta = 0.13$	$\epsilon = -0.68$ $\theta = 0.11$
Lognormal $\frac{\Delta I_{D,SAT}}{I_{D,SAT}}$	$\frac{1}{2}+\frac{1}{2}\mathrm{erf}\left(\frac{\ln(\delta I_{D,SAT}/I_{D,SAT})-\epsilon}{\sqrt{2}\theta}\right)$	$\epsilon = -0.47$ $\theta = 0.15$	$\epsilon = -0.61$ $\theta = 0.16$
GEV ΔVth_{LIN}	$e^{-\left(\left(1+\xi\left(\frac{\delta Vth_{LIN}-\alpha}{\beta}\right)\right)^{-\frac{1}{\xi}}\right)}$	$\xi = 0.54$ $\alpha = 0.46$ $\beta = 0.13$	$\xi = 0.4$ $\alpha = 0.37$ $\beta = 0.14$
GEV ΔVth_{SAT}	$e^{-\left(\left(1+\xi\left(\frac{\delta Vth_{SAT}-\alpha}{\beta}\right)\right)^{-\frac{1}{\xi}}\right)}$	$\xi = 0.51$ $\alpha = 0.49$ $\beta = 0.14$	$\xi = 0.38$ $\alpha = 0.4$ $\beta = 0.15$

To have a quantitative comparison, the number of effective charged traps per device is given in Fig. 7.17 for different time windows. The N_A and N_D are clearly higher for $\Delta Vth_{,SAT}$. Fig. 7.18 shows that the mean impact per trap, however, is similar for $\Delta Vth_{,SAT}$ and $\Delta Vth_{,LIN}$. Based on the extracted GEV per trap, Fig. 7.15(b) gives the predicted ΔVth at multiple σ at 10 sec. At 6σ, $\Delta Vth_{,SAT}$ and $\Delta Vth_{,LIN}$ reach 40.3 mV and 32.7 mV, respectively. Fig. 7.16(b) shows that the failure rates for $\Delta Vth_{,SAT}$ and $\Delta Vth_{,LIN}$ are 2×10^{-6} and 0.11×10^{-6} ppm at 50 mV. Therefore, they should also be modelled separately.

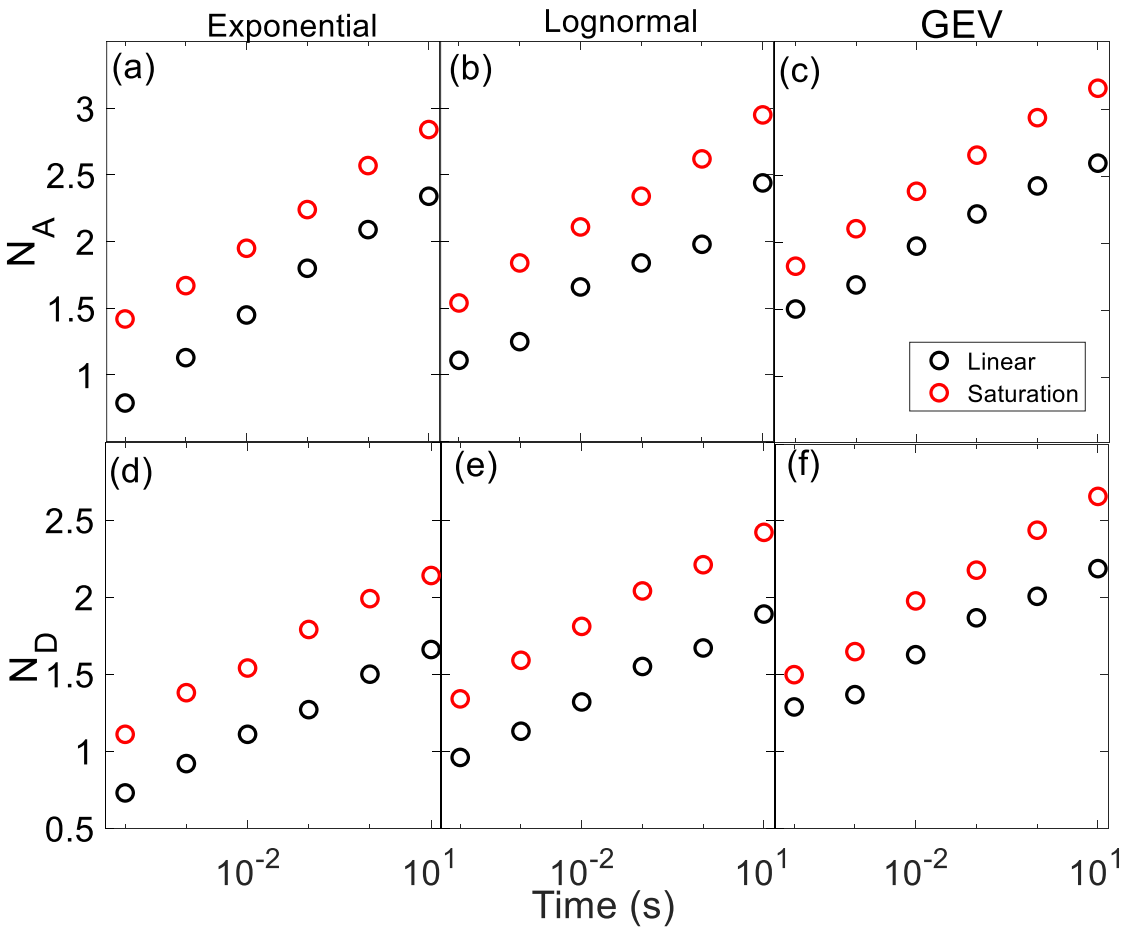


Fig. 7.17 ΔVth at different time windows: The extracted average number of acceptor-like (a)-(c) and donor-like (d)-(f) traps per device. The red and black symbols are for $\Delta Vth_{,SAT}$ and $\Delta Vth_{,LIN}$, respectively.

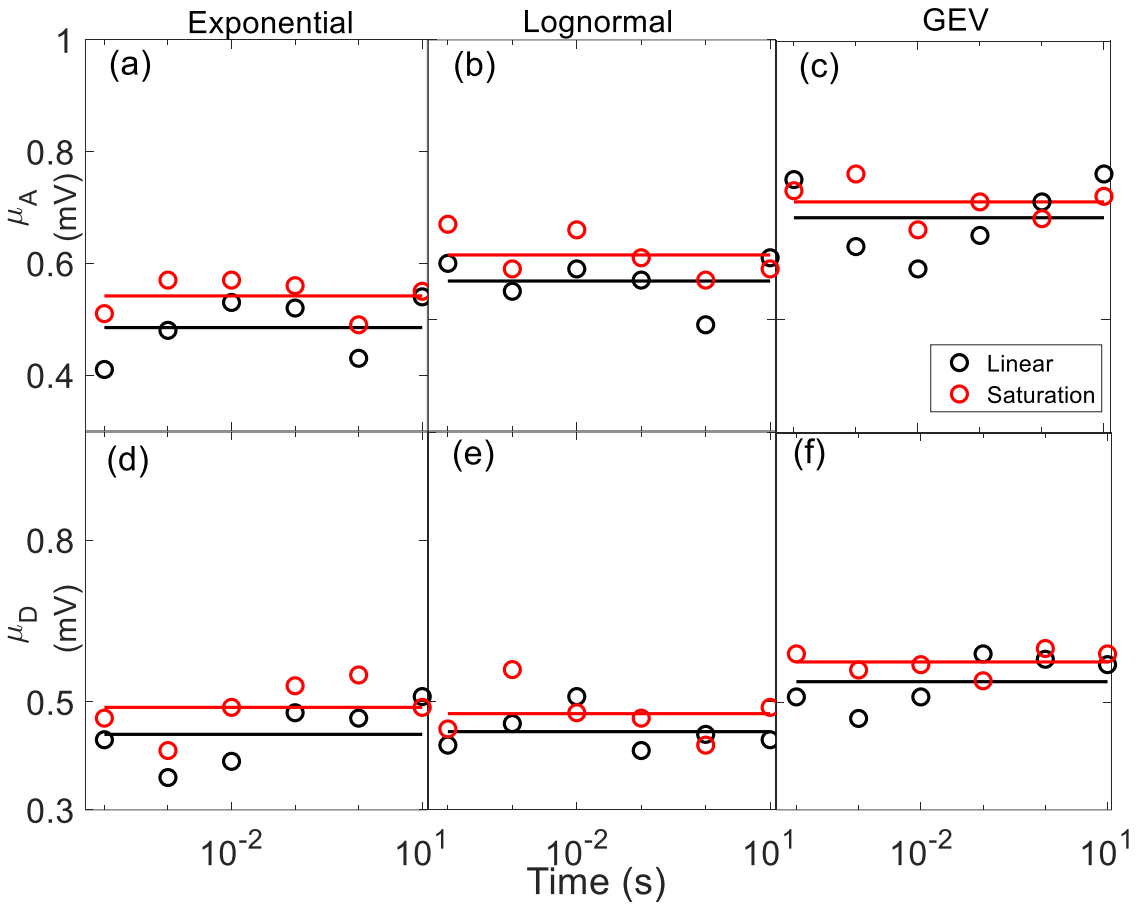


Fig. 7.18 ΔVth at different time windows: The extracted average impact per acceptor-like (a)-(c) and donor-like (d)-(f) trap. The red and black symbols are for $\Delta Vth_{,SAT}$ and $\Delta Vth_{,LIN}$, respectively.

7.3.3 Predicting long-term $\Delta I_{D,LIN}/I_{D,LIN}$

The principle: One advantage of the integral methodology is that it does not require the statistical distribution of capture and emission times, which are difficult to establish [21], [24], [27]. At a given time window, the statistical distribution of ΔVth such as that in Fig. 7.12(b) is assumed to originate from a set of 'effectively charged traps (ECT)'. As ECTs are always charged, there is no need to specify their capture and emission times. As the time window increases, slower traps become active, leading to higher N_A and N_D in Fig. 7.13 and Fig. 7.17. In this way, the impact of the trap's time constants on ΔVth is transformed to the dependence of N_A and N_D on time. If the kinetics of N_A and N_D can be established based on short-time test data, their long-term values can be estimated through extrapolation, similar to predicting the ageing-induced device lifetime. Once N_A and N_D are known, the long-term CDF of ΔVth can be directly evaluated, without using the time-consuming Monte Carlo simulation.

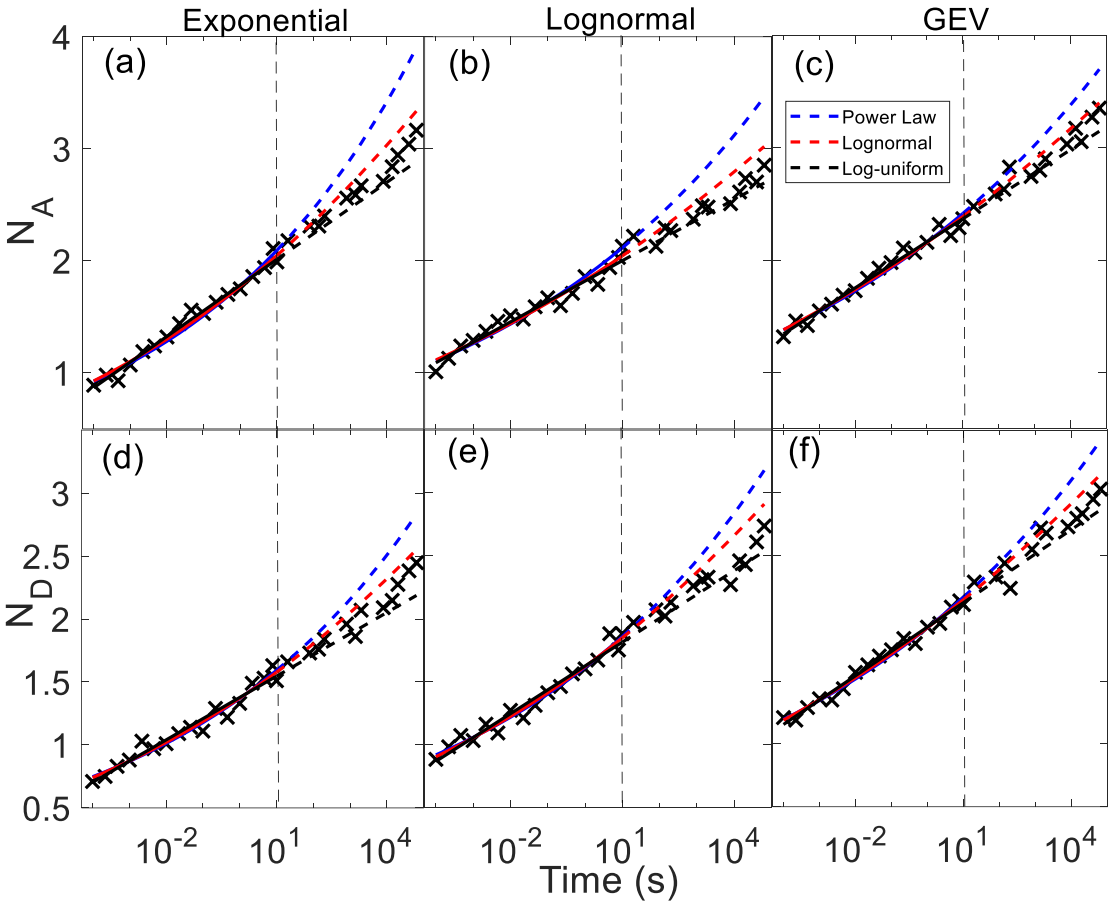


Fig. 7.19 $\Delta I_{D,LIN}/I_{D,LIN}$: Prediction of the average number of acceptor-like traps per device, N_A, in (a)-(c) and donor-like traps per device, N_D, in (d)-(f). Symbols are extracted from three different distributions: Exponential, Lognormal and GEV. Solid lines within 10 s represent the fitted kinetics with power law, lognormal and log-uniform. Dashed lines beyond 10 s represent the extrapolated kinetics. The test data beyond 10 s were not used for fitting.

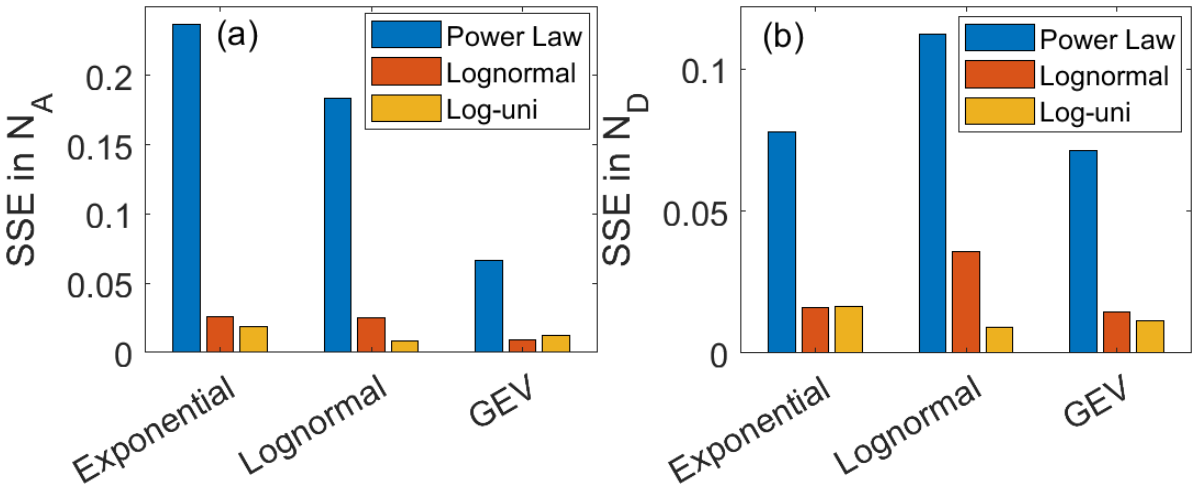


Fig. 7.20 The sum of squared errors of the prediction in Fig. 7.19: (a) for acceptor-like and (b) for donor-like traps. The lowest errors were obtained for Log-uniform kinetics with Lognormal CDF.

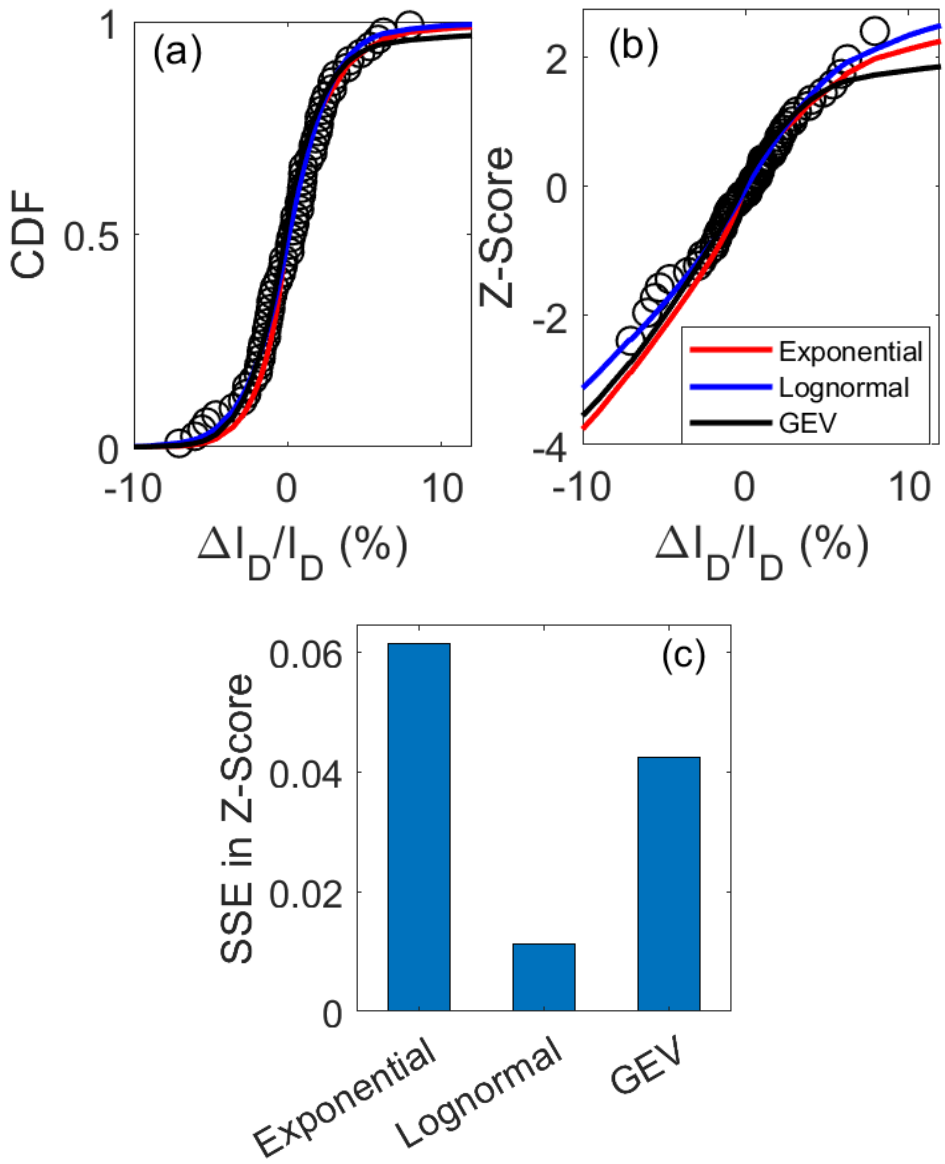


Fig. 7.21 A comparison of the measured and predicted CDF of $\Delta I_{D,LIN}/I_{D,LIN}$ at 6×10^4 sec. The CDF is plotted in linear (a) and Z-score (b) scales. (c) shows that the lowest error was obtained with Lognormal CDF per trap.

Applicability to $\Delta I_{D,LIN}/I_{D,LIN}$: Chapters 5 and 6 showed that the integral method was successfully used to predict the long-term CDF of ΔVth. Its applicability to predicting the long-term $\Delta I_{D,LIN}/I_{D,LIN}$ measured at Vg>Vth is investigated here. Two statistical distributions have been proposed for the trap's time constants by early works: Log-uniform or Log-normal [3], [21], [29], [61]. Their applicability to the kinetics of N_A and N_D will be tested. In addition, power law has been widely used as the kinetics for ageing [109], [135] and will also be tested.

In Fig. 7.19, the data obtained within 10 sec are used to fit the kinetics. The extracted kinetics are then extrapolated to 6×10^4 sec and compared with the test data. Fig. 7.20 shows that, for

both N_A and N_D, the power law gives the highest errors, and the Log-uniform kinetics gives the lowest errors within this time window. If extrapolated further into the future, Log-uniform kinetics will not saturate, but Lognormal kinetics will. Further work is needed to determine which one is more accurate over a typical device lifetime (e.g., 10 years).

With the predicted N_A and N_D at 6×10^4 sec and the mean impact per trap given in Fig. 7.14, the CDF of $\Delta I_{D,LIN}/I_{D,LIN}$ is evaluated based on the Poisson distribution of the number of traps per device and log-normal CDF of the impact per trap. The predicted CDF of $\Delta I_{D,LIN}/I_{D,LIN}$ is compared with test data in Fig. 7.21 and good agreement is obtained for the Log-normal CDF per trap. Hence, the integral methodology can be applied to predict long-term $\Delta I_{D,LIN}/I_{D,LIN}$. The two distributions selected for $\Delta I_{D,LIN}/I_{D,LIN}$ (Vg = 0.9V and Vd = 0.1V) are,

1. Impact per trap: Lognormal Distribution
2. Long-Term RTN Kinetic: Log-uniform Distribution.

7.4 Conclusions

The key parameters required for circuit simulation include driving current and threshold voltage under both linear and saturation modes. For the first time, this work used the integral methodology to assess their RTN-induced statistical distributions per trap based on test data.

Unlike early works that overestimated ΔVth by using $\Delta I_D/g_m$, ΔVth is directly measured from the pulse I_D-Vg at Vg=Vth in this study. It is shown that $\Delta I_D/I_D$ and ΔVth follow different CDF: Log-normal best describes $\Delta I_D/I_D$, while GEV is the best for ΔVth, under both linear and saturation conditions.

Despite the reduced field between gate and conduction channel under saturation, both $\Delta I_{D,SAT}/I_{D,SAT}$ and $\Delta Vth_{,SAT}$ are actually higher than their linear counterparts, indicating hot carriers assisting trap-filling. The results show that the RTN in each of the key parameters should be modelled separately. The work also shows that the integral methodology can be used to predict the long-term $\Delta I_{D,LIN}/I_{D,LIN}$, based on the CDF extracted from a short time window and Log-uniform kinetics.

CHAPTER 8: CONCLUSIONS AND FUTURE WORK FOR RTN MODELING

The main goal of this research work is to develop new RTN modelling techniques with a verified capability for long-term prediction and to improve the existing understanding of RTN. The new model developed is generically applicable for evaluating the probability of a certain ΔVth or $\Delta I_D/I_D$ occurring in devices in circuits. Prediction is essential as commercial devices are expected to operate for 10 years. This chapter summarizes the achievement of this project. Potential future work will be discussed in Section 8.2

8.1 Contributions to RTN modelling

A key limiting factor for industry is the lack of an RTN model that generates the probability of a given $\Delta I_D/I_D$ or ΔVth occurring. In Chapter 4, a new integral methodology was proposed which significantly advances RTN modelling. Early works applied a bottom-up approach to characterize the properties of individual traps, such as time constants, amplitude, and the number of traps per device. Such an approach has been shown to have major limitations. Based on the integral method proposed here, the burden of extracting the properties of individual traps is removed by:

i. Characterizing the combined impact of all traps on a device,
ii. Integrating the RTN of multiple devices into one dataset for statistical analysis.

Although the ΔVth of one device can fluctuate, the collective dataset gives a stable statistical distribution at a given time point when the number of devices is sufficiently high. This stable distribution can be assumed to originate from a set of effective charged traps (ECTs), which are always charged. For a longer time window, the number of ECTs will be higher. The dynamic charging and discharging of traps are transformed into modelling the kinetics of ECT and the need to characterize the capture and emission time of individual traps is removed. Based on Maximum Likelihood Estimation (MLE), the following key information can be extracted from the test data:

i. the statistical distribution of impact per trap,
ii. the time-dependent kinetics of ECTs,
iii. the probability of ΔVth hitting a certain mV.

Early works typically validate a model by showing that it can fit test data well. A good fitting, however, does not warrant that the model can predict future RTN. There is little information available on whether the existing RTN models can predict future RTN. In this project, long-term RTN measurement has been predicted up to $6x10^4$ sec and the capability to predict RTN based on the model extracted from short-time RTN data is verified. The results from DC measurement under Vg and Vd of 0.5 V and 0.1 V respectively show that the lowest error is obtained for log-uniform kinetics with GEV δVth distribution.

RTN fluctuations can be either positive or negative, which are referred to in this project as acceptor-like and donor-like trap, respectively. Early works mainly focused on ΔVth or $\Delta I_D/I_D > 0$. Based on the RTN measurements, however, ΔVth can be either > 0 or < 0. It was found that there is little correlation between $\Delta Vth > 0$ and $\Delta Vth < 0$. One advance made by this project is to model RTN in both positive and negative directions simultaneously, as detailed in Chapter 5.

Another major advance of this work is in AC RTN modeling. As is widely known, commercial devices mostly operate under a kind of AC waveform, instead of DC waveform. A detailed AC biased RTN measurement and analysis is performed to improve the knowledge in this area. It has been found that DC RTN properties can be different from AC RTN properties. The results of AC RTN short- and long-term measurement are given in Chapter 6. First, the frequency impact of applied Vg on individual traps is explored and the differences between acceptor-like and donor-like traps responses to frequency are investigated. The emission time of individual traps of both types is sharply reduced with an increase in frequency. However, the amplitude of single traps is insensitive to frequency. Our study found that the time-dependent kinetics and δVth per trap distribution are the same as those for DC RTN, i.e., log-uniform for time-dependent kinetics and GEV for δVth distribution. Good prediction by the model extracted from a short-time range of experimental data has been proven. These results demonstrate the applicability of the integral methodology to AC RTN analysis.

Next, the accurate extraction of ΔVth is important, as error in ΔVth can lead to severe over- or under-estimation of future ΔVth. The two most common ΔVth extraction methods are evaluated: one uses $\Delta Vth = \Delta I_D/g_m$ and the other measures ΔVth at Vg = Vth. It is found that $\Delta Vth = \Delta I_D/g_m$ substantially overestimates ΔVth, if ΔI_D is measured under a Vg substantially higher than Vth.

Lastly, in addition to experiments and analysis within the linear region of transistor operation, many real-world circuits have devices operating in the saturation region. One commonly used circuit where this is demonstrated is a 6T-SRAM circuit. To explore this important area of operation, measurement, and analysis of devices under the saturation condition of Vd = Vg = $0.9V$ was carried out. ΔVth and $\Delta I_D/I_D$ were evaluated to compare and explore the differences between linear and saturation modes of operation. It was found that the mean $I_{D,SAT}$ is a factor of 4.8 higher than $I_{D,LIN}$ and $\Delta I_{D,SAT}$ has a wider statistical spread along the x-axis. $\Delta I_D/I_D$ and ΔVth under linear and saturation were projected to multiple σ based on extracted impact per trap. The impact of linear and saturation operation on $\Delta I_D/I_D$ and ΔVth at different standard deviations was compared their differences were shown to be substantial at 5σ. It is concluded that linear and saturation mode measurements must be modelled separately, as the properties of RTN under saturation conditions cannot be inferred from the properties of RTN under linear conditions. Furthermore, based on the results presented in this work, the impact per trap distribution for $\Delta I_D/I_D$ is found to follow a Log-normal distribution, unlike ΔVth, which follows a GEV distribution. However, both $\Delta I_D/I_D$ and ΔVth kinetics follow a Log-uniform distribution.

8.2 Future Works

The new methodologies presented in this work have opened up many areas of potential further work for RTN and degradation modelling. One basic improvement that can be made is to increase the sampling rate of RTN datapoint collection. The current measurement setup uses 1 MSample/s for RTN current fluctuation, I_D, sampling for a 10-sec time window. This limits the fastest trap the measurement setup can detect. 1 µs per data point means that the fastest capture or emission event that can be detected is ~1 µs, but it is well known that the switching behaviour of fast traps can be less than 1 µs. However, the trade-off of measuring with a faster sampling rate is a large number of data points; having very large amounts of output data results in impractically slow analysis.

The accuracy of the RTN integral model can be further improved by applying RTN data from various unexplored measurement conditions. RTN parameters at different conditions, such as different bias voltages and temperatures, can be collected to find the relevant dependencies. These findings can then be included in the model to improve prediction accuracy across the bias and temperature space.

The capability of the integral model can be further tested, especially on arbitrary periodic workload data, which were not measured in this project. It can be done by performing RTN measurements under arbitrary workload, followed by modelling the collected data with the integral model, to find out if the integral model is applicable to the RTN under arbitrary periodic workloads.

Fully implementing a model of RTN impact into circuit simulation would be a major milestone, which the work presented here has made significant progress towards. A 6T SRAM circuit would be a prime example of a real circuit where this modelling could be demonstrated, since a 6T SRAM is formed by both nMOSFET and pMOSFET and the transistors can operate under both saturation and linear operation regions. Measurements of RTN should be taken following the integral methodology to extract the key parameters under saturation and linear operation, then applied to each of the transistors in 6T SRAM. The RTN impact can be explored in a few aspects, such as Static Read/Write Margin or Dynamic Read/Write Margin.

The ideal extension of this work would be to integrate the RTN integral methodology with both as-fabricated device-to-device variation (time-zero DDV) models and other time-dependent variation models. The time-zero DDV can originate from a number of sources, such as line-edge roughness and gate-work function variation. Time-dependent variation can originate from device ageing, including NBTI, PBTI, and HCA, etc. A fully integrated degradation model up to industrial standards could provide circuit designers with an estimation of different degradation mechanisms to evaluate designed circuits in terms of reliability and performance.

REFERENCES

[1]. 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, pp. 367-468, 1989.

[2]. A. Asenov, R. Balasubramaniam, A. R. Brown, and J. H. Davies, "RTS amplitudes in decananometer MOSFETs: 3-D simulation study," *IEEE Transactions on Electron Devices,* vol. 50, pp. 839-845, 2003.

[3]. R. Wang, S. Guo, Z. Zhang, Q. Wang, D. Wu, J. Wang, *et al.*, "Too noisy at the bottom?—Random telegraph noise (RTN) in advanced logic devices and circuits," in *2018 IEEE International Electron Devices Meeting (IEDM)*, 2018, pp. 17.2. 1-17.2. 4.

[4]. A. Manut, R. Gao, J. F. Zhang, Z. Ji, M. Mehedi, W. D. Zhang, *et al.*, "Trigger-when-charged: A technique for directly measuring RTN and BTI-induced threshold voltage fluctuation under use-Vdd," *IEEE Transactions on Electron Devices,* vol. 66, pp. 1482-1488, 2019.

[5]. F. M. Puglisi and P. Pavan, "Guidelines for a reliable analysis of random telegraph noise in electronic devices," *IEEE Transactions on Instrumentation and Measurement,* vol. 65, pp. 1435-1442, 2016.

[6]. M. B. da Silva, T. H. Both, and G. I. Wirth, "Random telegraph noise modeling for circuit analysis: RTN in ring oscillators," *IEEE Journal of the Electron Devices Society,* vol. 10, pp. 459-465, 2022.

[7]. J. F. Zhang, R. Gao, M. Duan, Z. Ji, W. Zhang, and J. Marsland, "Bias temperature instability of mosfets: Physical processes, models, and prediction," *Electronics,* vol. 11, p. 1420, 2022.

[8]. J. Fang and S. S. Sapatnekar, "Incorporating hot-carrier injection effects into timing analysis for large circuits," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* vol. 22, pp. 2738-2751, 2014.

[9]. C.-C. Chen, T. Liu, and L. Milor, "System-level modeling of microprocessor reliability degradation due to bias temperature instability and hot carrier injection," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems,* vol. 24, pp. 2712-2725, 2016.

[10]. M. Duan, J. Zhang, J. Zhang, W. Zhang, Z. Ji, B. Benbakhti, *et al.*, "Interaction between hot carrier aging and PBTI degradation in nMOSFETs: Characterization, modelling and lifetime prediction," in *2017 IEEE International Reliability Physics Symposium (IRPS)*, 2017, pp. XT-5.1-XT-5.7.

[11]. J. F. Zhang, "Traps," *Wiley Encyclopedia of Electrical and Electronics Engineering,* pp. 1-10, 1999.

[12]. A. M. Islam and H. Onodera, "Worst-case performance analysis under random telegraph noise induced threshold voltage variability," in *2018 28th International Symposium on Power and Timing Modeling, Optimization and Simulation (PATMOS)*, 2018, pp. 140-146.

[13]. K. K. Hung, P. K. Ko, C. Hu, and Y. C. Cheng, "A unified model for the flicker noise in metal-oxide-semiconductor field-effect transistors," *IEEE Transactions on Electron Devices*, vol. 37, pp. 654-665, 1990.

[14]. C. Y.-P. Chao, H. Tu, T. M.-H. Wu, K.-Y. Chou, S.-F. Yeh, C. Yin, *et al.*, "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.

[15]. 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

[16]. 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.

[17]. 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.

[18]. G. I. Wirth, R. da Silva, and B. Kaczer, "Statistical model for MOSFET bias temperature instability component due to charge trapping," *IEEE Transactions on Electron Devices*, vol. 58, pp. 2743-2751, 2011.

[19]. B. Kaczer, T. Grasser, P. J. Roussel, J. Franco, R. Degraeve, L.-A. Ragnarsson, *et al.*, "Origin of NBTI variability in deeply scaled pFETs," in *2010 IEEE International Reliability Physics Symposium*, 2010, pp. 26-32.

[20]. Q. Tang and C. H. Kim, "Characterizing the impact of RTN on logic and SRAM operation using a dual ring oscillator array circuit," *IEEE Journal of Solid-State Circuits*, vol. 52, pp. 1655-1663, 2017.

[21]. M. Mehedi, K. H. Tok, J. F. Zhang, Z. Ji, Z. Ye, W. Zhang, *et al.*, "An assessment of the statistical distribution of Random Telegraph Noise Time Constants," *IEEE Access*, vol. 8, pp. 182273-182282, 2020.

[22]. S. Realov and K. L. Shepard, "Random telegraph noise in 45-nm CMOS: Analysis using an on-chip test and measurement system," in *2010 International Electron Devices Meeting*, 2010, pp. 28.2. 1-28.2. 4.

[23]. N. Tega, H. Miki, M. Yamaoka, H. Kume, T. Mine, T. Ishida, *et al.*, "Impact of threshold voltage fluctuation due to random telegraph noise on scaled-down SRAM," in *2008 IEEE International Reliability Physics Symposium*, 2008, pp. 541-546.

[24]. H. Miki, M. Yamaoka, N. Tega, Z. Ren, M. Kobayashi, C. D'Emic, *et al.*, "Understanding short-term BTI behavior through comprehensive observation of gate-voltage dependence of RTN in highly scaled high-κ/metal-gate pFETs," in *2011 Symposium on VLSI Technology-Digest of Technical Papers*, 2011, pp. 148-149.

[25]. N. Tega, H. Miki, T. Osabe, A. Kotabe, K. Otsuga, H. Kurata, *et al.*, "Anomalously Large Threshold Voltage Fluctuation by Complex Random Telegraph Signal in Floating Gate Flash Memory," in *2006 International Electron Devices Meeting*, 2006, pp. 1-4.

[26]. H. Qiu, K. Takeuchi, T. Mizutani, T. Saraya, J. Chen, M. Kobayashi, *et al.*, "Statistical analyses of random telegraph noise amplitude in ultra-narrow (deep sub-10nm) silicon nanowire transistors," in *2017 Symposium on VLSI Technology*, 2017, pp. T50-T51.

[27]. T. Nagumo, K. Takeuchi, T. Hase, and Y. Hayashi, "Statistical characterization of trap position, energy, amplitude and time constants by RTN measurement of multiple individual traps," in *2010 International Electron Devices Meeting*, 2010, pp. 28.3. 1-28.3. 4.

[28]. N. Tega, H. Miki, F. Pagette, D. J. Frank, A. Ray, M. J. Rooks, *et al.*, "Increasing threshold voltage variation due to random telegraph noise in FETs as gate lengths scale to 20 nm," in *2009 Symposium on VLSI Technology*, 2009, pp. 50-51.

[29]. K. Ito, T. Matsumoto, S. Nishizawa, H. Sunagawa, K. Kobayashi, and H. Onodera, "Modeling of random telegraph noise under circuit operation—Simulation and measurement of RTN-induced delay fluctuation," in *2011 12th International Symposium on Quality Electronic Design*, 2011, pp. 1-6.

[30]. M. Mehedi, K. H. Tok, Z. Ye, J. F. Zhang, Z. Ji, W. Zhang, *et al.*, "On the accuracy in modeling the statistical distribution of random telegraph noise amplitude," *IEEE Access*, vol. 9, pp. 43551-43561, 2021.

[31]. K. i. Sonoda, K. Ishikawa, T. Eimori, and O. Tsuchiya, "Discrete dopant effects on statistical variation of random telegraph signal magnitude," *IEEE Transactions on Electron Devices*, vol. 54, pp. 1918-1925, 2007.

[32]. 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.

[33]. 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.

[34]. K. Fukuda, Y. Shimizu, K. Amemiya, M. Kamoshida, and C. Hu, "Random telegraph noise in flash memories-model and technology scaling," in *2007 IEEE International Electron Devices Meeting*, 2007, pp. 169-172.

[35]. A. Ghetti, C. M. Compagnoni, A. S. Spinelli, and A. Visconti, "Comprehensive analysis of random telegraph noise instability and its scaling in deca–nanometer flash memories," *IEEE Transactions on Electron Devices*, vol. 56, pp. 1746-1752, 2009.

[36]. J. P. Campbell, J. Qin, K. Cheungl, L. Yu, J. Suehlel, A. Oates, *et al.*, "The origins of random telegraph noise in highly scaled SiON nMOSFETs," in *2008 IEEE International Integrated Reliability Workshop Final Report*, 2008, pp. 105-109.

[37]. X. Chen, Y. Wang, Y. Cao, and H. Yang, "Statistical analysis of random telegraph noise in digital circuits," in *2014 19th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2014, pp. 161-166.

[38]. K. Kandiah and F. Whiting, "Low frequency noise in junction field effect transistors," *Solid-State Electronics*, vol. 21, pp. 1079-1088, 1978.

[39]. C. Claeys, M. De Andrade, Z. Chai, W. Fang, B. Govoreanu, B. Kaczer, *et al.*, "Random telegraph signal noise in advanced high performance and memory devices," in *2016 31st Symposium on Microelectronics Technology and Devices (SBMicro)*, 2016, pp. 1-6

[40]. N. Ashraf and D. Vasileska, "1/f Noise: threshold voltage and ON-current fluctuations in 45 nm device technology due to charged random traps," *Journal of computational electronics*, vol. 9, pp. 128-134, 2010.

[41]. G. Wirth, "Time-dependent random threshold voltage variation due to random telegraph noise," *IEEE Transactions on Electron Devices*, vol. 68, pp. 17-23, 2020.

[42]. A. Ghetti, C. M. Compagnoni, F. Biancardi, A. L. Lacaita, S. Beltrami, L. Chiavarone, *et al.*, "Scaling trends for random telegraph noise in deca-nanometer Flash memories," in *2008 IEEE International Electron Devices Meeting*, 2008, pp. 1-4.

[43]. K. P. Cheung and J. P. Campbell, "On the magnitude of random telegraph noise in ultra-scaled MOSFETs," in *2011 IEEE International Conference on IC Design & Technology*, 2011, pp. 1-4.

[44]. P. Ren, P. Hao, C. Liu, R. Wang, X. Jiang, Y. Qiu, *et al.*, "New observations on complex RTN in scaled high-κ/metal-gate MOSFETs—The role of defect coupling under DC/AC condition," in *2013 IEEE International Electron Devices Meeting*, 2013, pp. 31.4. 1-31.4. 4.

[45]. K. Ohmori, W. Feng, S. Sato, R. Hettiarachchi, M. Sato, T. Matsuki, *et al.*, "Direct real-time observation of channel potential fluctuation correlated to random telegraph noise of drain current using nanowire MOSFETs with four-probe terminals," in *2011 Symposium on VLSI Technology-Digest of Technical Papers*, 2011, pp. 202-203.

[46]. Z. Zhang, S. Guo, X. Jiang, R. Wang, R. Huang, and J. Zou, "Investigation on the amplitude distribution of random telegraph noise (RTN) in nanoscale MOS devices," in *2016 IEEE International Nanoelectronics Conference (INEC)*, 2016, pp. 1-2.

[47]. M. Yamaoka, H. Miki, A. Bansal, S. Wu, D. Frank, E. Leobandung, *et al.*, "Evaluation methodology for random telegraph noise effects in SRAM arrays," in *2011 International Electron Devices Meeting*, 2011, pp. 32.2. 1-32.2. 4.

[48]. K. Takeuchi, T. Nagumo, K. Takeda, S. Asayama, S. Yokogawa, K. Imai, *et al.*, "Direct observation of RTN-induced SRAM failure by accelerated testing and its application to product reliability assessment," in *2010 Symposium on VLSI Technology*, 2010, pp. 189-190.

[49]. M. Luo, R. Wang, S. Guo, J. Wang, J. Zou, and R. Huang, "Impacts of random telegraph noise (RTN) on digital circuits," *IEEE Transactions on Electron Devices*, vol. 62, pp. 1725-1732, 2014.

[50]. K. Takeuchi, T. Nagumo, and T. Hase, "Comprehensive SRAM design methodology for RTN reliability," in *2011 Symposium on VLSI Circuits-Digest of Technical Papers*, 2011, pp. 130-131.

[51]. M.-L. Fan, V. P.-H. Hu, Y.-N. Chen, P. Su, and C.-T. Chuang, "Analysis of single-trap-induced random telegraph noise on FinFET devices, 6T SRAM cell, and logic circuits," *IEEE Transactions on Electron Devices*, vol. 59, pp. 2227-2234, 2012.

[52]. L. Gerrer, J. Ding, S. M. Amoroso, F. Adamu-Lema, R. Hussin, D. Reid, *et al.*, "Modelling RTN and BTI in nanoscale MOSFETs from device to circuit: A review," *Microelectronics Reliability*, vol. 54, pp. 682-697, 2014.

[53]. M.-L. Fan, S.-Y. Yang, V. P.-H. Hu, Y.-N. Chen, P. Su, and C.-T. Chuang, "Single-trap-induced random telegraph noise for FinFET, Si/Ge Nanowire FET, Tunnel FET, SRAM and logic circuits," *Microelectronics Reliability*, vol. 54, pp. 698-711, 2014.

[54]. K. Zhang, *Embedded memories for nano-scale VLSIs* vol. 2: Springer, 2009.

[55]. P. Machanick, "The case for SRAM main memory," *ACM SIGARCH Computer Architecture News*, vol. 24, pp. 23-30, 1996.

[56]. ITRS 2009 web site: https://www.semiconductors.org/resources/2009-international-technology-roadmap-for-semiconductors-itrs/

[57]. N. V. Amarasinghe and Z. Çelik-Butler, "Complex random telegraph signals in 0.06 μm2 MDD n-MOSFETs," *Solid-State Electronics*, vol. 44, pp. 1013-1019, 2000.

[58]. C. M. Compagnoni, R. Gusmeroli, A. S. Spinelli, and A. Visconti, "RTN VT Instability From the Stationary Trap-Filling Condition: An Analytical Spectroscopic Investigation," *IEEE transactions on electron devices*, vol. 55, pp. 655-661, 2008.

[59]. R. E. Howard, W. Skocpol, L. D. Jackel, P. M. Mankiewich, L. A. Fetter, D. Tennant, *et al.*, "Single electron switching events in nanometer-scale Si MOSFET's," *IEEE transactions on electron devices*, vol. 32, pp. 1669-1674, 1985.

[60]. K. Kobayashi, M. Islam, T. Matsumoto, and R. Kishida, "Random Telegraph Noise Under Switching Operation," *Noise in Nanoscale Semiconductor Devices*, pp. 285-333, 2020.

[61]. M. Tanizawa, S. Ohbayashi, T. Okagaki, K. Sonoda, K. Eikyu, Y. Hirano, *et al.*, "Application of a statistical compact model for random telegraph noise to scaled-SRAM Vmin analysis," in *2010 Symposium on VLSI Technology*, 2010, pp. 95-96.

[62]. 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 *2014 IEEE International Reliability Physics Symposium*, 2014, pp. XT. 17.1-XT. 17.5.

[63]. T. Nagumo, K. Takeuchi, S. Yokogawa, K. Imai, and Y. Hayashi, "New analysis methods for comprehensive understanding of random telegraph noise," in *2009 IEEE International Electron Devices Meeting (IEDM)*, 2009, pp. 1-4.

[64]. T. Grasser, "Stochastic charge trapping in oxides: From random telegraph noise to bias temperature instabilities," *Microelectronics Reliability*, vol. 52, pp. 39-70, 2012.

[65]. T. H. Both, G. F. Furtado, and G. I. Wirth, "Modeling and simulation of the charge trapping component of BTI and RTS," *Microelectronics Reliability*, vol. 80, pp. 278-283, 2018.

[66]. Z. Zhang, S. Guo, X. Jiang, R. Wang, Z. Zhang, P. Hao, *et al.*, "New insights into the amplitude of random telegraph noise in nanoscale MOS devices," in *2017 IEEE International Reliability Physics Symposium (IRPS)*, 2017, pp. 3C-3.1-3C-3.5.

[67]. S. Ichino, T. Mawaki, A. Teramoto, R. Kuroda, S. Wakashima, T. Suwa, *et al.*, "Statistical analyses of random telegraph noise in pixel source follower with various gate shapes in CMOS image sensor," *ITE Transactions on Media Technology and Applications*, vol. 6, pp. 163-170, 2018.

[68]. M. Toledano-Luque, B. Kaczer, E. Simoen, P. J. Roussel, A. Veloso, T. Grasser, *et al.*, "Temperature and voltage dependences of the capture and emission times of individual traps in high-k dielectrics," *Microelectronic Engineering*, vol. 88, pp. 1243-1246, 2011.

[69]. K. Abe, A. Teramoto, S. Sugawa, and T. Ohmi, "Understanding of traps causing random telegraph noise based on experimentally extracted time constants and amplitude," in *2011 International Reliability Physics Symposium*, 2011, pp. 4A. 4.1-4A. 4.6.

[70]. D. Veksler, G. Bersuker, L. Vandelli, A. Padovani, L. Larcher, A. Muraviev, *et al.*, "Random telegraph noise (RTN) in scaled RRAM devices," in *2013 IEEE International Reliability Physics Symposium (IRPS)*, 2013, pp. MY. 10.1-MY. 10.4.

[71]. D. Veksler, G. Bersuker, B. Chakrabarti, E. Vogel, S. Deora, K. Matthews, *et al.*, "Methodology for the statistical evaluation of the effect of random telegraph noise (RTN) on RRAM characteristics," in *2012 International Electron Devices Meeting*, 2012, pp. 9.6. 1-9.6. 4.

[72]. N. Tega, H. Miki, Z. Ren, C. D'Emic, Y. Zhu, D. Frank, *et al.*, "Reduction of random telegraph noise in high-κ/metal-gate stacks for 22 nm generation FETs," in *2009 IEEE International Electron Devices Meeting (IEDM)*, 2009, pp. 1-4.

[73]. F. M. Puglisi, A. Padovani, L. Larcher, and P. Pavan, "Random telegraph noise: Measurement, data analysis, and interpretation," in *2017 IEEE 24th International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, 2017, pp. 1-9.

[74]. M. Maestro, J. Diaz, A. Crespo-Yepes, M. Gonzalez, J. Martin-Martinez, R. Rodriguez, *et al.*, "New high resolution Random Telegraph Noise (RTN) characterization method for resistive RAM," *Solid-State Electronics*, vol. 115, pp. 140-145, 2016.

[75]. S. Dongaonkar, M. Giles, A. Kornfeld, B. Grossnickle, and J. Yoon, "Random telegraph noise (RTN) in 14nm logic technology: High volume data extraction and analysis," in *2016 IEEE Symposium on VLSI Technology*, 2016, pp. 1-2.

[76]. C. Marquez, N. Rodriguez, F. Gamiz, and A. Ohata, "Systematic method for electrical characterization of random telegraph noise in MOSFETs," *Solid-State Electronics*, vol. 128, pp. 115-120, 2017.

[77]. W. Fang, E. Simoen, M. Aoulaiche, J. Luo, C. Zhao, and C. Claeys, "Distinction between silicon and oxide traps using single-trap spectroscopy," *physica status solidi (a)*, vol. 212, pp. 512-517, 2015.

[78]. X. Zhan, C. Shen, Z. Ji, J. Chen, H. Fang, F. Guo, *et al.*, "A Dual-Point Technique for the Entire I D–V G Characterization Into Subthreshold Region Under Random Telegraph Noise Condition," *IEEE Electron Device Letters*, vol. 40, pp. 674-677, 2019.

[79]. S. Realov and K. L. Shepard, "Analysis of random telegraph noise in 45-nm CMOS using on-chip characterization system," *IEEE Transactions on Electron Devices*, vol. 60, pp. 1716-1722, 2013.

[80]. G. Pedreira, J. Martín-Martínez, J. Diaz-Fortuny, P. Saraza-Canflanca, R. Rodríguez, R. Castro-López, *et al.*, "A new time efficient methodology for the massive characterization of RTN in CMOS devices," in *2019 IEEE International Reliability Physics Symposium (IRPS)*, 2019, pp. 1-5.

[81]. J. Martin-Martinez, J. Diaz, R. Rodriguez, M. Nafria, and X. Aymerich, "New weighted time lag method for the analysis of random telegraph signals," *IEEE Electron Device Letters*, vol. 35, pp. 479-481, 2014.

[82]. T. Grasser, *Noise in nanoscale semiconductor devices*: Springer Nature, 2020.

[83]. 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*, vol. 613, pp. 2-5, 2016.

[84]. F. M. Puglisi and P. Pavan, "RTN analysis with FHMM as a tool for multi-trap characterization in HfO x RRAM," in *2013 IEEE International Conference of Electron Devices and Solid-State Circuits*, 2013, pp. 1-2.

[85]. H. Miki, N. Tega, M. Yamaoka, D. Frank, A. Bansal, M. Kobayashi, *et al.*, "Statistical measurement of random telegraph noise and its impact in scaled-down high-κ/metal-gate MOSFETs," in *2012 International Electron Devices Meeting*, 2012, pp. 19.1. 1-19.1. 4.

[86]. H. Awano, H. Tsutsui, H. Ochi, and T. Sato, "Multi-trap RTN parameter extraction based on Bayesian inference," in *International Symposium on Quality Electronic Design (ISQED)*, 2013, pp. 597-602.

[87]. R. Wang, S. Guo, Z. Zhang, J. Zou, D. Mao, and R. Huang, "Complex random telegraph noise (RTN): What do we understand?," in *2018 IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, 2018, pp. 1-7.

[88]. F. M. Puglisi and P. Pavan, "Factorial hidden Markov model analysis of random telegraph noise in resistive random access memories," *ECTI Transactions on Electrical Engineering, Electronics, and Communications*, vol. 12, pp. 24-29, 2014.

[89]. F. M. Puglisi, L. Larcher, A. Padovani, and P. Pavan, "A complete statistical investigation of RTN in HfO 2-based RRAM in high resistive state," *IEEE Transactions on Electron Devices*, vol. 62, pp. 2606-2613, 2015.

[90]. T. Gong, Y. Wang, H. Yu, Y. Xu, P. Jiang, P. Yuan, *et al.*, "Investigation of endurance behavior on HfZrO-based charge-trapping FinFET devices by random telegraph noise and subthreshold swing techniques," *IEEE Transactions on Electron Devices*, vol. 68, pp. 3716-3719, 2021.

[91]. F. M. Puglisi, A. Padovani, P. Pavan, and L. Larcher, "Advanced modeling and characterization techniques for innovative memory devices: The RRAM case," in *Advances in Non-Volatile Memory and Storage Technology*, ed: Elsevier, 2019, pp. 103-135.

[92]. M. Mehedi, *Characterisation and modelling of Random Telegraph Noise in nanometre devices*: Liverpool John Moores University (United Kingdom), 2022.

[93]. S. Chitrashekaraiah, S. Guo, R. Herberholz, D. Vigar, and M. Redford, "Addressable test structures for MOSFET variability analysis," in *2012 IEEE International Conference on Microelectronic Test Structures*, 2012, pp. 31-35.

[94]. Datasheet: “Keysight N6780 Series Source/Measure Units (SMUs) for the N6700 Modular Power System”.

[95]. R. Gao, “Bias Temperature Instability Modelling and Lifetime Prediction on Nano-scale MOSFETs”, PhD thesis, Liverpool John Moore’s University, July 2018.

[96]. A. 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

[97]. G. Ghibaudo, “New method for the extraction of MOSFET parameter,” Electronic Letters, vol.24, pp. 543-545, 1988.

[98]. M. Tsuno, M. Suga, M. Tanaka, K. Shibahara, M. Miura-Mattausch, and M. Hirose, "Physically-based threshold voltage determination for MOSFET's of all gate lengths," *IEEE transactions on electron devices,* vol. 46, pp. 1429-1434, 1999.

[99]. P. Weckx, B. Kaczer, C. Chen, J. Franco, E. Bury, K. Chanda, *et al.*, "Characterization of time-dependent variability using 32k transistor arrays in an advanced HK/MG technology," in *2015 IEEE International Reliability Physics Symposium*, 2015, pp. 3B. 1.1-3B. 1.6.

[100]. F. Edition, A. Papoulis, and S. U. Pillai, *Probability, random variables, and stochastic processes*: McGraw-Hill Europe: New York, NY, USA, 2002.

[101]. G. C. Chow, "Maximum-likelihood estimation of misspecified models," *Economic Modelling*, vol. 1, pp. 134-138, 1984.

[102]. M. Toledano-Luque, B. Kaczer, J. Franco, P. J. Roussel, T. Grasser, T. Y. Hoffmann, *et al.*, "From mean values to distributions of BTI lifetime of deeply scaled FETs through atomistic understanding of the degradation," in *2011 Symposium on VLSI Technology-Digest of Technical Papers*, 2011, pp. 152-153.

[103]. J. Heyes, ‘‘Normal probability plots,’’ ASQC Statist. Division Newslett., vol. 13, no. 1, pp. 5–6, 1992.

[104]. N. T. Thomopoulos and N. T. Thomopoulos, "Exponential," *Statistical Distributions: Applications and Parameter Estimates*, pp. 21-29, 2017.

[105]. H. Quan and J. Zhang, "Estimate of standard deviation for a log-transformed variable using arithmetic means and standard deviations," *Statistics in medicine*, vol. 22, pp. 2723-2736, 2003.

[106]. J. R. M. Hosking, J. R. Wallis, and E. F. Wood, "Estimation of the Generalized Extreme-Value Distribution by the Method of Probability-Weighted Moments," *Technometrics*, vol. 27, pp. 251-261, 1985.

[107]. A. J. Scholten, L. F. Tiemeijer, R. Van Langevelde, R. J. Havens, A. T. Zegers-van Duijnhoven, and V. C. Venezia, "Noise modeling for RF CMOS circuit simulation," *IEEE Transactions on Electron Devices*, vol. 50, pp. 618-632, 2003.

[108]. M. Nour, Z. Çelik-Butler, A. Sonnet, F.-C. Hou, S. Tang, and G. Mathur, "A stand-alone, physics-based, measurement-driven model and simulation tool for random telegraph signals originating from experimentally identified MOS gate-oxide defects," *IEEE Transactions on Electron Devices*, vol. 63, pp. 1428-1436, 2016.

[109]. J. Zhang, Z. Ji, and W. Zhang, "As-grown-generation (AG) model of NBTI: A shift from fitting test data to prediction," *Microelectronics Reliability*, vol. 80, pp. 109-123, 2018.

[110]. H. Yang, M. Robitaille, X. Chen, H. Elgabra, L. Wei, and N. Y. Kim, "Random telegraph noise of a 28-nm cryogenic MOSFET in the coulomb blockade regime," *IEEE Electron Device Letters*, vol. 43, pp. 5-8, 2021.

[111]. Y. Ye, C.-C. Wang, and Y. Cao, "Simulation of random telegraph noise with 2-stage equivalent circuit," in *2010 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2010, pp. 709-713.

[112]. P. Weckx, M. Simicic, K. Nomoto, M. Ono, B. Parvais, B. Kaczer, *et al.*, "Defect-based compact modeling for RTN and BTI variability," in *2017 IEEE International Reliability Physics Symposium (IRPS)*, 2017, pp. CR-7.1-CR-7.6.

[113]. R. Gao, M. Mehedi, H. Chen, X. Wang, J. Zhang, X. Lin, *et al.*, "A fast and test-proven methodology of assessing RTN/fluctuation on deeply scaled nano pMOSFETs," in *2020 IEEE International Reliability Physics Symposium (IRPS)*, 2020, pp. 1-5.

[114]. M. Duan, J. Zhang, Z. Ji, W. Zhang, B. Kaczer, T. Schram, *et al.*, "Time-dependent variation: A new defect-based prediction methodology," in *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*, 2014, pp. 1-2.

[115]. P. Saraza-Canflanca, J. Martin-Martinez, R. Castro-Lopez, E. Roca, R. Rodriguez, F. V. Fernandez, *et al.*, "Statistical characterization of time-dependent variability defects using the maximum current fluctuation," *IEEE Transactions on Electron Devices*, vol. 68, pp. 4039-4044, 2021.

[116]. M. Duan, J. F. Zhang, Z. Ji, W. D. Zhang, B. Kaczer, T. Schram, *et al.*, "New analysis method for time-dependent device-to-device variation accounting for within-device fluctuation," *IEEE transactions on electron devices*, vol. 60, pp. 2505-2511, 2013.

[117]. J. Zhang, S. Taylor, and W. Eccleston, "Electron trap generation in thermally grown SiO2 under Fowler–Nordheim stress," *Journal of applied physics*, vol. 71, pp. 725-734, 1992.

[118]. T. Grasser, K. Rott, H. Reisinger, M. Waltl, P. Wagner, F. Schanovsky, *et al.*, "Hydrogen-related volatile defects as the possible cause for the recoverable component of NBTI," in *2013 IEEE International Electron Devices Meeting*, 2013, pp. 15.5. 1-15.5. 4.

[119]. R. Gao, A. B. Manut, Z. Ji, J. Ma, M. Duan, J. F. Zhang, *et al.*, "Reliable time exponents for long term prediction of negative bias temperature instability by extrapolation," *IEEE Transactions on Electron Devices*, vol. 64, pp. 1467-1473, 2017.

[120]. T. Grasser, H. Reisinger, P.-J. Wagner, F. Schanovsky, W. Gös, and B. Kaczer, "The time dependent defect spectroscopy (TDDS) for the characterization of the bias temperature instability," in *2010 IEEE International Reliability Physics Symposium*, 2010, pp. 16-25.

[121]. J. Zou, R. Wang, N. Gong, R. Huang, X. Xu, J. Ou, *et al.*, "New insights into AC RTN in scaled high-κ/metal-gate MOSFETs under digital circuit operations," in *2012 Symposium on VLSI Technology (VLSIT)*, 2012, pp. 139-140.

[122]. J. Zou, R. Wang, M. Luo, R. Huang, N. Xu, P. Ren, *et al.*, "Deep understanding of AC RTN in MuGFETs through new characterization method and impacts on logic circuits," in *2013 Symposium on VLSI Technology*, 2013, pp. T186-T187.

[123]. M. Duan, J. Zhang, Z. Ji, J. Ma, W. Zhang, B. Kaczer, *et al.*, "Key issues and techniques for characterizing time-dependent device-to-device variation of SRAM," in *2013 IEEE International Electron Devices Meeting*, 2013, pp. 31.3. 1-31.3. 4.

[124]. M. Duan, J. F. Zhang, Z. Ji, W. D. Zhang, B. Kaczer, and A. Asenov, "Key issues and solutions for characterizing hot carrier aging of nanometer scale nMOSFETs," *IEEE Transactions on Electron Devices*, vol. 64, pp. 2478-2484, 2017.

[125]. T. Grasser, H. Reisinger, K. Rott, M. Toledano-Luque, and B. Kaczer, "On the microscopic origin of the frequency dependence of hole trapping in pMOSFETs," in *2012 International Electron Devices Meeting*, 2012, pp. 19.6. 1-19.6. 4.

[126]. T. Grasser, K. Rott, H. Reisinger, P.-J. Wagner, W. Gös, F. Schanovsky, *et al.*, "Advanced characterization of oxide traps: The dynamic time-dependent defect spectroscopy," in *2013 IEEE International Reliability Physics Symposium (IRPS)*, 2013, pp. 2D. 2.1-2D. 2.7.

[127]. A. Asenov, B. Cheng, D. Dideban, U. Kovac, N. Moezi, C. Millar, *et al.*, "Modeling and simulation of transistor and circuit variability and reliability," in *IEEE Custom Integrated Circuits Conference 2010*, 2010, pp. 1-8.

[128]. R. Akimoto, R. Kuroda, A. Teramoto, T. Mawaki, S. Ichino, T. Suwa, *et al.*, "Effect of drain-to-source voltage on random telegraph noise based on statistical analysis of MOSFETs with various gate shapes," in *2020 IEEE International Reliability Physics Symposium (IRPS)*, 2020, pp. 1-6.

[129]. C.-H. Pao, M.-L. Fan, M.-F. Tsai, Y.-N. Chen, V. P.-H. Hu, P. Su, *et al.*, "Impacts of random telegraph noise on the analog properties of FinFET and trigate devices and Widlar current source," in *2012 IEEE International Conference on IC Design & Technology*, 2012, pp. 1-4.

[130]. A. Yonezawa, A. Teramoto, R. Kuroda, H. Suzuki, S. Sugawa, and T. Ohmi, "Statistical analysis of random telegraph noise reduction effect by separating channel from the interface," in *2012 IEEE International Reliability Physics Symposium (IRPS)*, 2012, pp. 3B. 5.1-3B. 5.7.

[131]. K. Ota, R. Ichihara, M. Suzuki, M. Saitoh, and Y. Mitani, "Random telegraph noise after hot carrier injection in tri-gate nanowire transistor," in *2019 Electron Devices Technology and Manufacturing Conference (EDTM)*, 2019, pp. 169-171.

[132]. K. Takeuchi, T. Nagumo, S. Yokogawa, K. Imai, and Y. Hayashi, "Single-charge-based modeling of transistor characteristics fluctuations based on statistical measurement of RTN amplitude," in *2009 Symposium on VLSI Technology*, 2009, pp. 54-55.

[133]. 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 *2014 Symposium on VLSI Technology (VLSI-Technology): Digest of Technical Papers*, 2014, pp. 1-2.

[134]. Z. Ji, S. Hatta, J. Zhang, J. Ma, W. Zhang, N. Soin, *et al.*, "Negative bias temperature instability lifetime prediction: Problems and solutions," in *2013 IEEE International Electron Devices Meeting*, 2013, pp. 15.6. 1-15.6. 4.

[135]. R. Gao, Z. Ji, S. Hatta, J. Zhang, J. Franco, B. Kaczer, *et al.*, "Predictive As-grown-Generation (AG) model for BTI-induced device/circuit level variations in nanoscale technology nodes," in *2016 IEEE International Electron Devices Meeting (IEDM)*, 2016, pp. 31.4. 1-31.4. 4.

LIST OF PUBLICATIONS

[1]. **K. H. Tok**, M. Mehedi, J. F. Zhang, J. Brown, Z. Ye, Z. Ji, *et al.*, "An integral methodology for predicting long-term RTN," *IEEE Transactions on Electron Devices,* vol. 69, pp. 3869-3875, 2022.

[2]. **K. H. Tok**, J. F. Zhang, J. Brown, Z. Ye, Z. Ji, W. Zhang, *et al.*, "AC RTN: Testing, modeling, and prediction," *IEEE Transactions on Electron Devices,* vol. 69, pp. 5780-5786, 2022.

[3]. **K. H. Tok**, J. F. Zhang, J. Brown, Z. Ji, W. Zhang, and J. S. Marsland, "Characterizing and Modeling RTN Under Real Circuit Bias Conditions," *IEEE Transactions on Electron Devices,* 2023.

[4]. **K. H. Tok**, M. Mehedi, J. F. Zhang, Z. Ye, Z. Ji, W. Zhang, *et al.*, "Criteria for selecting statistical distribution for the Amplitude of Random Telegraph Noise," in *2022 IEEE 16th International Conference on Solid-State & Integrated Circuit Technology (ICSICT)*, 2022, pp. 1-4.

[5]. **K. H. Tok**, J. F. Zhang, J. Brown, Z. Ji, and W. Zhang, “Extracting statistical distributions of RTN originating from both acceptor-like and donor-like traps,” IEEE 15th International Conference on ASIC (ASICON), 2023.

[6]. M. Mehedi, **K. H. Tok**, J. F. Zhang, Z. Ji, Z. Ye, W. Zhang, *et al.*, "An assessment of the statistical distribution of Random Telegraph Noise Time Constants," *IEEE Access,* vol. 8, pp. 182273-182282, 2020.

[7]. M. Mehedi, **K. H. Tok**, Z. Ye, J. F. Zhang, Z. Ji, W. Zhang, *et al.*, "On the accuracy in modeling the statistical distribution of random telegraph noise amplitude," *IEEE Access,* vol. 9, pp. 43551-43561, 2021.

[8]. M. Mehedi, **K. H. Tok**, J. F. Zhang, Z. Ji, Z. Ye, W. Zhang, *et al.*, "An integrated method for extracting the statistical distribution of RTN time constants," in *2021 IEEE 14th International Conference on ASIC (ASICON)*, 2021, pp. 1-4.

[9]. J. F. Zhang, M. Duan, M. Mehedi, **K. H. Tok**, Z. Ye, Z. Ji, *et al.*, "Defect loss and its physical processes," in *2020 IEEE 15th International Conference on Solid-State & Integrated Circuit Technology (ICSICT)*, 2020, pp. 1-4.

APPENDIX:

Matlab Commands used in the methodology:

Commands	Function
input()	Request User Input
mle()	Maximum Likelihood Estimation: mle(data, pdf equation, start value, Lower Boundary, Upper Boundary)
integral()	integral(Equation, Lower Boundary, Upper Boundary, Absolute Tolerance, Relative Tolerance)
poissppdf()	Poisson Distribution
lognpdf / logncdf	Lognormal Distribution: PDF / CDF
gevpdf / gevcdf	GEV Distribution: PDF / CDF

Inputs

Variables	Function
verify	User's inputs will be saved in 'Verify' automatically and carried forward to distribution script.
custpdf / custcdf	Corresponding PDF and CDF file for probability distribution.
lower / upper	Lower and upper boundaries of MLE Fitting.
startval	Start value that user has inputted.
param	MLE estimation function with return of MLE fitted parameters.
parameters	Parameters for Probability Calculation (User Inputs)
output	Output: calculated CDF values

```
%% Load File Name
file = input('Please enter the data filename: ','s');
filename = file;
load(filename);
z=data;
%% Maximum nA nD and Thermal Noise
nAInput = input('Please enter the maximum number of Traps for Acceptor (nA): ');
MaxnA = nAInput;
nDInput = input('Please enter the maximum number of Traps for Donor (nD): ');
MaxnD = nDInput;
ThermalInput = input('Thermal Noise: With (1) | Without (0): ');
ThermalCheck = ThermalInput;
tic;
%%
save ('verify.mat','MaxnA','MaxnD','ThermalCheck');
if ThermalCheck == 1
    sig = 0.1;
else
    sig = 0;
end
```

Appendix A1 Matlab-Coding for the Inputs.

```
prompt = input('Please select distribution: Expo (1) | Lognormal (2)| GEV (3): ');
distribution = prompt;
keepRunning = true;
while (keepRunning)
    %expo
    if  distribution == 1
        custpdf = @eval_custpdfexpo;
        custcdf = @eval_custcdfexpo;
            prompt1 = input('Please select MLE Fit (1) | Probability Calculation (2): ');
            selection = prompt1;
            keepRunning1 = true;
            if selection == 1
                startvalinput = input('Please enter the start values in [mu(A) mu(D) N(A) N(D)]: ');
                    startval = startvalinput;
                    startval(1,length(startvalinput)+1) = sig;
                lower = [startval(1,1)-0.2 startval(1,2)-0.2 startval(1,3)-0.5 startval(1,4)-0.5 startval(1,5)-0.05];
                upper = [startval(1,1)+0.3 startval(1,2)+0.3 startval(1,3)+0.5 startval(1,4)+0.5 startval(1,5)+0.05];
                                param = mle(z,'pdf',custpdf,'cdf',custcdf,'start',startval,'Options',statset('MaxFunEvals',...
                                    10000,'MaxIter',10000),'LowerBound',lower,'UpperBound',upper);
                 output = custcdf(z,param(1,1),param(1,2),param(1,3),param(1,4),param(1,5));
                 keepRunning1 = false;
            elseif selection == 2
                 para = input('Please enter the parameters in bracket [mu(A) mu(D) N(A) N(D)]: ');
                    parameters = para;
                output = custcdf(z,parameters(1,1),parameters(1,2),parameters(1,3),parameters(1,4),sig);
                keepRunning1 = false;
            else
                selection = input('Please select MLE Fit (1) | Probability Calculation (2): ');
            end
        keepRunning = false;
```

Appendix A2 Matlab-Coding for the Distribution Selections.

Probability Calculation: Eaxponential | Lognormal | GEV

Variables	Function
derlang / derlang_CDF	Erlang Distribution (Convolution of Exponential Distribution)
donorS0 / donorL0	Difference of Exponential Distribution
ANOF	Numerical Integration Equation for Acceptor and Donors
A	Table for Numerical Equation
AA	Transform 'A' into a single array
fsum	Summation of array 'AA'
ADT	Convolution of ΔVth_{RTN} and $\Delta Vth_{Thermal}$
A'n'	Numerical Equation for Acceptor Traps (A1, A2…A12)
D'n'	Numerical Equation for Donor Traps (D1, D2 … D12)
A'n'D'n'	Numerical Equation for Acceptor & Donor Traps (A1D1, A1D2...A12D12)

The current version has the numerical equation up to 12 acceptor and 12 donor traps (A12D12), which can estimate up to the average number of traps, N_A & N_D of 4 traps. For a higher average number of traps, numerical equation for 'n' numbers of acceptor and donor traps must be added manually by user.

Derivation of Equation (4.4):

x and y are two independent random variables with probability distribution functions of pdf_X and pdf_Y, respectively. If z = x + y, to have a z for a given x requires y to be y = z - x. As the probability for x and y = z - x to occur are $pdf_x(x)$ and $pdf_y(z - x)$, respectively, the probability for them to occur simultaneously will be the product of $pdf_x(x)$ and $pdf_y(z - x)$. Since z requires their simultaneous occurrence, the probability for z to occur once, $pdf_1(z)$, will be,

$$pdf_1(z) = pdf_x(x)pdf_y(z - x)$$

We can have the same z for different x, so long as y = z - x. As a result, z can occur multiple times for different x and y combinations. If x can vary from -∞ to ∞, the summation of probability for z to occur in this range of x will be,

$$pdf(z) = \int_{-\infty}^{\infty} pdf_x(x) pdf_y(z-x) dx$$

By definition [100], this is the convolution of x and y.

Justification of Lower and Upper Boundary for Convolution

Acceptor:

$pdf(z) = \int_{-\infty}^{\infty} pdf_x(x) pdf_y(z-x) dx$

Acceptor + Donor:

$pdf(z) = \int_{-\infty}^{\infty} pdf_x(x) pdf_y(x-z) dx$

Exponential & Lognormal: $pdf_x(x) = 0, if\ x < 0$

Acceptor:

$$z - x > 0 \rightarrow z > x$$

$$pdf_x(x < 0) = 0$$

Let z = ΔVth and x = δVth.

Lower Boundary: 0, Upper Boundary = ΔVth.

Acceptor + Donor:

$$x - z > 0 \rightarrow x > z$$

$if\ \Delta Vth > 0$, Lower Boundary = ΔVth, Upper Boundary = ∞.

$$pdf_x(x < 0) = 0$$

$if\ \Delta Vth < 0$, Lower Boundary = 0, Upper Boundary = ∞.

GEV: $pdf_x(x) = 0, if\ x < t_A$

Acceptor:

$$z - x > t_a \rightarrow z - t_A > x$$

$$pdf_x(x < t_A) = 0$$

Lower Boundary = t_A, Upper Boundary = $\Delta Vth - t_A$.

Acceptor + Donor:

$$pdf_x(x < t_A) = 0$$

$if\ \Delta Vth < (t_A - t_D)$, Lower Boundary = t_A, Upper Boundary = ∞.

$$pdf_x[(x - z) < t_D] = 0\ x - z > t_D \rightarrow x > z + t_D$$

$if\ \Delta Vth > (t_A - t_D)$, Lower Boundary = $\Delta Vth + t_D$, Upper Boundary = ∞.

Tutorial:

MLE: 400 devices

```
Please enter the data filename: 400DevicesExample
Please enter the maximum number of Traps for Acceptor (nA): 7
Please enter the maximum number of Traps for Donor (nD): 7
Thermal Noise: With (1) | Without (0): 0
Please select distribution: Expo (1) | Lognormal (2)| GEV (3): 1
Please select MLE Fit (1) | Probability Calculation (2): 1
Please enter the start values in [mu(A) mu(D) N(A) N(D)]: [1 0.7 2.6 1.5]
```

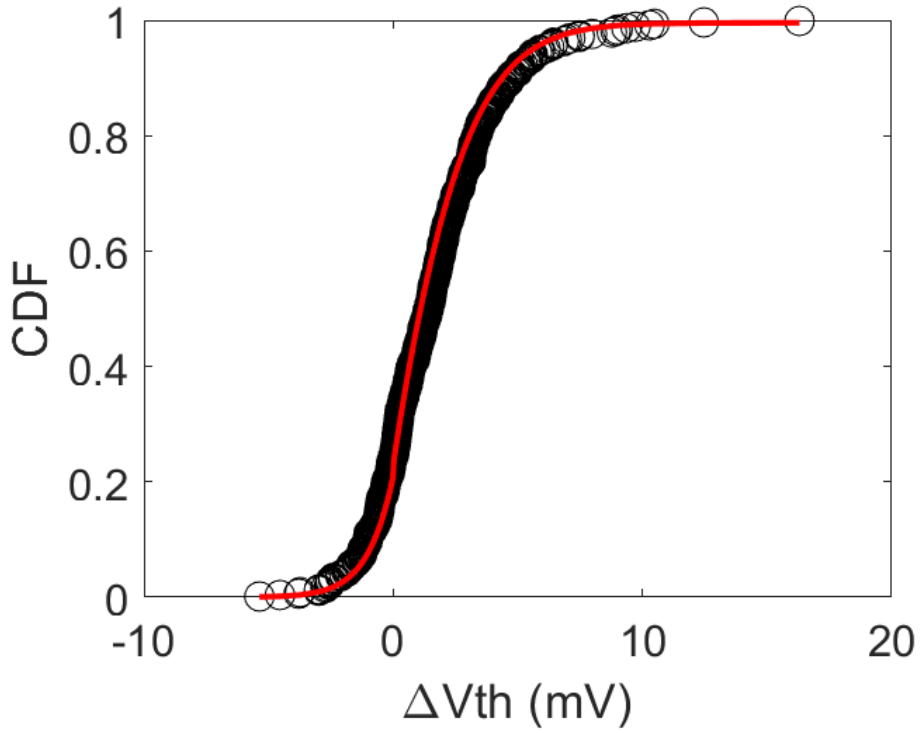
Appendix A3 400 Devices: MLE Inputs Example.

MLE Extracted Parameters: [mu(A) = 0.9, mu(D) = 0.6, N(A) = 2.5, N(D) =1.4]

Probability Calculation (Parameters from MLE Example Above)

```
Please enter the data filename: 400DevicesExample
Please enter the maximum number of Traps for Acceptor (nA): 7
Please enter the maximum number of Traps for Donor (nD): 7
Thermal Noise: With (1) | Without (0): 0
Please select distribution: Expo (1) | Lognormal (2)| GEV (3): 1
Please select MLE Fit (1) | Probability Calculation (2): 2
Please enter the parameters in bracket [mu(A) mu(D) N(A) N(D)]: [0.9 0.6 2.5 1.4]
Elapsed time is 22.371841 seconds.
```

Appendix A4 400 Devices: Probability Calculation Inputs Example.



Appendix A5 400 Devices: MLE and Probability Calculation Fitting.

MLE: 60 devices

```
Please enter the data filename: 60DevicesExample
Please enter the maximum number of Traps for Acceptor (nA): 4
Please enter the maximum number of Traps for Donor (nD): 4
Thermal Noise: With (1) | Without (0): 0
Please select distribution: Expo (1) | Lognormal (2)| GEV (3): 1
Please select MLE Fit (1) | Probability Calculation (2): 1
Please enter the start values in [mu(A) mu(D) N(A) N(D)]: [0.65 0.65 1 1.2]
Elapsed time is 455.086147 seconds.
```

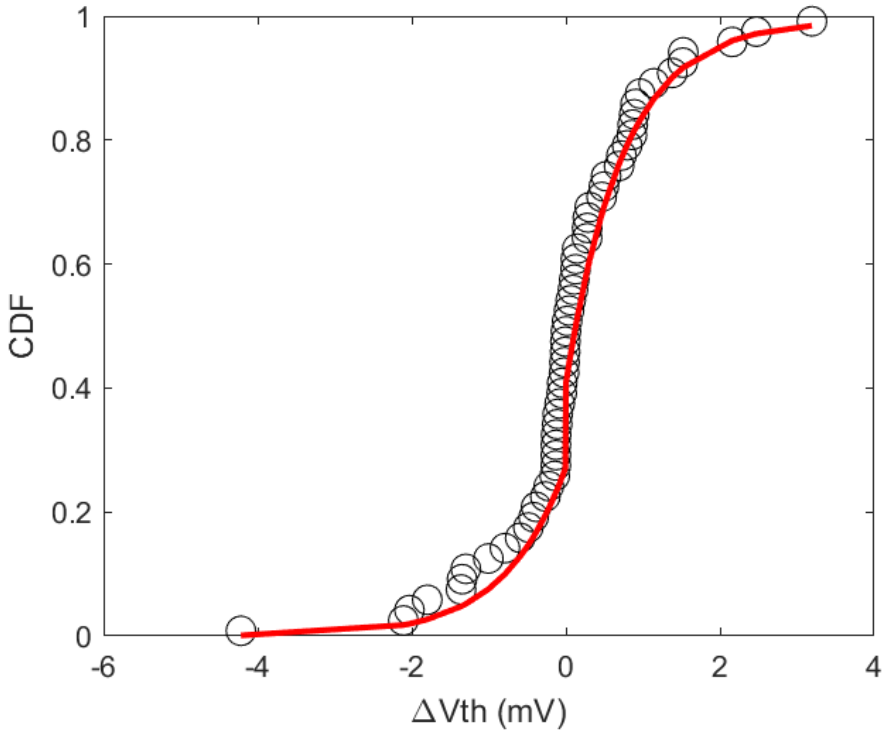
Appendix A6 60 Devices: MLE Inputs Example.

MLE Extracted Parameters: [mu(A) = 0.66, mu(D) = 0.55, N(A) = 0.91, N(D) = 1.1]

Probability Calculation (Parameters from MLE Example Above)

```
Please enter the maximum number of Traps for Acceptor (nA): 4
Please enter the maximum number of Traps for Donor (nD): 4
Thermal Noise: With (1) | Without (0): 0
Please select distribution: Expo (1) | Lognormal (2)| GEV (3): 1
Please select MLE Fit (1) | Probability Calculation (2): 2
Please enter the parameters in bracket [mu(A) mu(D) N(A) N(D)]: [0.66 0.55 0.91 1.1]
Elapsed time is 14.109437 seconds.
```

Appendix A7 60 Devices: Probability Calculation Inputs Example.



Appendix A8 60 Devices: MLE and Probability Calculation Fitting.