UNIVERSITY OF Southampton

University of Southampton Research Repository

Copyright © and Moral Rights for this thesis and, where applicable, any accompanying data are retained by the author and/or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This thesis and the accompanying data cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content of the thesis and accompanying research data (where applicable) must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holder/s.

When referring to this thesis and any accompanying data, full bibliographic details must be given, e.g.

Thesis: Author (Year of Submission) "Full thesis title", University of Southampton, name of the University Faculty or School or Department, PhD Thesis, pagination.

Data: Author (Year) Title. URI [dataset]

University of Southampton

Faculty of Physical Science and Engineering

School of Electronics and Computer Science

Manipulation of the Random telegraph signals for quantum random number generation

By

Fayong Liu

Thesis for the degree of Doctor of Philosophy

September 2019

University of Southampton

Abstract

Faculty of Physical Science and Engineering

School of Electronics and Computer Science

Thesis for the degree of Doctor of Philosophy

Manipulation of the Random telegraph signals for quantum random number generation

By

Fayong Liu

The random telegraph signals (RTSs) has become a serious reliability issue during the scaling of the CMOS technology. From the fundamental physics point of view, the RTS is generated by the defects at the Si/SiO_2 interface or in the gate oxide layer, they can not be eliminated completely. The study on RTSs topic has led a paradigm shift from eliminating the RTSs to utilizing the RTSs. The RTSs phenomenon provides two contrasting states by the quantum effect. It also has shown that the capture and emission time are related to the bias conditions. Principally, the RTSs can be a promising candidate for the new QRNG source. However, the RTSs are usually generated by randomly distributed traps, it prevents the people to have a certain RTSs source. In this case, there is a high demand to develop a certain and stable RTSs source.

In this PhD project, we successfully fabricated the RTSs source with the artificial quantum dot in the multi-gate silicon nanowire transistors. The RTSs generated from the artificial quantum dot could be manipulated by the bias condition of the multi-gate. The probability of the two states in the typical RTSs could be balanced equally by the voltage control. In addition, the RTSs phenomenon in conventional MOSFETs was investigated at different temperatures. The quantum confinement in our silicon nanowires was characterized.

Table of Contents

Table of Contents v
Table of Tables ix
Table of Figures xi
Research Thesis: Declaration of Authorship xxiii
Acknowledgements xxv
Chapter 1 Introduction 1
1.1 Motivation 1
1.2 Aim 3
1.3 Outline of the thesis 3
Chapter 2 Background and literature review 5
2.1 Random telegraph signal phenomenology 5
2.1.1 The SRH framework 5
2.1.2 Tunneling transitions of the RTSs at different temperatures. 6
2.1.3 The RTS amplitude 7
2.1.4 RTSs in the gate leakage current 8
2.1.5 1/f noise and RTS 8
2.2 Quantum random number generators 9
2.2.1 QRNGs based on the radioactive decay 9
2.2.2 QRNGs based on the noise 10
2.2.3 QRNG based on photon 10
2.3 RTSs in the silicon devices using as QRNGs source 12

2.4 Summary 14

Chapter 3 Random telegraph signals in MOSFET at low temperature........ 15

3.1 Methods........ 15

3.2 Low-temperature measurement system and sample packaging........ 20

3.3 RTSs in a wide-channel MOSFET (W/L=10um/75nm)........ 25

3.3.1 Measurement........ 25

3.3.2 Discussion 35

3.4 RTN in a narrow-channel MOSFET (W/L=60nm/60nm)........ 45

3.4.1 Measurement system upgrade 45

3.4.2 Measurements and discussion 49

3.5 Establish the relationship between the Coulomb diamond, RTSs and QRNG........ 60

3.6 summary 61

Chapter 4 Fabrication of the multi-gate silicon nanowire transistor devices 62

4.1 Introduction 62

4.2 Fabrication process flow 62

4.2.1 Patterning silicon nanowire........ 62

4.2.2 Gate oxide formation 65

4.2.3 Dopant diffusion 66

4.2.4 Patterning the First gate........ 67

4.2.5 Patterning Top gate........ 68

4.2.6 Patterning metal contact........ 69

4.3 Quantum confinement in the silicon nanowires 71

4.3.1 Measurement and discussion 72
4.4 Summary.. 81
Chapter 5 Manipulate the RTSs in the artificial quantum dot with the multi-gate silicon nanowire transistor 83
5.1 Introduction 83
5.2 Device structure... 83
5.3 Measurement and results 84
5.4 Summary.. 93
Chapter 6 Application of the RTSs for QRNG source...94
6.1 Introduction 94
6.2 Device structure... 94
6.3 Measurement and results 95
6.4 Randomness test 107
6.5 Summary.. 109
Chapter 7 Conclusions 110
Appendix A Publication List 113
Bibliography 115

Table of Tables

Table 1 *The comparison of the features between this work and the existing optical QRNGs.* 4

Table of Figures

Figure 2.1 The structure of the triple gates silicon nanowire transistor 13

Figure 3.1 Device layout in the optical microscope. The core of the device was located in the middle with 4 metal pads connected. From left to right, they were Gate, Source, Drain and Bulk electrodes. 15

Figure 3.2 (Left) MEMS prober station with all high-quality triaxial cables feedthrough setting. (Right) Cascade highly reliable probe connecting with device. 16

Figure 3.3 Drain current with gate voltage dependence, drain voltage (Vds) was set to 50mV and 1V: (Left) Linear scale to investigate the threshold voltage. (Right) Log scale to investigate the subthreshold slope 16

Figure 3.4 Drain current with drain voltage dependence. Gate voltage (Vgs) was set to 0V, 0.2V, 0.4V, 0.6V, 0.8V and 1V.. 17

Figure 3.5 Drain current with gate voltage dependence. Drain voltage (V_D) was set to 50mV. The Sex selected points were marked as red diamonds, which represented the I_D in 1 pA, 10 pA, 100 pA, 1 nA, 10 nA, and 100 nA levels.. 18

Figure 3.6 Drain current with gate voltage dependence. Drain voltage (V_D) was set to 50mV. The Sex selected points were marked as red diamonds, which represented the I_D in 1 pA, 10 pA, 100 pA, 1 nA, 10 nA and 100 nA levels. (a-f)The time domain measurement results with the first 50s. (g-l)The probability distribution of the I_D. 19

Figure 3.7 The cryogen-free high field measurement system 20

Figure 3.8 The schematic structure of the cryogen-free high field measurement system 21

Figure 3.9 (Left) The schematic structure of the He-3 insert. (Right) He-3 insert with electrical connections 22

Figure 3.10 The performance of the He-3 insert. The lowest temperature was 298 mK. It can be stable around 40 minutes from 17:37 to 18:18. 22

Figure 3.11 Maximum 3mm 5mm small chips after dicing. The cleaned samples were stored in a jelly sample box.* 23

Figure 3.12 The MOSFETs samples mounted to a 6-pin chip carrier with a yellow no-conductive epoxy. 24

Figure 3.13 The optical microscope view of the sample which was mounted and successfully bonded to a 6-pin chip carrier with 4 gold wires. 24

Figure 3.14 Drain current with gate voltage dependence for the pMOSFET (10 μm long and 75 nm wide). V_D was set to 50mV to make a stable measurement temperature. 25

Figure 3.15 The contour plot for drain current versus gate voltage and drain voltage of the pMOSFET at 2K. The bluepoints were selected conditions to extract the gate voltage and drain voltage. Then the RTSs time-domain measurement was carried on the voltage conditions of these points. The results were shown in Figure 3.18-3.20 and 3.22-3.24. 26

Figure 3.16 Horizontal peaks happened at the edges of the Coulomb diamond areas 27

Figure 3.17 One I_DV_G curve (V_D=-13.5 mV) extracted from Figure 3.15. The step changes of the current can be found around V_G=-0.65 V (In the Area 1 of Figure 3.16). 27

Figure 3.18 The time-domain measurement of I_D with the bias condition of V_D=-13.5 mV and V_G=-0.64 V. Three types of RTSs were found. 28

Figure 3.19 (a-g) The time-domain measurement results of the V_G dependence for the RTSs. (h-n) The probability distribution of the current versus its valve. The high state of RTS1 was marked as red "H"; the low state of RTS1 was marked as blue "L". 30

Figure 3.20 The time-domain measurement results in Figure 3.19 (c) were divided into (a) (from 0 to 1605s) and (b) (from 1605s to 3500s).of the V_G dependence for the RTSs. (c,d) The probability distribution of the current versus its valve. The high state of RTS1 was marked as red "H"; the low state of RTS1 was marked as blue "L". 31

Figure 3.21 The "break-out box" to make the connect between measurement equipment and insert probe. The green tube was the bottom part of the inset probe, where the chip carrier mounted on. The "break-out box" could directly measure the sample before loading into the cryostat. 32

Figure 3.22 (b-f) The time-domain measurement results of the V_D dependence for the RTSs. (h-l) The probability distribution of the current versus its valve. The high state of RTS1 was marked as red "H"; the low state of RTS1 was marked as blue "L". (a,g) The RTS5 appeared. The high state of the RTS5 was marked as red "H'"; the low state of the RTS5 was marked as blue "L'". 33

Figure 3.23 (a-c) The time-domain measurement results of the V_D dependence for the RTSs. (d-f) The probability distribution of the current versus its valve. 33

Figure 3.24 The time-domain measurement results in the Coulomb diamonds. 34

Figure 3.25 The contour plot for drain current versus gate voltage and drain voltage of the pMOSFET at 2K. 35

Figure 3.26 The equivalent circuit model for the single hole transistor 36

Figure 3.27 (a) The schematic diagram for the structure model of the device. The quantum dot was formed by the remote Poly-Si grains. (b) The band structure diagram in the

channel. The E_f was the Fermi energy level. The E_{VB} was the valence band energy level. 37

Figure 3.28 The schematic diagram for the structure model of the device. Many different quantum dots were formed by the randomly distributed Poly-Si grains 38

Figure 3.29 The average high-state current (τH) and the average low-state current (τL) as a function of gate voltage for the RTS1 and RTS2. 39

Figure 3.30 The $\Delta I/I$ as a function of gate voltage for the RTS1 and RTS2. 40

Figure 3.31 The physical model to explain the original source to form the RTS1 and RTS2. 40

Figure 3.32 The schematic diagram of the potential in the channel cross-section from source to drain. 41

Figure 3.33 The V_G dependence of P_H and *P_L in RTS1. V_D bias was -13.5 mV.* 42

Figure 3.34 The V_D dependence of P_H and *P_L in RTS1. V_G bias was -640 mV.* 42

Figure 3.35 The potential diagram with the hole trapping and de-trapping from the oxide layer to the gate electrode. 43

Figure 3.36 The V_G dependence of P_H and *P_L in RTS2. V_D bias was -13.5 mV.* 44

Figure 3.37 The V_D dependence of P_H and *P_L in RTS2. V_G bias was -640 mV.* 44

Figure 3.38 New "Break-out" box with full metal shield. 46

Figure 3.39 The original connection panel layout with all singles mixed in one 12-pin Fisher connector[134]. 46

Figure 3.40 The modified connection panel layout with 4-SMA connectors and 2-triaxial connectors to separate the singles by independent guarding and shielding. 47

Figure 3.41 (Left) The original chip carrier mounted tube, which could mount 2 chip carriers at the same time. (Right) The re-designed chip carrier mounted tube, which only had 2 chip carrier position, but low noise level. 47

Figure 3.42 (a) 4 SMA cables and 2 triaxial cables making the electrical connection between the insert probe and outside equipment. (b) SMA cables connecting to triaxial cables by guard disconnected adaptors. (c) The long triaxial cables through the ceiling to reach the Agilent B1500. (d) Triaxial cables connecting with Agilent B1500 SMU ports. 48

Figure 3.43 The contour plot for I_D versus V_G and V_D of the nMOS at 5K. 49

Figure 3.44 The contour plot for I_D versus V_G and V_D of the nMOS at 10K. 49

Figure 3.45 The contour plot for I_D versus V_G and V_D of the nMOS at 20K. 50

Figure 3.46 The contour plot for I_D versus V_G and V_D of the nMOS at 40K. 50

Figure 3.47 The contour plot for I_D versus V_G and V_D of the nMOS at 60K. 50

*Figure 3.48 (a-g) The time-domain measurement results of the V_G dependence for the RTSs at 10 K. (h-n) The probability distribution of the current versus its valve. The high state of RTS was marked as red "H"; the low state of RTS was marked as blue "L".*52

Figure 3.49 The V_G dependence of P_H and *P_L in RTSs. V_D bias was 20 mV.* 53

Figure 3.50 The τH and the τL as a function of gate voltage for the RTS. 54

Figure 3.51 The $\Delta I/I$ as a function of gate voltage for the RTS. 54

Figure 3.52 The physical model of the RTS. (a) The unoccupied state of the additional energy state. (b) The occupied state of the additional energy state. 55

Figure 3.53 (a-f) The time-domain measurement results of the V_G dependence for the RTSs at 40 K. (g-l) The probability distribution of the current versus its valve. The high state of RTS was marked as red "H"; the low state of RTS was marked as blue "L". 56

Figure 3.54 (a-g) The time-domain measurement results of the V_G dependence for the RTSs at 60 K. (h-n) The probability distribution of the current versus its valve. The high state of RTS was marked as red "H"; the low state of RTS was marked as blue "L". 57

Figure 3.55 (a) The time-domain measurement results of the V_G dependence for the RTSs at 100 K. (b) The probability distribution of the current versus its valve. The high state of RTS was marked as red "H"; the low state of RTS was marked as blue "L". 58

Figure 3.56 (a) The time-domain measurement results of the V_G dependence for the RTSs at 150 K. (b) The probability distribution of the current versus its valve. The high state of RTS was marked as red "H"; the low state of RTS was marked as blue "L". 58

Figure 3.57 (a) The time-domain measurement results of the V_G dependence for the RTSs at 200 K. (b) The probability distribution of the current versus its valve. The high state of RTS was marked as red "H"; the low state of RTS was marked as blue "L". 59

Figure 3.58 The comparison of the I_D and I_S of the RTS at the same time scale. (a) The I_D time-domain measurement results of the V_G dependence for the RTS (b) The I_S time-domain measurement results of the V_G dependence for the RTS. 59

Figure 4.1 The fabrication process flow chart 62

Figure 4.2 Schematic diagram after E-beam lithography 63

Figure 4.3 Schematic diagram after RIE 63

Figure 4.4 Schematic diagram after HF etching 63

Figure 4.5 Schematic diagram after TMAH etching and undercut part 64

Figure 4.6 SEM view of the nanowire after TMAH etching 64

Figure 4.7 SEM view of the nanowire after removing the hard mask 65

Figure 4.8 The nanowire dimensions from the TCAD simulation (HSQ mask: 30 nm)..............65

Figure 4.9 The nanowire dimensions from the TCAD simulation (HSQ mask: 50 nm)..............65

Figure 4.10 The nanowire dimensions from the TCAD simulation (HSQ mask: 75 nm) 66

Figure 4.11 SEM image for the diffusion windows location..66

Figure 4.12 Schematic diagram after depositing the Poly-Si layer..67

Figure 4.13 Schematic diagram after ICP etching 67

Figure 4.14 SEM image after ICP etching 68

Figure 4.15 Schematic diagram after patterning the Top gate...68

Figure 4.16 Optical image after patterning the Top gate...69

Figure 4.17 Optical image for the contact windows...69

Figure 4.18 (Left) Optical image for the metal contact. (Right) Optical image for one device ..70

Figure 4.19 Optical image for the damaged metal layer. 70

Figure 4.20 The schematic diagram of the silicon nanowire transistor and the layer structure. 71

Figure 4.21 The optical image of the silicon nanowire transistor. 71

Figure 4.22 (a) The planar view of the device region. (b) The planar view of nanowire region. (c) The cross-sectional view in AA direction (blue dash line). (d) The cross-sectional view in BB* direction (yellow dash line).* 72

Figure 4.23 Drain current versus First gate voltage characteristics on drain voltage at 50mV and 1V with different channel width. 73

Figure 4.24 Threshold voltage of the silicon nanowire transistors with different width. 73

Figure 4.25 Quantum confinement in the smaller dimensional silicon devices. 76

Figure 4.26 The diameter dependence of the energy bandgap in a silicon nanowire by fitting the results from the literature. 77

Figure 4.27 2D transistor model in TCAD. 77

Figure 4.28 I_D *versus* V_{FG} *characteristics on* V_D*=50 mV with different bandgaps and affinity.* . 78

Figure 4.29 The schematic diagram to show the bandgap expansion in different part of the nanowire channel from the cross section of view 79

Figure 4.30 The electric field distribution in the cross section of the nanowire from the TCAD simulation (HSQ mask: 30 nm). In the left, V_{FG}*=0 V. In the right,* V_{FG}*=0.5 V* 79

Figure 4.31 The electric field distribution in the cross section of the nanowire from the TCAD simulation (HSQ mask: 50 nm). In the left, V_{FG}*=0 V. In the right,* V_{FG}*=0.5 V* 79

Figure 4.32 The electric field distribution in the cross-section of the nanowire from the TCAD simulation (HSQ mask: 75 nm). In the left, V_{FG}*=0 V. In the right,* V_{FG}*=0.5 V* 80

Figure 4.33 The multi-channel transport model to explain the SS degradation 80

Figure 5.1 The 3D schematic structure of the device. TG was not shown [142]. 83

Figure 5.2 The cross section of view from AA' direction in Figure 5.1[142]. 84

Figure 5.3 (a)The I_D *versus* V_{LG} *characteristics with* V_{TG}*=0 V and* V_{RG} *floating at* V_D*=50 mV. (b) The* I_D *versus* V_{RG} *characteristics with* V_{TG}*=0 V and* V_{LG} *floating at* V_D*=50 mV [142].* 84

Figure 5.4 (a) The I_D versus V_{TG} characteristics with different V_{LG}= V_{RG} at V_D=50 mV in log scale [142]. 85

Figure 5.5 The marked area in Figure 5.4 where the RTSs could be observed [142]. 86

Figure 5.6 (a-f) The time-domain measurements on V_{TG} dependence with V_{LG}=V_{RG}=0 V and V_D=50 mV. (g-l) The corresponding probability distribution of the I_D [142]. 87

Figure 5.7 (a)The occupation of the I_D in "High" and "Low" states with different V_{TG}. (b)The schematic diagram of the energy level of the quantum dot on V_{TG} dependence[142]. 87

Figure 5.8 (a) The τH and τL with the V_{TG} dependence. (b) The ratio of the $\tau H \tau L$ with the V_{TG} dependence in the log scale by a linear fitting [142]. 88

Figure 5.9 (a) The energy level of the quantum dot on the V_{TG}. The dash line was the linear fitting for the extracted points from the measurement results. (b-d) The schematic diagram of the energy level in the quantum with different V_{TG} bias [142]. 89

Figure 5.10 The time-domain measurement results on V_{LG}=V_{RG} dependence with V_{TG}=50 mV and V_D=50 mV [142]. 90

Figure 5.11 (a-d) The typical probability distribution of the I_D in the selected V_{LG}=V_{RG} bias condition. (e-f) The schematic diagram of the different energy levels in the quantum dot corresponding to (a-d) [142]. 91

Figure 5.12 (a-d) The schematic diagram of the equivalent circuit for the multi-gate silicon nanowire transistor [142]. 91

Figure 5.13 (a) The energy level of the quantum dot with different First gates voltage. The filled circles and squares were extracted from the measurement data for RTS1 and RTS2 respectively. The lines were the linear fitting for the different energy level ($E0h, E1h$ and $E0l$) by the effective mass approximation [142]. 92

Figure 6.1 The structure of the triple gates silicon nanowire transistor with the atomically flat interface (a) The artificial quantum dot created by TG, LG and RG bias. (b) The schematic diagram of the device. 94

Figure 6.2 Dimension confirmation for the nanowire after TMAH in SEM image. 95

Figure 6.3 I_D-V_{TG} *characteristics with* $V_{LG}= V_{RG}=0V$ *at* $V_D=50$ *mV and 1 V*.................... 95

Figure 6.4 I_D-V_{LG} *characteristics with* $V_{TG}=0$ *V and* V_{RG} *floating at* $V_D=50$ *mV and 1 V* 96

Figure 6.5 I_D-V_{RG} *characteristics with* $V_{TG}=0$ *V and* V_{LG} *floating at* $V_D=50$ *mV and 1 V* 96

Figure 6.6 The Multi-level RTSs observed in low, medium, and high V_{TG} *region and clearly shift the Vth in the linear region.* 97

Figure 6.7 The time domain measurements with different V_{TG}. 97

Figure 6.8 The lag plots with time lag (Δt) at 108 ms and 100 s to show the local correlation of the I_D *with different* V_{TG} bias. 98

Figure 6.9 The time domain measurements with different $V_{RG}=V_{LG}$. 99

Figure 6.10 The schematic diagram of the energy level in the quantum dot. 100

Figure 6.11 The lag plots with time lag (Δt) at 108 ms and 100 s to show the local correlation of the I_D *with different* $V_{RG}=V_{LG}$ bias. 100

Figure 6.12 The variation tendency of probability distribution of I_D *in "high (h)" and "low (l)" states with* $V_{LG}= V_{RG}$ *from 50 mV to -200 mV in the low* V_{TG} *region.* 102

Figure 6.13 The variation tendency of the probability distribution of I_D *in "High (H)" and "Low (L)" states with* $V_{LG}= V_{RG}$ *from 200 mV to -400 mV in the high* V_{TG} *region.* 102

Figure 6.14 (a) The occupation rate of the three states on V_{RG}=V_{LG} dependence with V_{TG}=200 mV. (b)The occupation rate of the three states on V_{TG} dependence with V_{RG}= V_{LG}=-100 mV. 103

Figure 6.15 The time-domain measurement at V_{TG}=700 mV, V_{RG}= V_{LG}=0 V and V_D=50 mV. The RTS2 was marked as "H" and "L". The RTS1 was marked as "h", "l" and "l'". 104

Figure 6.16 The occupation rate of the "HIGH" (H) and "LOW" (L) state in RTS2 (a) on V_{RG}=V_{LG} dependence with V_{TG}=700 mV and (b) on V_{TG} dependence with V_{RG}= V_{LG}=-200 mV. 104

Figure 6.17 The occupation rate of the "high" (h), "low" (l) and "low2" (l') states of RTS1 (a) in the "HIGH" (H) state of RTS2 and (b) in the "LOW" (L) state of RTS2 on V_{TG} dependence with V_{RG}=V_{LG}=-200 mV. 105

Figure 6.18 The time-domain measurement on V_{TG}=300 mV, V_{RG}=V_{LG}=-106 mV and V_D=50 mV. 106

Figure 6.19 The lag plots with time lag (Δt) at (a) 24 ms and (b) 1 s to show the local correlation of the I_D with a certain voltage bias. 107

Figure 6.20 (a) The 0 and 1 digital numbers extracted from the re-sampling data. (b) The number count of 0 and 1. 107

Figure 6.21 (a) Self-correlation plots for 8-bit random numbers generated by our device. (b) Self-correlation plots for 8-bit random numbers generated by MATLAB. 108

Research Thesis: Declaration of Authorship

Print name: Fayong Liu

Title of thesis: Manipulation of the Random telegraph signals for quantum random number generation

I declare that this thesis and the work presented in it are my own and have been generated by me as the result of my own original research.

I confirm that:

1. This work is based on a research project by a candidate in collaboration with researchers in the University of Southampton;

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

3. Where I have consulted the published work of others, this is always clearly attributed;

4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

5. I have acknowledged all main sources of help;

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. Parts of this work have been published as:

   References [142],[143], and [154].

Signature: Date:

Acknowledgements

I would like to express my gratitude to my parents, Dr. Shaoqing Liu and Xiuping Shen, for their great financial and moral support. It was solid backing during my PhD life.

I would like to sincerely thank my supervisor, Prof. Shinichi Saito, for his great support in finance and excellent supervision both in my MSc and PhD courses. His rigorous academic attitude highly affected my self-improvement. He is not only an academic supervisor but also my career mentor.

I also would like to sincerely thank my second supervisor, Dr. Yoshishige Tsuchiya, for his full support in the experimental activities. His valuable experience helped me establish the standard methodology in the high-quality measurement in Nanoelectronics.

I also would like to sincerely thank Prof. Harvey Rutt, Dr. Jonathan Fletcher, Dr. Stephen Giblin and Dr. Masaya Kataoka for their creative support to help me establish the low-temperature and high-frequency measurement systems.

I also would like to sincerely thank Mr. Mark R. Long and Mr. James C. Chalk for their full support in the Mechanical Workshop.

I also would like to sincerely thank Dr. Muhammad K. Husain and Dr. Zuo Li for their hard work and help in the device fabrication.

I also would like to sincerely thank all the members of the Sustainable Electronics and Technology group for their help and advice, and all the technicians in the Southampton Nanofabrication Centre for their help and training.

I would like to appreciate all other colleagues, Dr. Abdelrahman Z. R. Al-Attili, Dr. Kapil Debnath, Dr. Yichi Zhang, Dr. Moise Sotto, Dr. Daniel Burt, Mr. Kouta Ibukuro, Mr. James Byers, Mr. Joseph W. Hillier to build a family-like working environment to help each other and learn from each other.

Chapter 1 Introduction

1.1 Motivation

The semiconductor industry development had been guided by Moore's law in the past 50 years [1] which has promoted the continuous technological progress in computing power. Nowadays, computer-based modern communications, designs, data analysis, healthcare, and artificial intelligence (AI) computing have been used as an essential part of everyone's life. The core in a personal electronic device even has similar computing power to one of the modern supercomputer cores. After Apple iPhone XS was launched in September 2018 and AMD Radeon VII graphics card was released in January 2019, both the mobile device and the desktop computer had been announced to step into 7 nm node generation. During the time while people are enjoying the latest technology and expecting the next amazing product, designing a reliable electrical system has become increasingly difficult. With aggressively scaling the Complementary Metal–Oxide–Semiconductor (CMOS) transistors, the reliability issues have become a nightmare for semiconductor engineers. In recent researches, random telegraph signals (RTSs) had attracted so much attention from the fundamental physics researchers as well as the device engineers. When the device size was approaching the level of atoms, RTS had emerged as a source of a series of reliability and variability issues which caused the failure of device [2,3,4]. The detailed reason was that the number of activated traps at the Si/SiO2 interface or in the dielectrics was reduced to a countable number during the scaling of the Metal-Oxide-Semiconductor-Field-Effect Transistor (MOSFET) area. In this condition, the carrier transport in the scaled transistors would be influenced just by one or a few activated defects. If one or a few traps randomly influenced the current in nearly 100% amplitude, the device would lose the function as a logic element [5].

RTS was first demonstrated by Ralls *et al.* in MOSFETs with 1 um channel length and 100nm channel width in 1984 [6]. It was simply illustrated as the drain current I_D or the drain voltage V_D which was randomly switched between two levels, "high" and "low" [7]. Researchers described the current which temporally switched between two levels as the high state "high" and the low state "low". The variation between high-level and low-level currents was characterized by the switching amplitude ΔI_D.

The fundamental physical properties of the RTS were usually interpreted as a single electron or hole captured and emitted by a trap, which fluctuated the current characteristics in orders of magnitude. In MOSFETs, these traps were usually located at the Si/SiO_2 interface or in the gate oxide layer [5,7], which were generated in the fabrication processes. Engineers had tried some strategies to suppress

the effect of the RTSs both in the process-level and in the circuit-level. For example, by changing the dielectric materials in the fabrication process, the oxide interface trap density for the non-nitrided transistors was one order of magnitude lower than the nitrided devices [8]. In the circuit level, by using the correlated multiple sampling differential averaging (CSMDA) technique, the RTS and thermal noise could be effectively reduced [9]. However, the CSMDA technique was not so effective for the multi-level RTSs. Actually, the charged trap could never be eliminated completely from the view of fundamental physics. Then it led a paradigm shift from fighting with the RTSs to utilizing the RTSs phenomenon.

The discrete nature of RTS provided a unique way to investigate the single-electron phenomenon caused by activated traps in the semiconductor devices. Researchers can not only extract the properties of the traps such as the energy level and capture cross-section [5], but also derive their physical position such as the trap depth and lateral position in the channel [5]. From the application perspective, the “high and “low” states of the RTSs provided the essential element for a quantum random number generator (QRNG) source, in which the switching behavior was dominated by the pure quantum mechanics. From the previous study of the RTSs [10,11,12], it had been shown that the capture and emission time had a strong dependence on the voltage bias, which gave a way to make an unbiased “high” and “low” states. In principle, the RTSs could be a promising candidate as a QRNG source.

The QRNGs produced random numbers with pure quantum mechanical effects [13]. A random number should be unpredictable and cannot be reproduced. Random numbers were the foundational element for information security and they built the blocks of encryption, key wrapping, signing, one-time codes, authentication, and other cryptographic applications [14]. Random numbers were also essential elements for computer assistant modeling and modern digital gaming or lottery. There is no doubt that the quality of the random numbers had a fatal impact on modern security. Attackers do not usually attempt to crack encryption, they simply steal or guess keys. The digital keys generated by the poor quality random numbers would be much easier to be cracked by the attackers. Comparing with the classical random number generators [15-18], the QRNGs delivered the higher quality random number due to their intrinsic randomness from the quantum physical processes.

Some commercial QRNGs products had been available in the market [19-24] and most of them were based on optical quantum phenomena. Today’s technology had provided a broad range of choices for both the light source and the precise detectors [25-27]. However, as long as the QRNGs were based on the optical quantum effects, the common issues such as unbalanced detectors [28], detector dead time [28-30], time precision [29], photon resolving capability [30], source instability

[31] and so on, could not be avoided. In addition, it increased the difficulty of integrating the QRNGs into the VLSI circuit while the light source was needed. According to most of the commercial QRNGs outputting random numbers by the USB link as an independent component, the discrete QRNG devices from the motherboard or central processing unit (CPU) would increase the risk that the attackers might modify the random numbers in the software level.

The QRNGs based on the RTSs would avoid the common issues of the optical QRNGs without using any light source and detectors. As the RTSs could be generated by the standard MOSFETs, it could be easily integrated into the VLSI design to form a private communication environment between the QRNGs and the logical units which provided a new way to enhance the security level at the hardware layer. However, there were few proposals for QRNG based on the RTSs demonstrated in the past decade years. The main reason was that the traps were randomly generated in the devices and the standard statistical investigation was quite time-consuming [32-37]. A certain RTSs source was highly demanded to develop toward the application of QRNGs.

1.2 Aim

The aim of this PhD project was to develop a RTS source with an artificial quantum dot in multi-gate silicon nanowire transistors. To achieve the final aim, the standard MOSFETs needed to be measured at different temperatures to investigate the fundamental physics of the RTS phenomenon and establish the characterization method to identify the RTSs. Based on the established method, the multi-gate silicon nanowire transistors needed to be designed and fabricated by collaborating with the senior researchers. Then the RTSs from the artificial quantum dot needed to be evaluated whether it could be a candidate as a QRNG source or not from the perspective of the fundamental physics.

1.3 Outline of the thesis

In chapter 2, the development of the QRNGs and the research progress in the RTS phenomenology were reviewed. Based on the different kinds of quantum effect phenomena, the advantages and challenges of the existing QRNGs were summarised. Then the design concept of the QRNG based on the RTSs was pointed out.

In chapter 3, the RTSs in the conventional MOSFETs were measured at different temperatures with a cryostat system. In order to get the high-resolution measurement results, the cryostat was re-designed to decrease the noise level. Two types of RTSs in both of the wide-channel MOSFETs and

narrow-channel MOSFETs were observed and investigated. The study established the theoretical basis to manipulate the RTSs in multi-gate transistor devices.

In chapter 4, the fabrication processes for the multi-gate the silicon nanowire devices were demonstrated in detail. After successfully finishing the fabrication of the devices, the transport properties of the silicon nanowires were investigated with a single gate. The quantum confinement in the special type of silicon nanowires was discussed.

In chapter 5, the main devices with the artificial quantum dot were characterized at room temperature. Not only the typical RTSs were observed in a certain bias condition, but also the probability of the "high" and "low" states could be manipulated by the multi-gate voltages. The functionalities of the top gate and the first gates were identified.

In chapter 6, the main devices were measured in different bias combinations to investigate the quality of the RTSs to be as a QRNG source. The way to find a better quality of RTSs generated by the artificial quantum dot was established. The challenges for our design to be a practical QRNG source were discussed based on the measurement results.

Chapter 7 concluded my PhD studies. The main features of the QRNG in this work was shown in the table below comparing to the existing optical QRNGs.

Table 1 The comparison of the features between this work and the existing optical QRNGs.

	This work	Optical RNG
Qubit	Single-electron	Single-photon
Speed	1bps (Potentially 10Gbps)	Mbps Gbps
Working Temperature	Room temperature	Room temperature
Size	Transistor scale	Equipment scale
Integration	CMOS compatible	No
Calibration	Gate bias calibration in hardware level	Mainly software level calibration

Chapter 2 Background and literature review

2.1 Random telegraph signal phenomenology

Since the report of the random telegraph signals (RTSs) was published in 1984 [6], most of the research on the RTSs were based on the MOSFETs structure [5]. The typical RTS phenomenon was that the drain current or drain voltage was temporally switched between two levels [7]. The RTS characterization mainly included three elements: the time in high state, the time in low state and the switching amplitude. Usually, the RTS phenomenon was explained according to the frame of the Shockley-Read-Hall (SRH) theory [6, 37]. The RTSs were driven by thermal energy at room temperature [38-41]. At low temperatures, while the thermal energy was decreased, the RTS was mainly characterized by the different tunneling transitions [42, 43]. In addition, the RTS amplitude could be extracted from the measurement results even without the statistical distribution, but it was quite difficult to derive a universal model to describe the RTS amplitude [5]. It was reported that the RTS amplitude could be related to the carrier number fluctuation in the channel [44], the change of the channel mobility [45-48], the gate voltage and drain voltage [49, 50], the impact of the device scaling [51-54] and so on. Moreover, the RTSs in the gate current was observed in the thin oxide layer devices [55]. It should be carefully distinguished with the channel RTS. The details were discussed in the following parts.

2.1.1 The SRH framework

The SRH theory could be used to describe the capture (τ_c) and emission (τ_e) time for the charge trap. By assuming that the charged trap was allocated in the interface between the channel and gate oxide layer in the n-channel devices, the average capture time for an electron from the inversion layer could be described as [6, 37],

$$\tau_c = \frac{1}{n\sigma_n v_{thn}} \tag{2.1}$$

The n was the volume concentration, which was proportional to the drain current. The v_{thn} was the thermal velocity for the electrons. The precondition to use this equation was that the inversion layer quantization could be neglected [37]. The ratio of the capture and emission time constant also could be described as [6],

$$\frac{\tau_c}{\tau_e} = g\ exp\left(\frac{E_T - E_F}{kT}\right) = \frac{1 - f_T}{f_T} \tag{2.2}$$

The E_F was the Fermi level in the surface. The E_T was the energy level of the charge trap. The k was the Boltzmann constant. The f_T was the occupancy of the charge trap by an electron. The g was the trap degeneracy, which was usually assumed to 1. By considering [7],

$$n = N_C \, exp\left(\frac{E_C - E_F}{kT}\right) \tag{2.3}$$

Where N_C was the density of the states and E_C was the bottom of the conduction band, the τ_e could be derived as,

$$\tau_e = \frac{1}{N_C \sigma_n v_{thn}} \, exp\left(\frac{E_T - E_F}{kT}\right) \tag{2.4}$$

In the p-channel devices, we could replace the σ_n by σ_p, the v_{thn} by v_{thp} (the thermal velocity for holes) and $E_C - E_F$ by $E_T - E_V$. The E_V was the top of the valance band.

In the SRH theory, it also demonstrated the way to identify the low state and high state with either the capture or the emission behaviors. As an example, for the acceptor charge trap in an nMOS, the charge state was switched from more positive to more negative by trapping an electron. By changing the local potential in the channel, the capture of an electron was corresponding to the low current state, while the emission of the electron was corresponding to the high current state [10]. By obtaining the τ_c and τ_e parameters, the feature of the charge trap such as energy level, barrier height and location could be derived by the SRH approaching method [37, 46, 56-59].

However, some results had shown that the RTSs that were influenced by the Coulomb blockade effects were exceeded of the validity range of the SRH theory [60, 61]. The Coulomb blockade happened on the way where the electron transfer from the semiconductor to the charged trap. It formed a single electron transistor system, which should be taken into account while calculating the τ_c and τ_e parameters [61]. Especially at the cryogenic temperature, while the thermal energy was quite small, the electron or hole might be bound to the charge trap in a bi-stable configuration [62, 63]. In this case, the typical RTSs could not be observed because no carrier was exchanged between the semiconductor and the charge trap.

2.1.2 Tunneling transitions of the RTSs at different temperatures.

At room temperature, the typical RTSs behavior was recognized as being thermally activated [6-7, 37]. It implied that the elastic tunneling effect was smaller enough to be negligible at the energy level around E_F [65]. By decreasing the temperature, it was observed that the thermally activated transitions were gradually changed to the temperature-independent tunneling transitions [43]. The temperature-independent tunneling transitions were considered between the localized state and

the two-dimensional electron gas (2DEG) in the channel, which dictated by the Fermi's golden rule [42]. While the temperature was decreased to the liquid helium temperatures, the tunneling two-state system was identified at the Si/SiO_2 interfaces due to the degeneration of the 2DEG [12, 66]. It was suggested that the two-level system was only caused by the energy exchange between the charged trap and the 2DEG without carriers exchange [66]. That induced the change for the inversion layer electrons in the scattering cross-section, which led the channel resistance changing. It also observed that the energy separation between the two-level states could be controlled by the gate voltage [67]. At the milli-kelvin temperature, the tunneling transitions between the single defect state to the 2DEG showed the finite temperature counterpart of the Fermi-edge singularity [67]. A peak of the tunneling rate could be observed while the energy level of the defect was aligned to the Fermi level of the channel, which was consistent with the Fermi-edge singularity theory at the finite temperature [67].

2.1.3 The RTS amplitude

The RTS amplitude was usually normalized to $\Delta I/I_D$ [5]. By assuming that the channel was uniform, the first model to describe the RTS amplitude was demonstrated that the local resistivity (*R*) was enhanced by the trap capturing an electron [44]. That resulted,

$$\frac{\Delta R}{R} = \frac{1}{WL}\frac{q\beta}{(C_{OX} + C_D - \beta Q_n)} \tag{2.5}$$

The β was equal to q/kT. The *W* was the width of the channel. The *L* was the length of the channel. The C_{OX} was the oxide capacitance per unit of area. The C_D was the depletion capacitance per unit of area. The Q_n was the channel charge density in $/cm^2$. It implied that the normalized RTS amplitude was affected by the total number of the carriers (*N*) in the channel, due to [44, 49]

$$\frac{1}{N} = \frac{q}{WLQ_n} \tag{2.6}$$

Although the model could explain the gate voltage dependence of the RTS amplitude, it could explain the large RTS amplitude, especially in the weak inversion due to the image charge screening effect [68-70].

It was also demonstrated that normalized RTS amplitude was related to the change of the channel mobility ($\Delta\mu$) [45,47]. After the oxide trap captured a carrier, the local electrical potential would be changed. This induced the scattering increase in the local cross-section, and then changed the channel mobility. If assuming the channel was uniform and the change was uniform, the relationship between the normalized RTS amplitude and the mobility could be [45, 47],

$$\frac{\Delta I}{I} = \frac{\Delta n_S}{n_S} + \frac{\Delta \mu}{\mu} \quad (2.7)$$

The n_S was the surface carrier density. There was evidence to show that the Δn_S was dominated by the $\Delta \mu$ [46, 48].

In addition, it was indicated that the RTS amplitude was not only dependent on the gate voltage but also dependent on the drain voltage [46, 49-50, 71-72]. By exchanging the role of the source and drain, the asymmetry RTS amplitudes were usually observed, which could be used to extract the lateral trap position [49,71] and to be as a local probe for the surface potential in the channel [72].

Moreover, it was confirmed that the RTS amplitude was increased significantly during device scaling [51-54]. In the ultra-narrow transistors, the charged trap could be just located in the strategic position to block the channel current by changing the surface potential. However, in the thin gate oxide devices, the screening of the metal gate influenced the RTS amplitude drastically [5]. It had been confirmed that the image charge effect from the gate electrode decreased the RTS amplitude significantly [73].

2.1.4 RTSs in the gate leakage current

Since the thin gate oxide layer was widely implemented in the advance CMOS technology, the RTS in the gate leakage current attracted more attention both from the researchers and from the engineers [55, 74-81]. Even in the high-κ material transistors, the RTSs in the gate current was also observed [82-85]. It was pointed out that the gate leakage current in high state was corresponding to the charge trap capturing the electron, which was opposite to the channel RTS [55]. The reason was suggested as the trapped electron in the oxide layer increased the transmission probability to cross the dielectric layer by the image force reduction [5]. In this case, while the RTSs were observed in the drain current, the source, drain and gate leakage should be all plotted out. It should be clearly distinguished between the RTS in the gate leakage current and the RTS in the channel current.

2.1.5 1/f noise and RTS

The 1/f noise was a noise signal with the functionalized behavior in frequency spectrum that the power spectral density was proportional to frequency inversely, shown as:

$$S(f) \propto \frac{1}{f^{\alpha}}$$

Where S is the power spectral density, f is the noise frequency and $0 < \alpha < 2$. The 1/f noise was an intermediate between the white noise and Brownian motion noise, which was neither correlation in time nor correlation between increments. In this case, the 1/f noise can not be generated by integration or differentiation from the common signals. And there were no simple formulas to generate the 1/f signals. Enormous experimental data had exhibited that the 1/f noise was commonly observed in both electronic and optical devices [145,146]. However, as the 1/f noise was defined from the observed features, the physical mechanism has not been clarified. In the nanoscale electronic devices, the RTS has been proposed to be the possible origin for the 1/f noise in the microscopic view [147]. From another point of view, the 1/f noise could be the macroscopic feature of the summation of many RTSs in the large scale devices.

2.2 Quantum random number generators

In the 1980s, the pioneers including Paul Benioff, Yuri Manin, Richard Feynman and David Deutsch initially established the quantum computing research field [86-88], which theoretically gave the hope to have the computing capability to against the exponential growth of Hilbert space with increasing the number of particles. Quantum computing was based on quantum effect phenomena including superposition, entanglement and so on [89]. After nearly 40 years' developing, the experimental achievements in quantum technology areas in the last decades had shown that quantum physics could offer some new ways in computation, security, and cryptography. One important, emerging and more closed to daily life quantum technology was quantum random number generation. Quantum random number generators (QRNGs) were devices that used quantum mechanical effects to produce random numbers and applied to the simulation or cryptography. Comparing with classical physics, quantum physics was fundamentally random. Its intrinsic randomness became the overwhelming superiority in random number generator design. The main types of QRNGs were summarized in the following parts with their fundamental physics, strength, and limitation.

2.2.1 QRNGs based on the radioactive decay

The first QRNGs was designed based on the radioactive decay [90-91]. The core of this kind of QRNGs mainly included the radioactive samples and the Geiger-Muller (GM) tubes [92]. In the GM tube, the ionization event was produced by a single particle and was amplified in the Townsend avalanche. The device could generate the pulse for each detected particle. The probability for the numbers of the atom to decay in a certain time interval would be an exponential random number, which followed the Poisson distribution. In addition, the time intervals between different detected pulses would be also an exponential random number. That time intervals would not be affected by

the previous results [93]. By replacing the GM detectors to the semiconductor detectors in the modern proposals [94-97], the QRNG based on the radioactive decay was still a convenient method to generate high-quality random numbers. However, there were obvious limitations in the practical using of the QRNG based on the radioactive decay. The first one was due to the radioactive source. In order to improve the speed of the random generation, the highly radioactive source like Nickel-63 [97], Americium-241 [96], Cesium-137 [98], Strontium-90 [99], Cobalt-60 [90] were introduced. Meanwhile, these radioactive substances required strict health and safety measures. This problem prevented the device to be commonly used and made it quite difficult to integrate into the computer circuit. The second one was that the bit generation rate was limited by the dead time from the detector recovery. The ions generated in the GM tube would stop the further avalanches which amplified each count until they recovered to the normal state [92]. So that the dead time was the essential time interval for the GM tube recovering to the full capability, which would be around microseconds level [90, 96-99]. Even in the semiconductor detectors, they also needed a microseconds time interval to the carriers replenishing after each detection [94-97]. The last but most one was that the detector could be degraded by the radiation. Especially in the semiconductor detectors, the radiation would damage the device surfaces and change the electrical properties [94-95].

2.2.2 QRNGs based on the noise

Most of the noise-based random number generators belong to classical physical random number generators, like thermal noise [15], avalanche noise [16], dark noise [17] and so on. Besides, the shot noise-based devices [100-103] could be considered as QRNGs. The shot noise was generated by the discrete carriers passing through the potential barrier, which showed the quantum fluctuations. However, the thermal noise was generated by the thermal fluctuation of the carrier, which had the temperature dependence. Practically, these two kinds of noise were difficult to separate and mixed together [104]. So that the QRNGs based on the shot noise would be affected by many environmental fluctuations. It implied that the quantum effects were not well controlled and isolated in the QRNGs based on shot noise. Due to the simple electrical circuits, some commercial QRNGs based on the shot noise had been available in the market [101].

2.2.3 QRNG based on photon

The optical QRNGs had the largest family in the latest research. The light provided the quantum states in many aspects with inherent randomness to be used as a QRNG source. The quantum states were usually described by the fock states and the coherent states [105-106]. The fock states contained numbers of photons sharing the same mode such as frequency, polarization, common

path, temporal profile and so on. The coherent states could be considered as the superposition of number states, which shared the properties with the classical light. The representative examples of the optical QRNGs were classified by the physical principles in the following parts.

a) Beam splitting

The beam splitting method was based on the single photon arriving at the beam splitter and passing with either of two different channels [107]. The beam splitter had equal transmissivity and reflectivity to provide equal probabilities for the two different paths. Two detectors were allocated in the two paths. By clicking the detectors with the arriving photon, the two detectors generated the "0" or "1" random number sequence. One problem for the optical path branching QRNGs was the same as the radioactive decay ones. The detectors needed a certain time to recover the sensitivity to the single phonon. It limited the maximum speed of the QRNGs in the Mbps level [108]. The other one was that the real beam splitter and detectors would cause a certain bias due to the different efficiency of the detector and the coupling ratios [12].

b) Photon counting

A large number of optical QRNGs was based on the photon counting in a certain time interval. The number of total photons that arrived at the detection in a fixed interval was an exponential random variable, which obeyed the Poisson distribution [109]. In order to assign the "0" and "1" bit, the post-processes were required to adjust the mean photon level of the source. In this case, the random number was not directly generated by the quantum effects. It was generated from the post-processes by setting a certain mean photon level to make sure the probability of "0" and "1" bit were almost equal [110]. Then the limitation for the QRNGs based on the photon counting was that it highly relied on the post-processes to get the equal probability of "0" and "1" bit. The detection rate was also limited by the dead time.

c) Time of photon arrival

The method for the QRNGs based on the time of photon arrival was similar to the radioactive decay generators. The detector would receive an average number of the photons from the LED incoherent light source and from the coherent laser source in a short time period respectively in one measurement cycle. Both of the arrival times followed the exponential distributions. The difference between the two arrival times was also an exponential random number. By comparing two of the time difference t_1 and t_2 in two measurement cycles, we could assign $t_1 > t_2$ as "1" bit, and $t_1 < t_2$ as "0" bit [111-112]. It could be found that the precise time tagging was highly recommended. The clock speed to digitizing the time arrivals could limit the precision of the measurement. As the photon detectors were utilized, the dead time could not be ignored.

d) Vaccum fluctuations

The QRNGs based on the quantum vacuum fluctuations was designed with the homodyne detection techniques [113]. The homodyne detector mixed the reference laser with the vacuum state in a balanced beam splitter. Two detectors were allocated in the output of the beam splitter. By subtracting and processing the results from two detectors, the final current output could be digitized to the random numbers. Although the rate of this kind of QRNGs could exceed the Mbps rate of photon detection method and could reach Gbps [114], the main measurement would be dominated by the shot-noise. Then the performance would be affected by the classical thermal noise. In addition, complex post-processes were required to digitize the current out to random numbers.

2.3 RTSs in the silicon devices using as QRNGs source

From the fundamental physics of view, the RTSs fluctuations in the MOSFETs were generated from the single electron or hole trapped or de-trapped by a trap in the oxide layer or at the Si/SiO_2 interfaces [5]. The observed the “high” and “low” state were truly random events, which were dominated by the quantum effects. Principally, if we could adjust the probability of the “high” and “low” states to be equal, and assigned the “high” state to “1” bit, the “low” state to “0” bit, the devices could be a promising QRNG source. Comparing with most popular optical QRNGs, the QRNGs based on RTSs do not need the high power light source and the photon detectors. It avoided the dead time as a common issue for most of the optical QRNGs [12]. In addition, the RTSs could be manipulated by the gate bias to achieve equal probabilities for each state [5]. So that it reserved the methods for calibration and maintenance from the original quantum process. Comparing to the QRNGs based on the photon counting, once the emission source was fabricated, the emission processes could not be controlled. The random numbers could only be generated from the complex post-processes. In this case, all the calibration and maintenance could only be implemented in the post-processes, not in the quantum processes. Moreover, the QRNGs source based on the RTSs was just transistor size. Without any light source and radioactive source, the transistor size device could be convenient to integrate to the electrical circuits, or even into the CPU directly. Comparing to the QRNGs based on the shot noise which suffering from the thermal noise, the QRNGs based on the RTSs could use the voltage bias to restrain the effect from the temperature. At different temperatures, the voltage bias could always manipulate the probability distribution of the “high” and “low” states.

It looks like the RTSs should be a perfect candidate as a QRNG source. However, it was rear to find papers to discuss the QRNGs using the RTS phenomenon. The typical one was proposed 17 years

ago in the International Electron Devices Meeting [115]. It was based on the single-electron transistor with back gate. Although recently researchers also proposed to use the RTS to generate random numbers [116, 148-150], they still based on the conventional one gate CMOS structure. And no commercial products were made based on the RTS till now. The main reason was that the original source of the RTSs was the defect, and the defects were randomly formed in the transistor devices during the fabrication processes. In this case, people need to measure thousands of devices to find the RTSs with the statistical investigation. From the engineering point of view, that means the stabilized RTSs source was not available now. Even the radiation could be utilized to introduce the interface-trap to the transistor devices, the energy level of the traps could not be controlled. In order to use the RTSs as a QRNG source, the RTSs source of certainty was highly demanded.

Based on the basic physical model of the RTSs, we proposed the multi-gate transistor design to form an artificial quantum dot in the channel, shown in Figure 2.2. The artificial quantum dot would work as a defect to trap and de-trap the electron and generate the RTSs. By manipulating the voltage bias of the multi-gate, we could control both the energy level in the quantum dot and the potential barriers of the quantum dot. In this case, the occupation and un-occupation of the electron were highly tunable. The drain current switching between two states could be a source of certainty for the quantum random number generation.

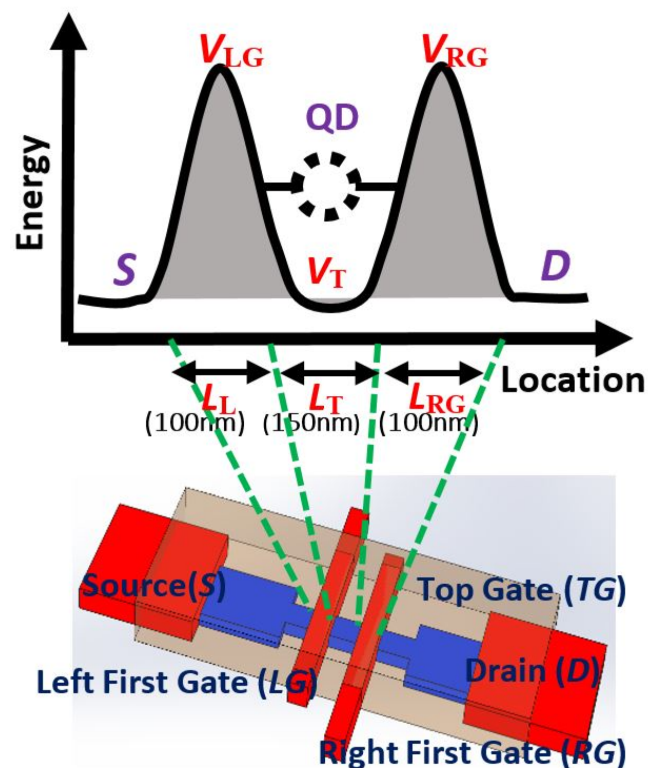


Figure 2.1 The structure of the triple gates silicon nanowire transistor

2.4 Summary

In summary, the phenomenology of the RTS in the MOSFETs was reviewed. We clarified the characterizing method to describe the feature of the RTS. The RTSs in the gate leakage current and in the channel current from source to drain should be distinguished. By reviewing the existing QRNGs, the disadvantages from these devices provided the way to propose the new QRNGs. Based on the standard single-electron transistor model, we proposed our multi-gate transistor design for a QRNG source.

Chapter 3 Random telegraph signals in MOSFET at low temperature

This chapter summarized the experimental RTS behavior in the conventional metal-on-silicon field-effect-transistors (MOSFET) in sub-100nm scale. For the wide-channel device, a pMOS with 10 µm channel width and 75 nm channel length was characterized. For the narrow-channel device, a nMOS with 60 µm channel width and 60 nm channel length was characterized. The key physics and fabrication processes of silicon quantum random number generation devices are based on the conventional MOSFET. The industrial-level MOSFETs were measured using a low noise cryostat to get a better understanding of RTS behavior and investigate the RTS characterization methodology before we fabricated the devices at the University of Southampton.

3.1 Methods

The samples were planar bulk silicon n-type and p-type MOSFET fabricated using 65nm technology by industrial collaborators, which is shown in Figure 3.1. The material of the gate electrode was high-doped polycrystalline silicon (Poly-Si). The gate insulator was made of silicon oxynitride (SiON), in which the equivalent oxide thickness was 2.4nm. The source, drain, gate and bulk electrodes of each transistor were connected with 4 metal pads, which dimension was 100 µm long and 100 µm wide. The pMOS and nMOS which have the same channel dimensions were located up and down next to each other.

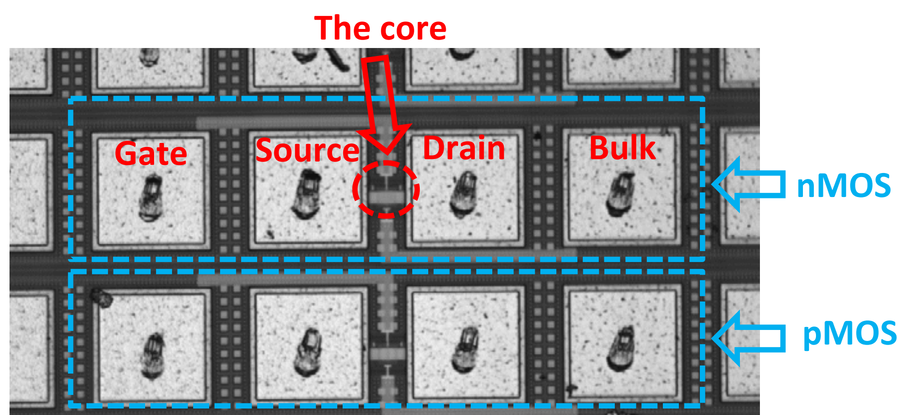


Figure 3.1 Device layout in the optical microscope. The core of the device was located in the middle with 4 metal pads connected. From left to right, they were Gate, Source, Drain and Bulk electrodes.

After setting up the low noise MEMS prober station (Figure 3.2), more than 200 devices including both pMOS and nMOS were measured to confirm the transistor performance at room temperature.

The measurement system was established with the Kelvin measurement method using high-quality triaxial cables. By utilizing the Agilent B1500A Semiconductor Device Analyser to calibrate the system, the system noises were well controlled below 100 fA.

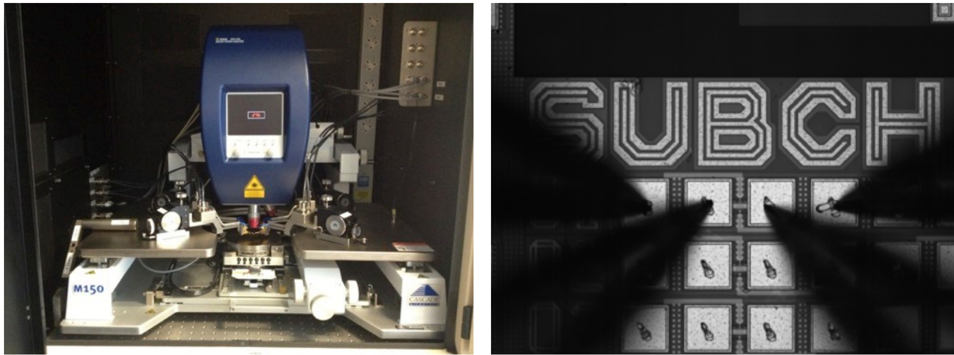


Figure 3.2 (Left) MEMS prober station with all high-quality triaxial cables feedthrough setting. (Right) Cascade highly reliable probe connecting with device.

The standard I_D-V_G and I_D-V_D characterizations were performed on both pMOS and nMOS at room temperature. Superior transistor performance has been achieved to confirm both the measurement setting and the device quality. As an example, an nMOS device (Channel length: 100nm, Channel width: 10μm) measurement results were shown in Figure 3.3-3.4. The drain current was normalized by channel width. From the linear scale of the I_D-V_G diagram, the threshold voltage can be calculated as 0.56V. Moreover, from the log scale of the I_D-V_G diagram, the subthreshold slope was extracted as 106mV/decade.

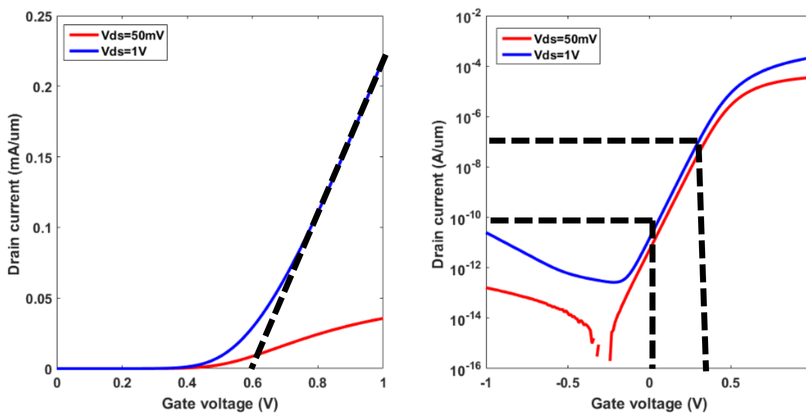


Figure 3.3 Drain current with gate voltage dependence, drain voltage (Vds) was set to 50mV and 1V: (Left) Linear scale to investigate the threshold voltage. (Right) Log scale to investigate the subthreshold slope

.

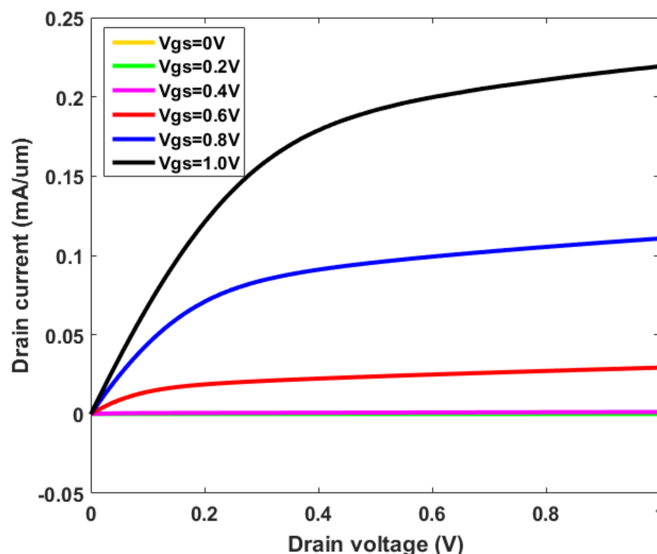


Figure 3.4 Drain current with drain voltage dependence. Gate voltage (Vgs) was set to 0V, 0.2V, 0.4V, 0.6V, 0.8V and 1V

At room temperature, the typical RTSs were mainly dominated by thermal energy. The carrier transitions would not correspond to the elastic tunneling transition, which happened at the energy level around E_f [5]. In addition, these devices were industrial-level quality and high-performance bare transistors. It was very rare to find a poor device which shown the RTS. Only one nMOS device which channel was 60 nm long and 60 nm width showed 2 level oscillation in the IdVd curve. Then, it was soon identified to be a very poor device in which the gate oxide was broken easily in the next standard measurement. The time-domain measurements of a selected good device were also implemented, which showed good transistor performance in I_D-V_G characterizations (Figure 3.5). In the subthreshold region, six points (Red diamonds in Figure 3.5) were selected to do the time domain measurements, which covered the region of the I_D from 1 pA to 100 nA. A certain V_G applied with V_D was 50 mV. 1001 points were measured in the same condition. We cannot see any typical RTSs phenomenon and mainly 1/f noise (Figure 3.6 a-f), and all the probability distribution of the I_D showed only one peak in each of them (Figure 3.6 g-l).

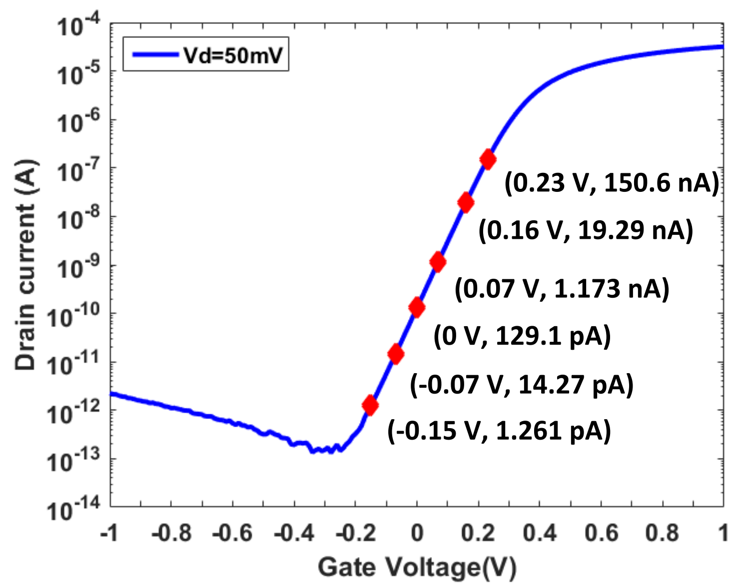


Figure 3.5 Drain current with gate voltage dependence. Drain voltage (V_D) was set to 50mV. The Sex selected points were marked as red diamonds, which represented the I_D in 1 pA, 10 pA, 100 pA, 1 nA, 10 nA, and 100 nA levels

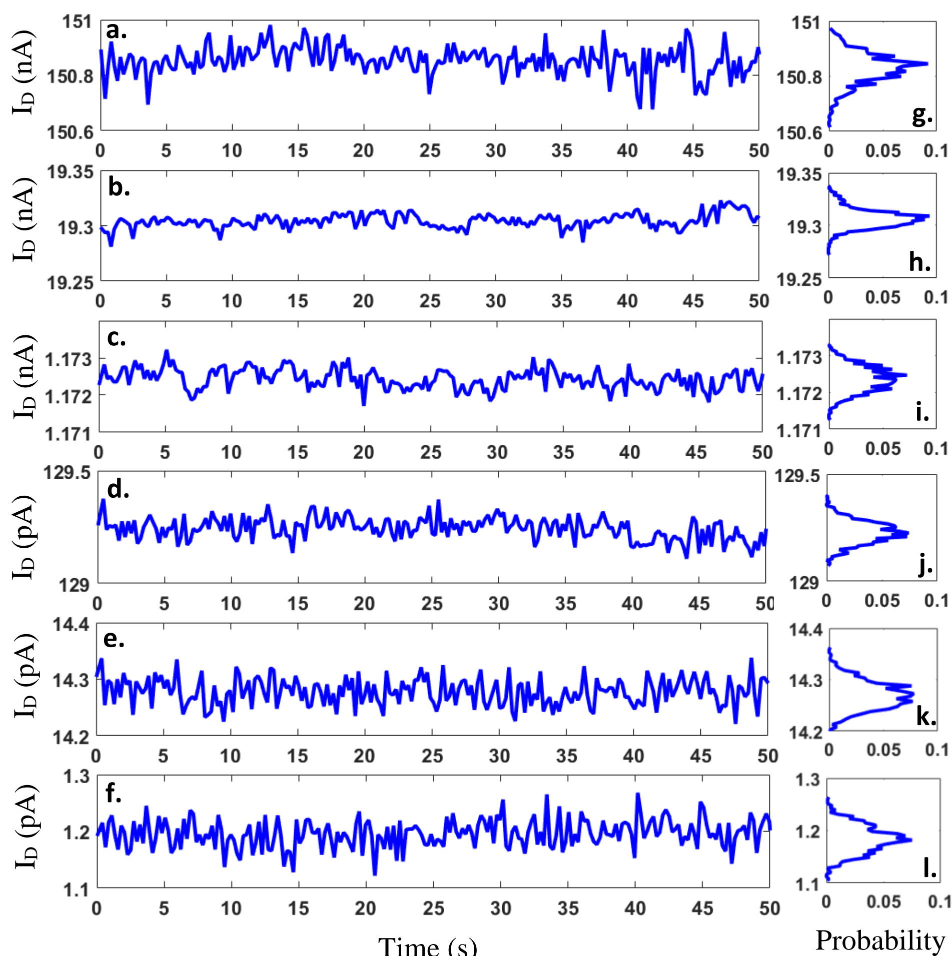


Figure 3.6 Drain current with gate voltage dependence. Drain voltage (V_D) was set to 50mV. The Sex selected points were marked as red diamonds, which represented the I_D in 1 pA, 10 pA, 100 pA, 1 nA, 10 nA and 100 nA levels. (a-f)The time domain measurement results with the first 50s. (g-l)The probability distribution of the I_D.

However, at low temperatures, the RTSs did not rely on the thermally activated tunneling transition [42, 43, 117]. Without the help of thermal energy for the charge transition, typical RTSs could be more likely to be found, and the amplitude of RTSs would be much larger than that at the higher temperature [42]. In this case, the main characterization of RTSs in the standard MOSFETs was measured in the cryostat at low temperatures.

3.2 Low-temperature measurement system and sample packaging

The low-temperature measurement system was the Cryogen free high field measurement system from the Cryogenic Ltd. Company, which enables us to perform the experiment down to 300 mili-Kelvin (Figure 3.7). As it was a quite common system with a detailed user manual, the details of the system will not be repeated here. The basic working principles will be briefly described.

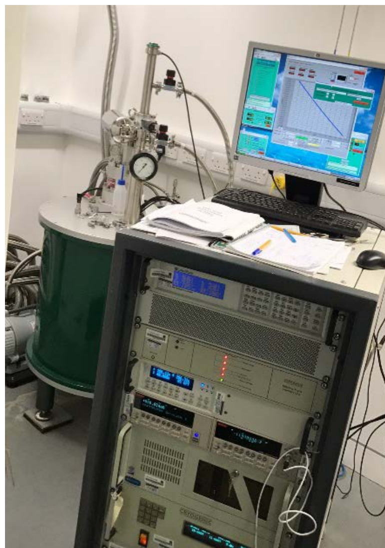


Figure 3.7 The cryogen-free high field measurement system

As most "Dry" cryogenic systems, there was no need to feed liquid Helium to the system during the measurement. This was the most outstanding advantage of the "Dry" cryogenic systems comparing to the "Wet" cryogenic systems, which consume the liquid Helium during the measurement. So that the "Dry" cryogenic systems almost had no time limitation for the measurement. For the RTS investigation, thousands of points need to be measured. The "Dry" cryogenic systems could be satisfied with this kind of long-time measurement. This cryogen-free high field measurement system was mainly comprised of the Cryo-cooler system, the Cryostat and magnet, and the Variable temperature insert (VTI) (Figure 3.8). To reach the lowest temperature, the cryo-cooler system was operated firstly. The helium compressor outputted the main power to cool down the cryostat and the magnet by an independent high-pressure helium circuit, which was isolated to the other parts of the system. The cryostat provided an isolated vacuum chamber to shield and support the VTI and superconducting magnet thermally. With the cooling power from the helium compressor, the VTI could be cooled down to 160 K. In this project, the magnet would not be used, so that it was ignored. The VTI part would mainly dominate the sample temperature from 1.6 K to 325 K. A bore form by the magnet surrounded and supported the VTI. Apart from the helium compressor cooling system, the VTI had another independent helium circulation cooling system. The "Helium Dump" was used to store the helium gas for the VTI helium circulation at the room temperature. After the VTI was cooled down below 160 K by the Cryo-cooler system, the "Oil-free pump" should be started to

continue cooling. The “Oil-free pump” pushed the helium gas out of the dump and came into the circulation from the “Helium Gas Inlet”. Then the helium passed through the “Charcoal Filter” to get impurities removed, and made the heat exchanger in the “40 K stage” to be cooled down to 40 K. In the following “4 K stage”, the helium gas was cooled further to below 4.2 K and finally condensed as helium liquid in the “He pot”. By controlling the “Needle Valve” to let the helium liquid flow into the sample chamber, the helium liquid expanded and cooled the sample chamber further to 1.6 K. Finally, the expanded helium gas flowed up to the top of the VTI and extracted by the “Oil-free pump” to get back to the “Helium Dump”. With this independent circulation cooling system, the sample chamber could be stable at 1.6 K to 325 K for a long time until the pump and compressor were shut down. In this case, the system could provide a stable temperature for an enough long time to characterize RTSs phenomena.

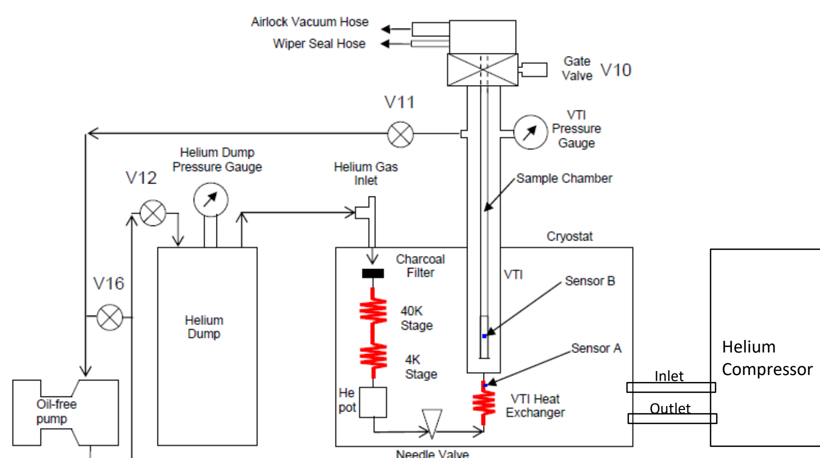


Figure 3.8 The schematic structure of the cryogen-free high field measurement system

With the special sorption pumped He-3 insert, the sample could be further cooled down to 300 milli-Kelvin (Figure 3.9). The He-3 insert included a “Main sorption pump” to decrease the pressure of the He-3 vapour. The sorption pump played the role of a He-3 gas absorber when it was lower than 40 K. When the sample temperature stayed around 1.6 K, we turned on the heater to increase the temperature of the sorption pump to 40 K, all the He-3 gas would release from the absorber and condense to liquid in the “2cc helium-3 pot” at the bottom of the insert. Then we turned off the heater. The He-3 vapor would begin to condense on the absorber inversely. In this case, the pressure in the “2cc helium-3 pot” would be decreased and the temperature inside would be finally decreased to 300 mK, until all the He-3 liquid evaporated.

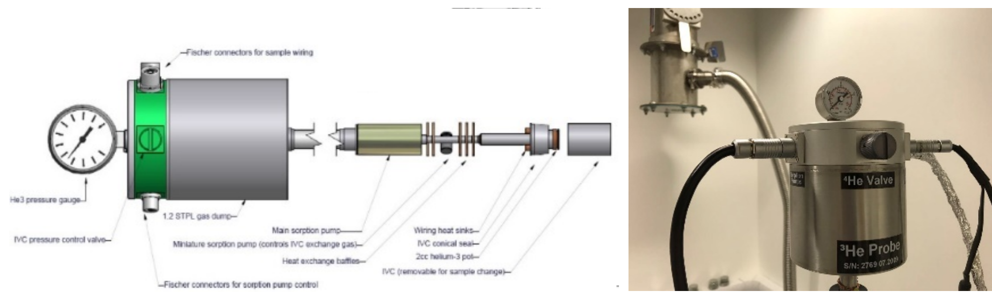


Figure 3.9 (Left) The schematic structure of the He-3 insert. (Right) He-3 insert with electrical connections

However, this He-3 insert had not been used and maintained in the past 7 years. When it was restarted to use, the lowest temperature (300 mK) could only be stable in 40 minutes (Figure 3.10). The time was too short for the RTSs measurements. Moreover, both the provided devices and the fabricated devices were aluminum pads. Below 1.2 K, the aluminum pads would become superconductors. The results could be more complicated and beyond the main aim for this project. So that the measurements were focused at the temperature from 2 K to room temperature. The measurements were also based on the Agilent B1500A Semiconductor Device Analyser.

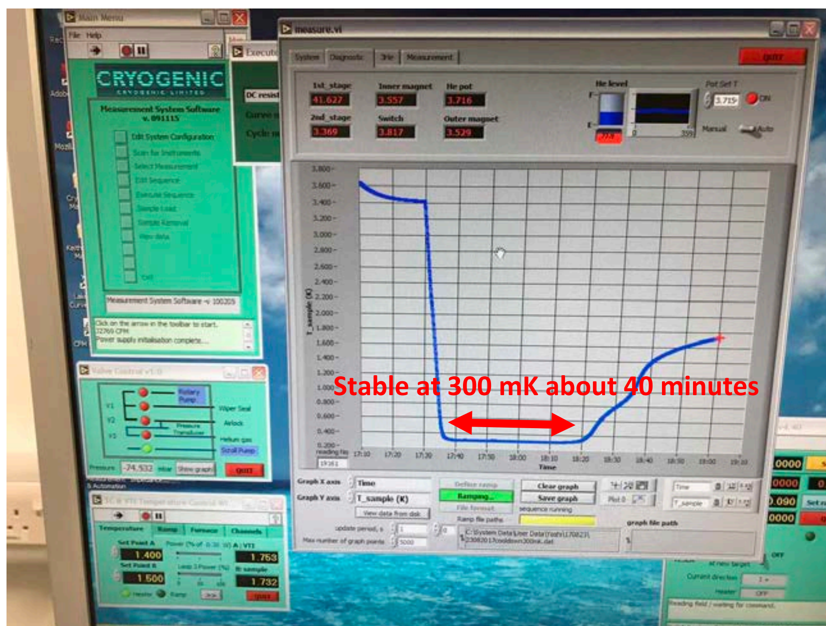


Figure 3.10 The performance of the He-3 insert. The lowest temperature was 298 mK. It can be stable around 40 minutes from 17:37 to 18:18.

Before loading to the cryogenic system, the devices should be diced and packaged to fit the loading insert of the cryostat. The samples need to be cut into 3mm x 5mm small chips by the dicing machine (Figure 3.11). This size was limited by the chip carriers, which can be loaded into the cryostat. To protect the sample surfaces, a photoresist (S1813) layer was coated before dicing. After dicing, Acetone followed by Isopropyl Alcohol (IPA) was used to remove the photoresist layer. After the De-ionized water-raising, the nitrogen gun was used to dry the sample.

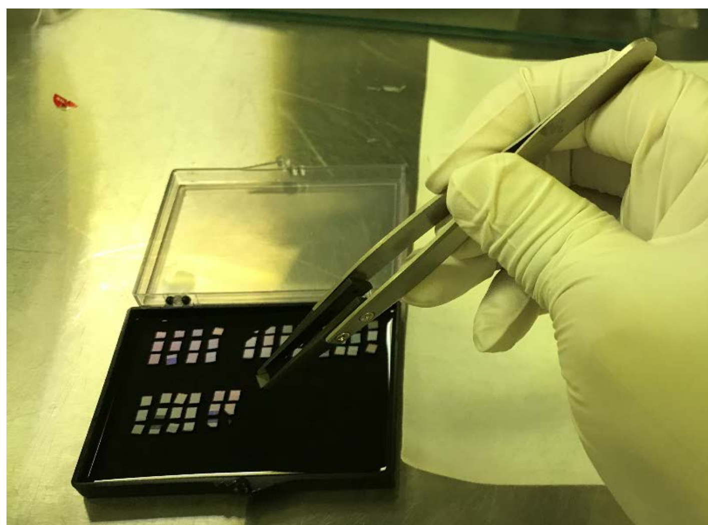


Figure 3.11 Maximum 3mm 5mm small chips after dicing. The cleaned samples were stored in a jelly sample box.*

The special 6-pin chip carrier was used to mount the sample with a non-conductive cryogenic epoxy (Figure 3.12). Then the device contact pads needed to be bonded to the pins on the chip carrier. As only 6 pins were on one chip carrier, and one device had 4 metal pads, we could only connect one device to the chip carrier once time. Besides, the wire bonding process was quite challenging due to the diameter of the gold wire was similar to the size of the metal pad device. Moreover, the distance between the two metal pads was quite closed. The bonding points were very easy to touch with each other during wire bonding and caused the failure. It could be seen in Figure 3.13 which was from a successfully bonded device. In this case, we prepared several samples that were mounted to the chip carrier before wire bonding. Finally, the chip carrier was mounted to the bottom of the insert probe that could be loaded into the cryostat.

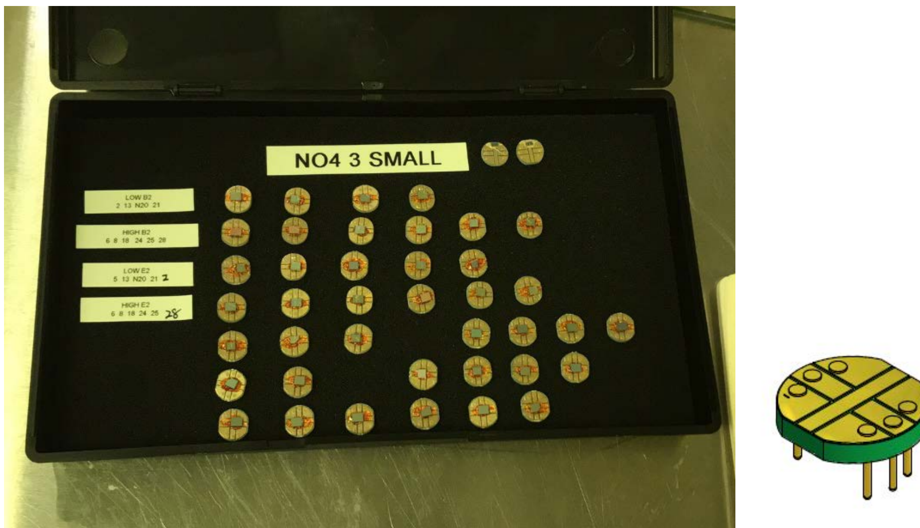


Figure 3.12 The MOSFETs samples mounted to a 6-pin chip carrier with a yellow no-conductive epoxy.

Figure 3.13 The optical microscope view of the sample which was mounted and successfully bonded to a 6-pin chip carrier with 4 gold wires.

3.3 RTSs in a wide-channel MOSFET (W/L=10um/75nm)

3.3.1 Measurement

The first successfully bonded device was a pMOSFET with 10 μm channel width and 75 nm channel length. It was loaded into the cryostat and cooled down to 2 K. The I_DV_G characteristics were measured at different temperatures. In Figure 3.14, the transistor performance was shown at three typical temperatures (300 K, 150K and 2K).

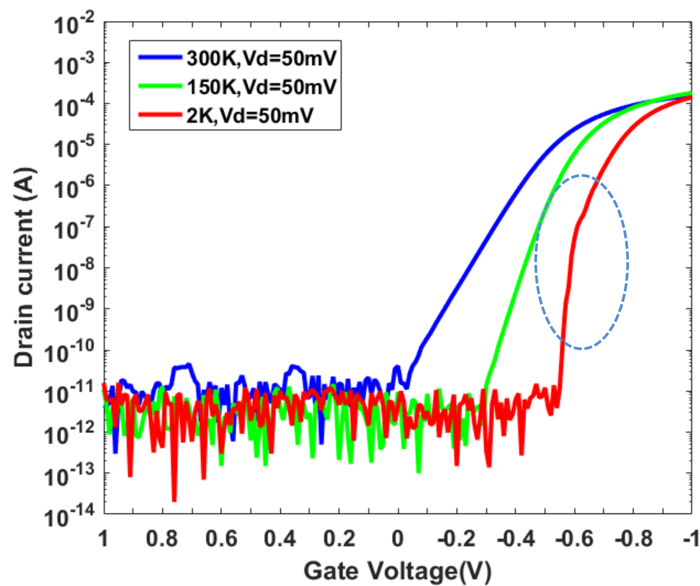


Figure 3.14 Drain current with gate voltage dependence for the pMOSFET (10 μm long and 75 nm wide). V_D was set to 50mV to make a stable measurement temperature.

Because this was a wide-channel pMOSFET, the ON current (V_G=1 V and V_D=1 V) was in the milli-ampere region at room temperature. Moreover, the effective mobility would be increased at low temperature [118]. The ON current would be increased further. To avoid the heat generated by the transistor affecting the measurement temperature, we kept a small V_D (50 mV) to measure the transistor performance in a low field transport regime. The sign of the Coulomb blockade phenomenon had appeared at 2 K in such a large area (wide but narrow) pMOSFET. Although finding the Coulomb blockade phenomenon was not the main aim of this project, it could provide a single electron/hole transport system built in the devices. As the RTSs phenomenon was related to single electron/hole trapping and de-trapping from a defect in the device, to find the Coulomb diamond could provide us an energy level reference of the device, which could be used to voltage parameters to investigate the RTSs phenomenon. According to the results in Figure 3.14, the channel resistance of the transistor was larger than 25.8 kΩ in the sub-threshold region from -0.55 V to -0.7 V, which was the essential condition to observe the Coulomb diamond. In this case, we

measured the drain current by sweeping the V_G from -0.55 V to -0.7 V and the V_D from -20 mV to 20 mV at 2 K.

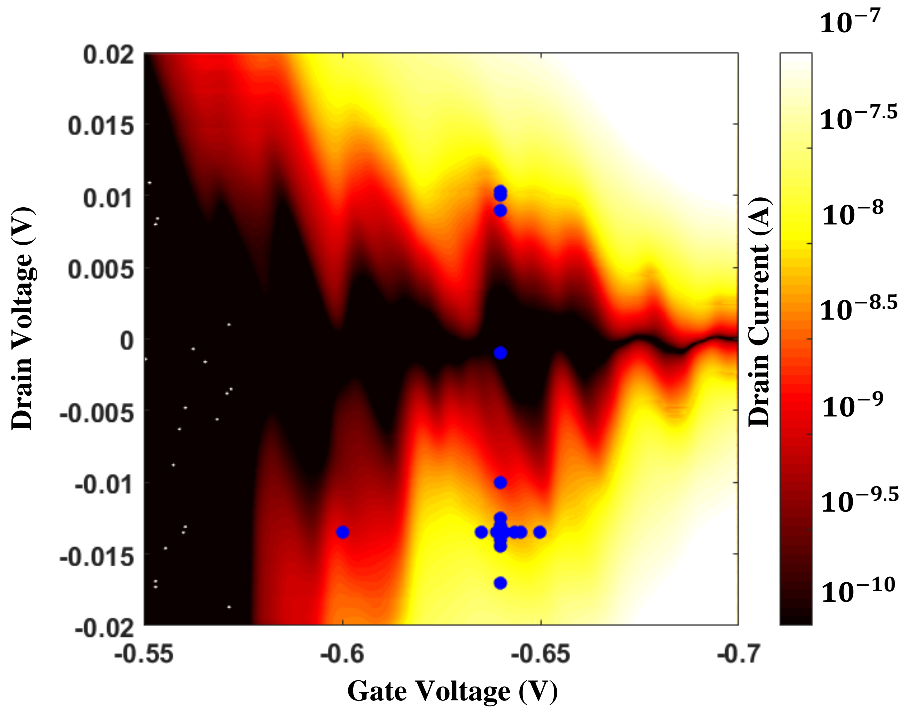


Figure 3.15 The contour plot for drain current versus gate voltage and drain voltage of the pMOSFET at 2K. The bluepoints were selected conditions to extract the gate voltage and drain voltage. Then the RTSs time-domain measurement was carried on the voltage conditions of these points. The results were shown in Figure 3.18-3.20 and 3.22-3.24.

The contour plot of the drain current was shown in Figure 3.15. It can be clearly seen that several Coulomb diamonds continuously appeared during the gate voltage sweep. The Coulomb diamond sizes were shrunk generally while increasing the $|V_G|$. Different Coulomb diamonds represented different quantum states in the single hole transport system. Thus different size of the Coulomb diamond was related to different charging energy of the different quantum states. It indicated that the changing of $|V_G|$ affected the coupling capacitance of the quantum dot. This phenomenon was quite different from the metallic tunneling junction mesoscopic model, in which the Coulomb diamond sizes were all the same. To understand the size decreasing Coulomb diamond, we need to investigate how the quantum dot was formed which would be discussed later. Firstly, we focused on the edges of the Coulomb diamond, where some horizontal peaks can be found. If we changed a better colour for contrast, the peaks were more clear (Figure 3.16). That means the current did not change continuously. The small step change of the current must happened in that area. One I_DV_G curve that crossed the “Area 1” was extracted from the contour plot data (Figure 3.17). The step-change could be confirmed.

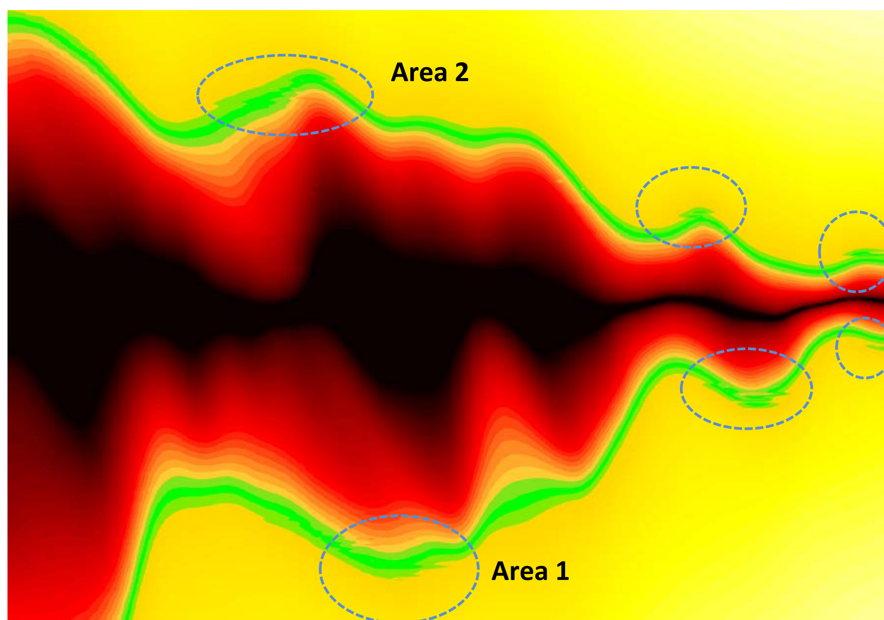


Figure 3.16 Horizontal peaks happened at the edges of the Coulomb diamond areas

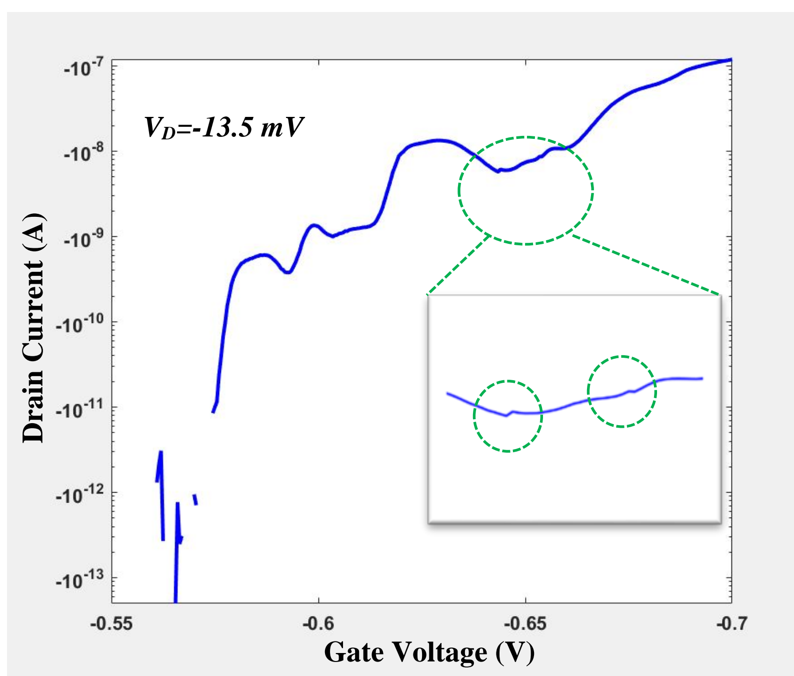


Figure 3.17 One I_DV_G curve (V_D=-13.5 mV) extracted from Figure 3.15. The step changes of the current can be found around V_G=-0.65 V (In the Area 1 of Figure 3.16).

Then the time-domain characteristics of I_D were implemented with the gate and the drain bias conditions located in the “Area 1” of Figure 3.16. The typical RTSs were finally observed in a wide and short pMOSFET at 2 K with the assist of Coulomb diamonds. One example of the typical result was shown in Figure 3. 18.

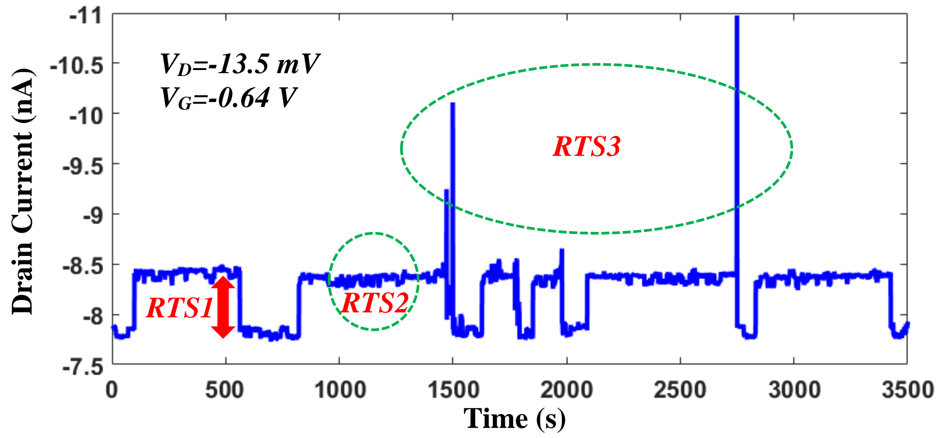


Figure 3.18 The time-domain measurement of I_D with the bias condition of V_D=-13.5 mV and V_G=-0.64 V. Three types of RTSs were found.

In Figure 3.18, 3500 points were collected within the time interval Δt=1 s. We can clearly find three kinds of RTSs that happened in this period. The RTS3 has the largest amplitude and happened quite rear, which would cause a large current peak and might damage the device. The RTS2 has the smallest amplitude and shortest average switching time, which had a limited effect on the total drain current. The RTS1 has an amplitude about 0.5 nA, and was switching in a reasonable time interval. So that the current step change in Figure 3.17 was mainly caused by the RTS1. As the RTSs were caused by a single electron or hole trapping and de-trapping from a trap (or defect), the time duration that the trap captured the hole or emitted the hole revealed the wave function of the trap. In quantum mechanics, it was described as the probability distribution of finding a particle that occupied a certain energy state. The occupied or unoccupied condition was reflected in the I_D in this case. So that the probability distribution of I_D ($P(I_D)$) that had a relationship with the wave function of the traps ($\varphi(I_D)$) is:

$$P(I_D) = |\varphi(I_D)|^2 \tag{3.1}$$

In Figure 3.19, the $P(I_D)$ was extracted from the data in Figure 3.18. The low state of RTS1 and the high state of RTS1 can be distinguished from the two main separated peaks. In each of the main peaks, it included two sub-peaks, which were related to the low and high state of RTS2 respectively. The RTS3 was very rare to happen so that the probability was almost zero. We could also find the probability distribution of RTS2 which had a high co-relationship with that of RTS1. When the RTS1 was in the high state, RTS2 had a high probability in the high state. When The RTS1 was in the low state, the probabilities of the RTS2 in the high and low state were almost equal. That means the two quantum states in the two traps, which caused the RTS1 and the RTS2, had a strong coupling relationship.

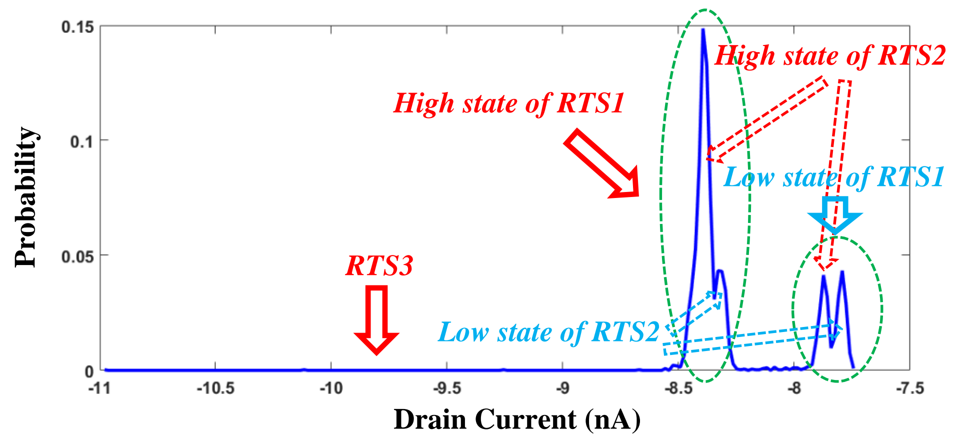
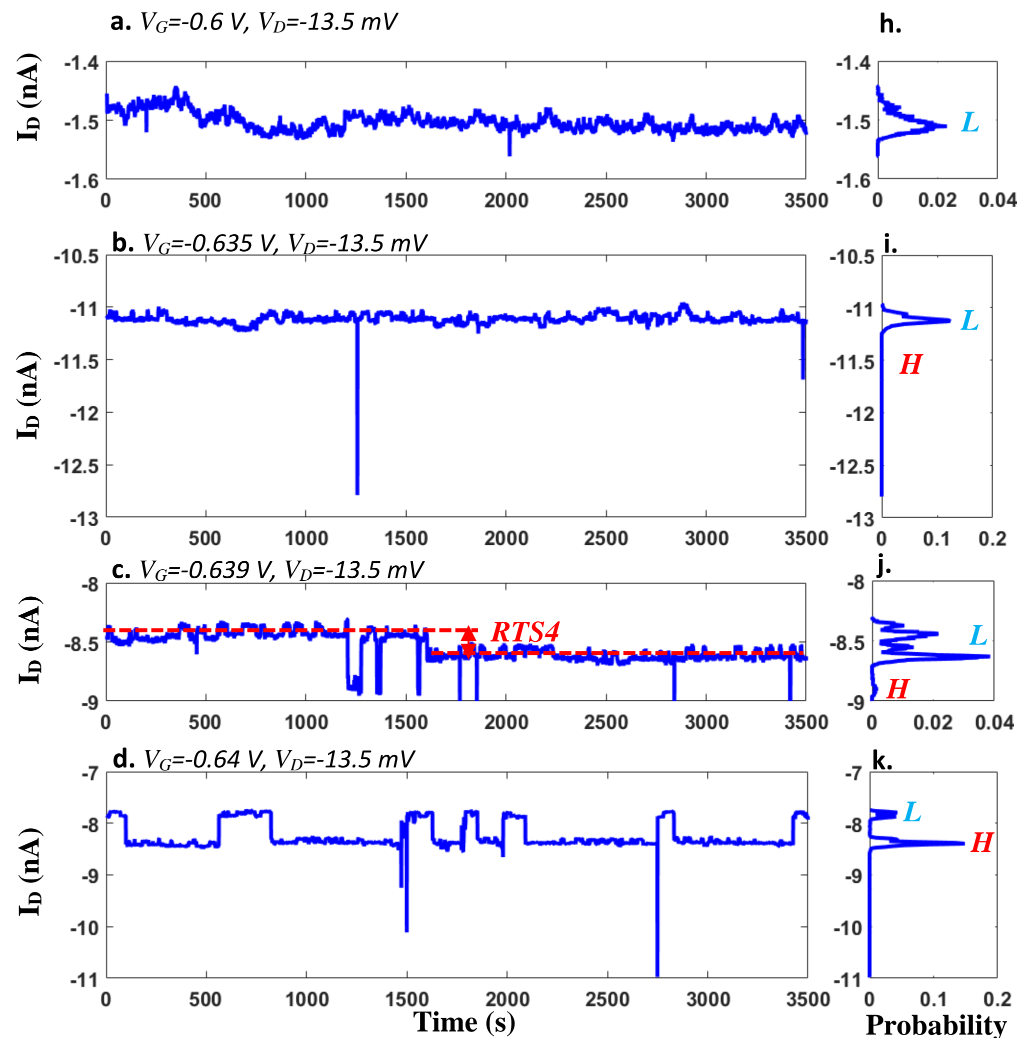


Figure 3.18 The probability distribution of the current versus its valve

In order to investigate the RTSs systematically, the time-domain measurement of the I_D was continued in the selected bias conditions that were marked as blue points in Figure 3.15. The bias condition with V_G=-0.64 V and V_D=-13.5 mV was focused as a centre point. By changing the V_G and V_D bias respectively, the dependence of the RTSs probability distribution on V_G and V_D could be illustrated. Firstly, the V_G bias was selected as -0.6 V, -0.635 V, -0.639 V, -0.64 V, -0.641 V, -0.645 V and -0.65 V with V_D=-13.5 mV as a constant. The results were shown in Figure 3.19 a-g.



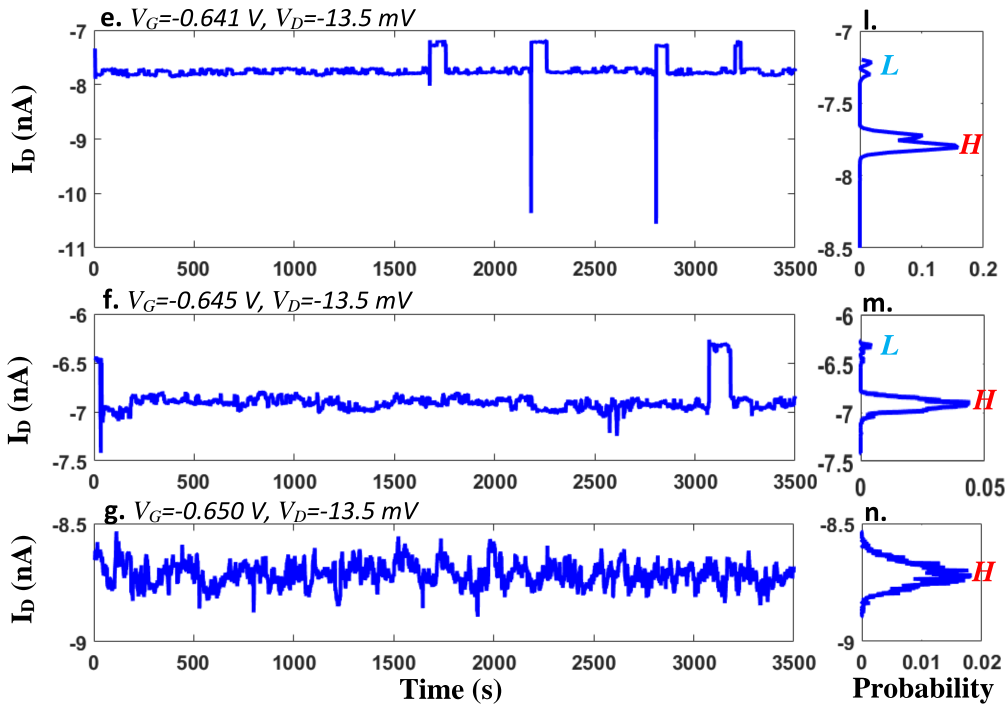


Figure 3.19 (a-g) The time-domain measurement results of the V_G dependence for the RTSs. (h-n) The probability distribution of the current versus its valve. The high state of RTS1 was marked as red "H"; the low state of RTS1 was marked as blue "L".

The corresponding $P(I_D)$ was extracted from the time-domain measurement results at each V_G bias respectively (Figure 3.19 h-n). It could be observed that the probability to find the current in a high state was increased by changing the V_G from -0.6 V to -0.65 V and this was mainly dominated by RTS1. Although RTS3 had been observed with huge amplitude and could be found in Figure 3.19 b, d, and e, it was still quite rare and was almost zero in the probability distribution calculation in Figure 3.19 i, k, and l. The RTS2 could always happen in both the high and low state of RTS1. On most occasions, the RTS2 were more likely to state at the high current level. However, when the bias condition made the RTS1 more likely to be in a high state, in the RTS1 low current state part, the RTS2 was more likely to be in a low state. This implied that the traps which caused the RTS1 and RTS2 were quite closed and the RTS1 could affect the RTS2 to some extent. In Figure 3.19 (c), another kind of RTSs was observed, which we named RTS4. The amplitude of the RTS4 was estimated between RTS1 and RTS2. The capture and emission time of RTS4 was quite long, which seemed to be more than 1 hour. To avoid the interference for the RTS4, the data in Figure 3.19 (c) was divided into 2 parts, the RTS4 low state, and the RTS4 high state, shown in Figure 3.20.

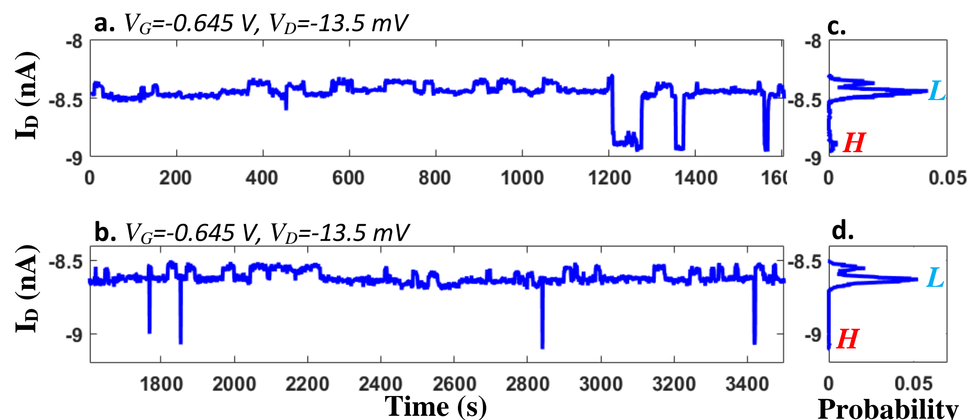


Figure 3.20 The time-domain measurement results in Figure 3.19 (c) were divided into (a) (from 0 to 1605s) and (b) (from 1605s to 3500s).of the V_G dependence for the RTSs. (c,d) The probability distribution of the current versus its valve. The high state of RTS1 was marked as red "H"; the low state of RTS1 was marked as blue "L".

As the experience from the measurement, if the wires were touched gently during the measurement, the current would have a high peak immediately. In addition, in the measurement room, it was vented all the time and was very crowded with noisy compressors and vacuum pumps. There was no guarantee that no vibration was transferred to the single wires. Moreover, between the Agilent B1500 semiconductor analyzer and the insert probe, there was a "break-out box" (Figure 3.21) to connect them together. The Agilent B1500 triaxle cables were terminated here by the triaxle to the banana adaptor. Only the single could pass without guard and shield layers. This "break-out box" was designed for the insert probe connecting with measurement equipment. Also, it could be used to test the sample performance after wire bonding before loading into the cryostat as shown in Figure 3.20. Because there was no guard and shield for the single in the "break-out box", the stochastic environmental noise could directly affect the signals, which was the limitation of the system. Although some actions had been done to reduce the environmental noise, it could not avoid the noise interruption from the surrounding. This was the reason why I upgraded the system into a better noise level condition after finishing the measurement for this device. The system upgrading was shown in chapter 3.4.1 later. In this case, the RTS3 would be treated as system noise and ignored for analyzing.

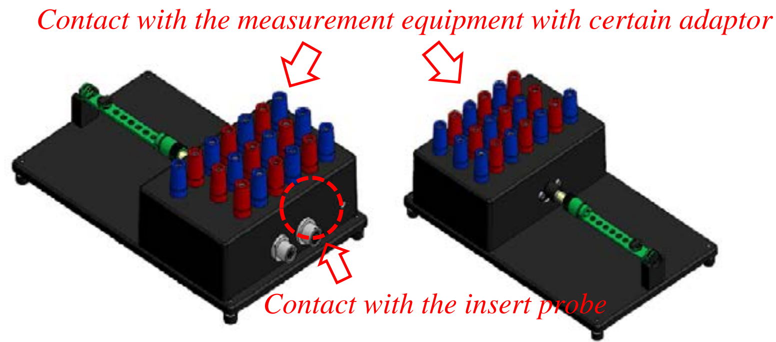
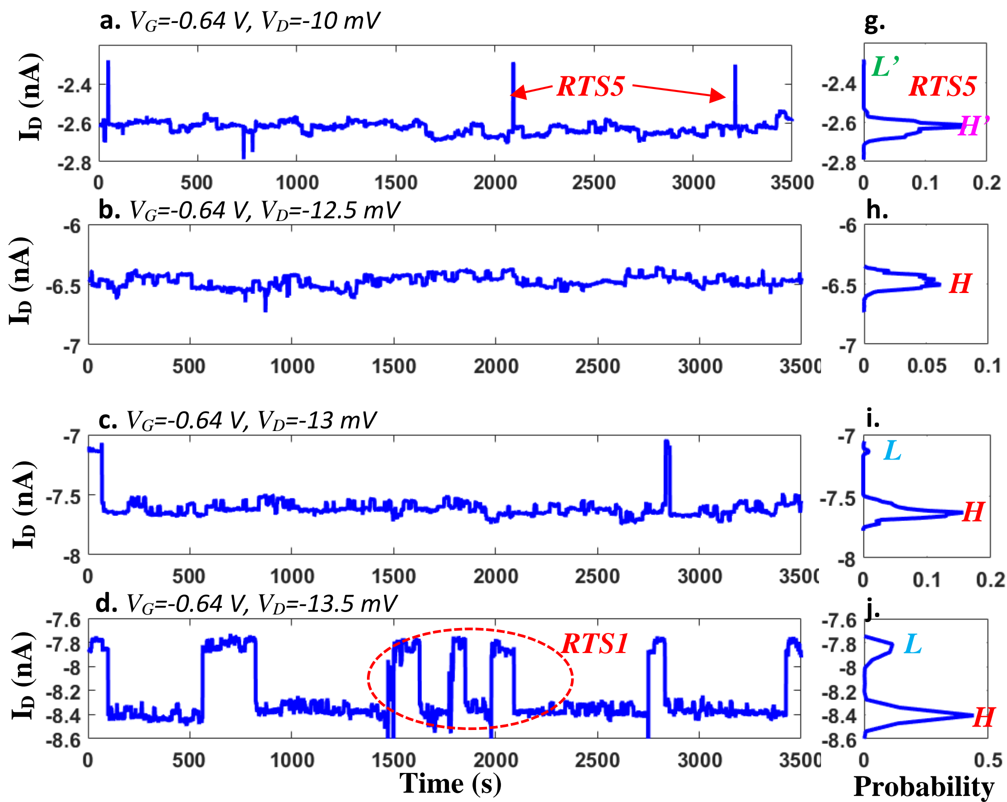


Figure 3.21 The "break-out box" to make the connect between measurement equipment and insert probe. The green tube was the bottom part of the inset probe, where the chip carrier mounted on. The "break-out box" could directly measure the sample before loading into the cryostat.

As the aim of measuring the RTSs in silicon devices in this project was to use the RTSs as a random number generation source. The RTSs that dominated the amplitude of the I_D was mainly focused. In this case, the RTS1 was focused on investigating the bias dependence of the RTSs. After the V_G bias dependence measurement, the V_D bias was selected as -10 mV, -12.5 mV, -13 mV, -13.5 mV, -14 mV and -14.5 mV with V_G=-640 mV as a constant. The results were shown in Figure 3.22 a-g.



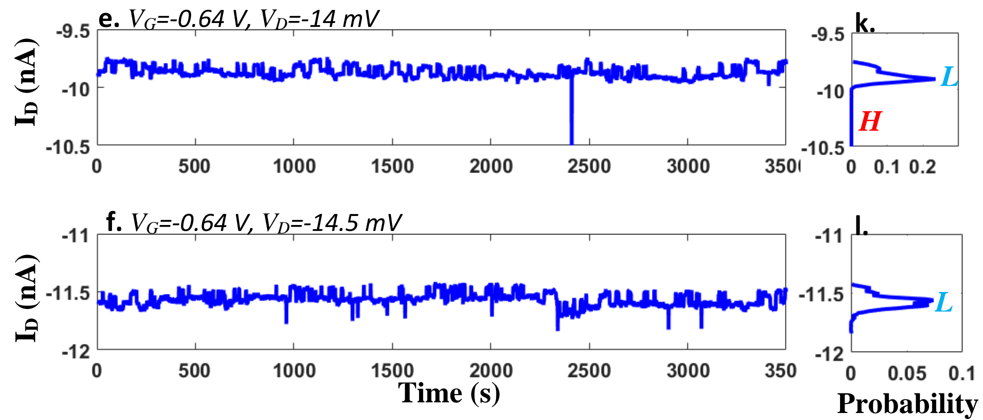


Figure 3.22 (b-f) The time-domain measurement results of the V_D dependence for the RTSs. (h-l) The probability distribution of the current versus its valve. The high state of RTS1 was marked as red "H"; the low state of RTS1 was marked as blue "L". (a,g) The RTS5 appeared. The high state of the RTS5 was marked as red "H'"; the low state of the RTS5 was marked as blue "L'".

It can be observed that the probability distribution of I_D for RTS1 was shifted from being more likely in a high state to being more likely in a low state during the V_D changing from -12.5 mV to -14.5 mV. However, if the V_D was decreased to -10 mV, another new RTS appeared which was named RTS5. The RTS amplitude was used to distinguish the RTS5 from the RTS1. It could be clearly found that the amplitude of RTS5 was smaller than 0.4 nA approximately from Figure 3.22 (a). However, the amplitude of RTS1 was more than 0.6 nA approximately. That means another RTS would dominate the I_D if we continued to decrease the V_D. Because we could also observe the horizontal peek on the top of the same Coulomb diamond (Area2) in Figure 3. 16, three points (V_D=9 mV, V_D=10 mV and V_D=10.3 mV) were selected to do the time-domain measurement with the V_G=-640 mV (Figure 3.23). It could be used to analyze the electrical symmetric of the trap.

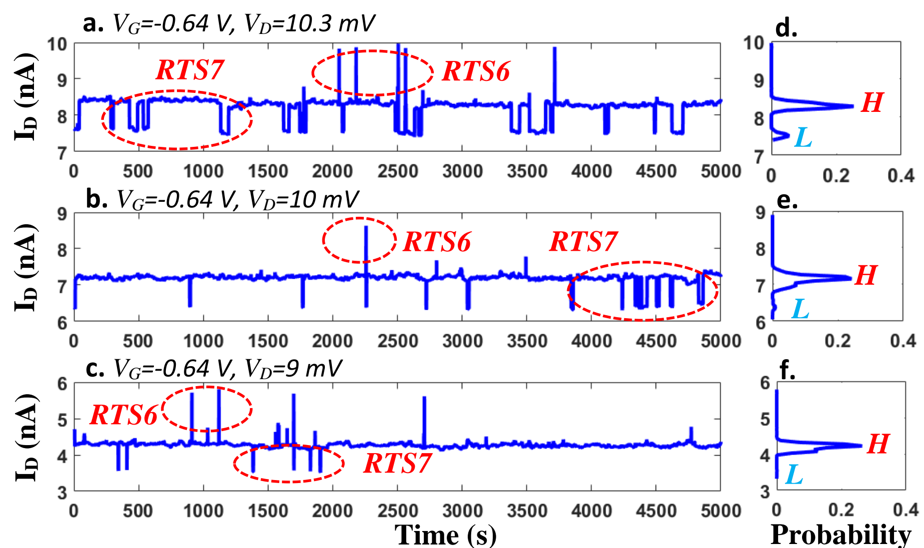


Figure 3.23 (a-c) The time-domain measurement results of the V_D dependence for the RTSs. (d-f) The probability distribution of the current versus its valve.

The RTSs were also observed on the top edge of the same coulomb diamond with RTS1, which were named RTS6 and RTS7. The amplitude of RTS7 was estimated to be about 0.7 nA, which was quite similar to RTS1. So that it might be caused by the same trap. The amplitude of RTS6 was estimated at about 1.5 nA. Therefore, it was more likely to activate a new trap at a different energy level. However, it was still too rare to be observed, RTS6 could not be used in statistic analysis. To check the bias conditions that located in the Coulomb diamonds as a reference, the measurement was carried out on V_D=-1 mV and V_G=-640 mV (Figure 3. 24). In the Coulomb diamond, the current was almost blocked and less than 1 nA. It was unable to analyze the RTSs.

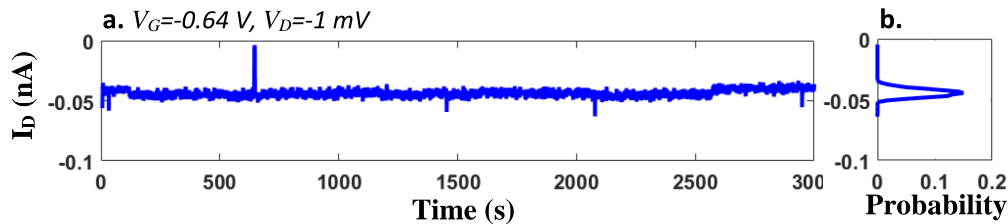


Figure 3.24 The time-domain measurement results in the Coulomb diamonds.

So far, we had observed 7 RTSs in this pMOS at 2K. If we continued to do the same measurement in other areas which showed the horizontal peaks in Figure 3.16, other kinds of RTSs might be observed. As the main target of these measurements was to establish the method to find RTSs in a silicon transistor and to investigate how the RTSs generated, there was no necessity to measure all possible RTSs at low temperatures. The discussion would focus on RTS1 to simplify the analysis.

3.3.2 Discussion

Because we found the RTSs with the assistance from the Coulomb diamonds, how these Coulomb diamonds were formed in a standard pMOS should be firstly investigated. The Coulomb diamonds marked in Figure 3.15 were shown in Figure 3.25. The threshold voltage was extracted from Figure 3.14 as -0.67 V on V_D=-50 mV at 2 K.

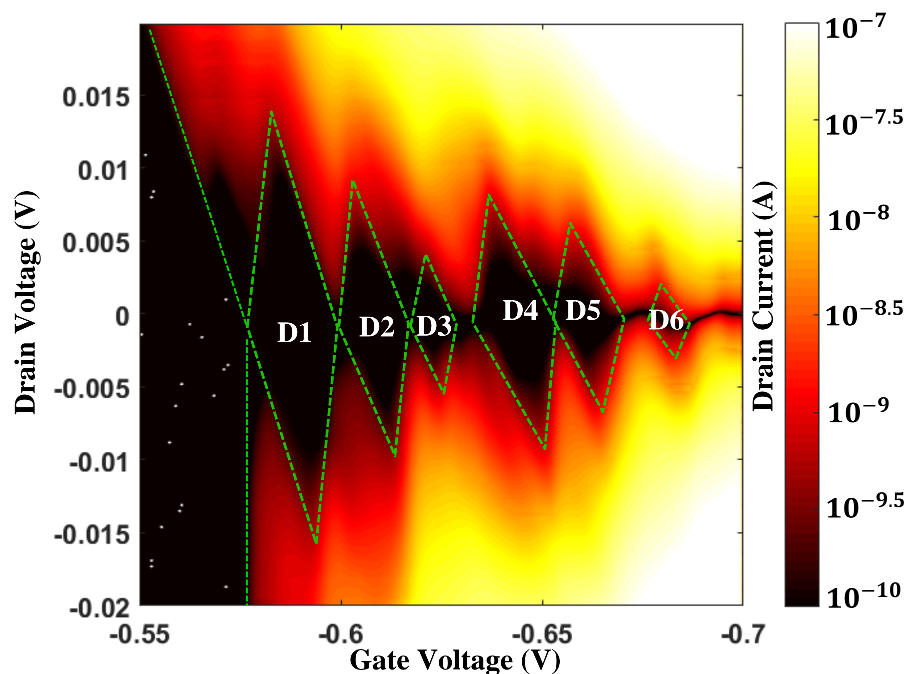


Figure 3.25 The contour plot for drain current versus gate voltage and drain voltage of the pMOSFET at 2K.

Six typical Coulomb diamonds were marked *D1*, *D2*, *D3*, *D4*, *D5*, and *D6*, and it can be observed that there were more Coulomb diamonds between them. It implied that the device was dominated by several quantum dots compared to the single dot in a silicon device [119]. The Coulomb diamonds sizes were decreasing from left to right as we mentioned before. In addition, some smaller Coulomb diamonds appeared among the six diamonds. Based on the carrier transport properties in the pMOS [120], the gate capacitance would be increased by sweeping $|V_G|$ to a larger value in the sub-threshold region due to the carrier concentration was increased. In this case, if the circuit model in this device was fitted to the standard mesoscopic model in Figure 3.26, all the coupling capacitances including gate coupling capacitance (C_G), source coupling capacitance (C_S) and drain capacitance (C_D) to the quantum would not increase [121]. In other words, all these coupling capacitances were related to channel inversion layer capacitance. After the threshold voltage, no diamond was found due to the channel was opened and the barrier resistances were smaller than the quantum resistance. As the RTSs measurements were based on *D4*, the *D4*

diamond was used to estimate the parameters for this single-hole transistor based on the mesoscopic model at that bias condition.

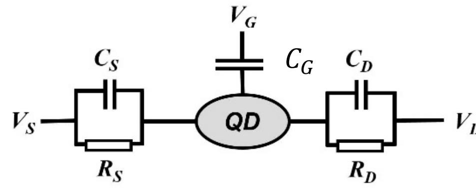


Figure 3.26 The equivalent circuit model for the single hole transistor

Supposing the Coulomb oscillation was periodic, the C_G could be extracted according to the gate voltage changing from the certain Coulomb diamond *D4* [120, 121].

$$C_G = \frac{e}{\Delta V_G} \tag{3.2}$$

where e was the elementary electric charge, which was $1.602176634 \times 10^{-19}$ C, and ΔV_G was the V_G changing from the vertex on the left side of Coulomb diamond *D4* to the vertex on the right side. If the quantum dot was assumed as a circle, the diameter (d) of the quantum dot could be estimated as:

$$d = \sqrt{\frac{4C_G t_{eff}}{\pi \varepsilon_0 \varepsilon_{SiO_2}}} \tag{3.3}$$

where ε_{SiO_2} was the dielectric constant of SiO_2, t_{eff} was the total effective thickness related to the total gate capacitance and t_{eff} should include not only the equivalent oxide thickness (t_{ox}) of the $SiON$ layer which was 2.4 nm, but also the additional thickness from inversion layer thickness (t_{inv}) near the $Si/SiON$ interface (about 2 nm) [121-124].

$$t_{eff} = \frac{\varepsilon_{SiO_2}}{\varepsilon_{Si}} t_{inv} + t_{ox} \tag{3.4}$$

Then t_{eff} was calculated as 3.1 nm and C_G could be estimated as 8.43 aF from Figure 3.25. Finally, d of the quantum dot related to the Coulomb diamond *D4* could be estimated at 31.9 nm. As the dimension of the quantum dot was closed to the size of the poly-Si grains that was about 50 nm [125], it was proposed to be defined by the surface roughness induced by the poly-Si grains in the poly-Si gate electrode [126]. The poly-Si grains on the poly-Si/$SiON$ interface could reduce the equivalent oxide thickness and increase the local gate capacitance. With a negative V_G bias, more holes would accumulate in this local area. In this case, the local channel energy level under the grain boundary area would be increased and form the quantum dot, which was shown in Figure 3.27.

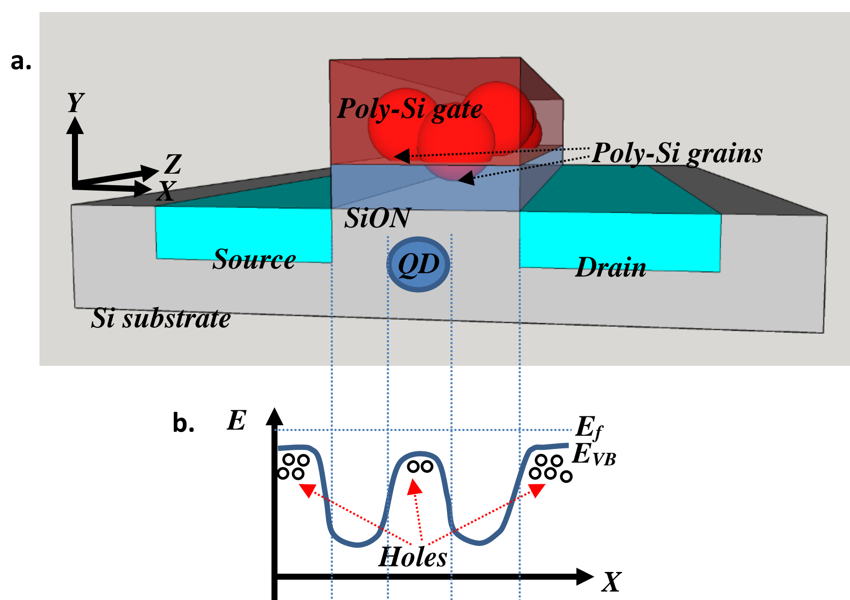


Figure 3.27 (a) The schematic diagram for the structure model of the device. The quantum dot was formed by the remote Poly-Si grains. (b) The band structure diagram in the channel. The E_f was the Fermi energy level. The E_{VB} was the valence band energy level.

As the channel length of this device was 75 nm, it was quite possible to form at least one quantum dot in the channel between Source and Drain. That could explain why the bright Coulomb diamonds could be observed, in which case the channel current was mainly dominated by one quantum dot. However, the channel width was 10 μm, the Poly-Si remote surface roughness should be everywhere caused by the grains in Figure 3.28. Many quantum dots with different sizes could be formed in the whole channel. Since the Poly-Si grains were randomly distributed, the coupling between different quantum dots in different bias conditions would be complex with mixed parallel and series relationship. That could be responsible for observing the blurry Coulomb diamonds between the six clear diamonds in Figure 3.25. It implied that if the narrow channel device was measured in the same condition, the clearer Coulomb diamonds should be observed which is mainly dominated by one quantum dot. That was one reason that the narrow channel device was measured in the next step.

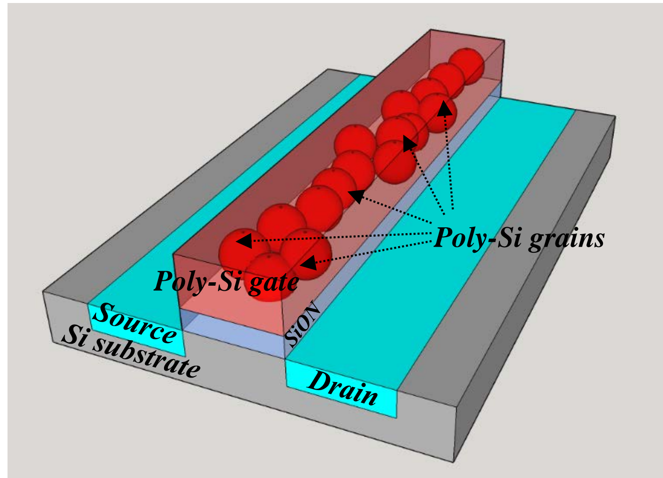


Figure 3.28 The schematic diagram for the structure model of the device. Many different quantum dots were formed by the randomly distributed Poly-Si grains

In order to simplify the model to investigate the relationship between the Coulomb blockade phenomenon and the RTSs phenomenon, we firstly could focus on the *D1*, *D2*, and *D3* that were shrunk continuously and next to each other. Since this pMOS transistor was not a standard single hole transistor [121], the gate bias did not only control the energy level in the quantum dot but also the barrier energy level. During the increasing of $|V_G|$, the holes would be accumulated and decrease the barrier height to make the capacitive coupling between C_S and C_D stronger [120]. This could simply explain the reason why the Coulomb diamond shrank gradually. On the other hand, the gradually shrunk Coulomb diamond phenomenon implied that in this bias window, the channel current was mainly dominated by one big quantum dot. If not, the Coulomb diamond size would not regularly decrease.

As a simplified model, it could be regarded as one quantum dot mainly controlling the transport in the channel in the gate bias window where the clear Coulomb diamond could be observed. Based on this model, the reason why the RTSs around the Coulomb diamond edges could be observed was investigated.

In order to understand where the RTSs came from, the average high-state time (τ_H) and the average low-state time (τ_L) as the function of gate bias was extracted from Figure 3.19 (c, d, e, f) for RTS1 and RTS2 shown in Figure 3.29.

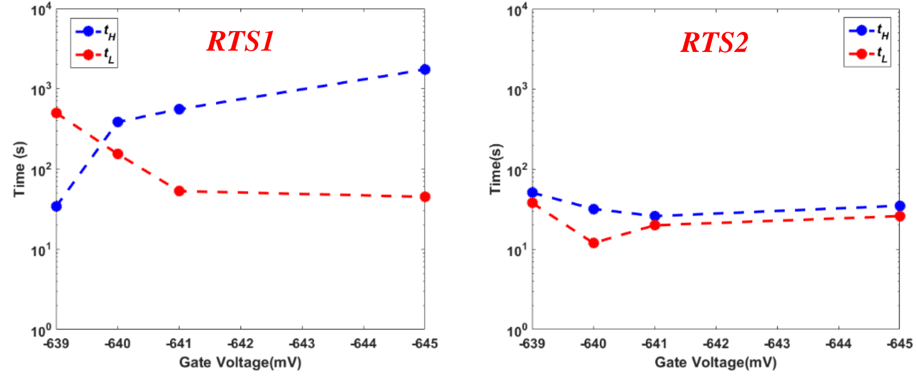


Figure 3.29 The average high-state current (τ_H) and the average low-state current (τ_L) as a function of gate voltage for the RTS1 and RTS2.

It could be found that RTS1 was a slow trap that should be located in the gate oxide [127]. The τ_H and τ_L of the RTS2 were much faster than the RTS1. According to the SRH theory, the capture time of the slow trap in the gate oxide was several orders of magnitude larger than the interface fast trap [6,37]. At the room temperature, the transition time for an interface trap whose energy level was located near the silicon mid-gap could be around 0.01s, whereas the transition time located near the band edges could be in microseconds [116]. However, the transition rate for an interface trap was exponentially decreased as the temperature decreased [116]. As we measured at 2K, the roughly transition time estimated for RTS2 was acceptable. So that they were consistent with the actual measurement data from Figure 3.29. The system noise and low measurement speed could also affect the transition time estimation for RTS2. In this case, RTS2 was a fast trap that was related to the interface state between the $SiON$ and Si substrate. The system noise was mainly from the temperature vibration in the cryostat and the no guard poor BNC cables, which made it difficult to get the precise valves from the raw data. The measurement speed was limited by the Agilent B1500.

If the RTS1 was supposed to be from the oxide trap, the amplitude of the RTS1 could be estimated by the number of carries in the Si channel, based on the equation [70, 128-129]:

$$\frac{\Delta I}{I} = \frac{1}{N_0} \tag{3.5}$$

$$N_0 = k_B T \frac{C_{gate}}{e^2} \tag{3.6}$$

Where N_0 was the number of carriers in the channel, k_B was the Boltzmann constant, T was the temperature value in Kelvin, e was the value of elementary charge and C_{gate} was the total gate capacitance, which can be estimated as

$$C_{gate} = \frac{\varepsilon_0 \varepsilon_{SiO_2} WL}{t_{ox}} \tag{3.7}$$

where ε_0 was the permittivity in a vacuum, ε_{SiO_2} was the dielectric constant of SiO_2, W was the width of the channel, L was the length of the channel and t_{ox} was the equivalent oxide thickness of the $SiON$ layer. $\Delta I/I$ for RTS1 could be estimated at 8.8% theoretically. Then the actual value $\Delta I/I$ both for RTS1 and RTS2 were extracted from Figure 3.19 (c, d, e, f) and shown in Figure 3.30. It can be safely estimated that RTS1 was from an oxide trap and RTS2 was from an interface trap by comparing our experiments and the others' previous work [5].

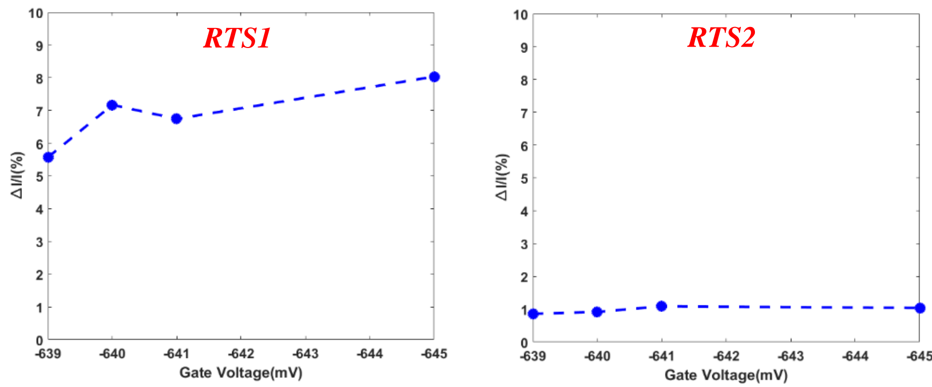


Figure 3.30 The $\Delta I/I$ as a function of gate voltage for the RTS1 and RTS2.

By combining the two types of traps and the single hole transistor model, the physical model could be established to describe the two types of RTSs, shown in Figure 3.31.

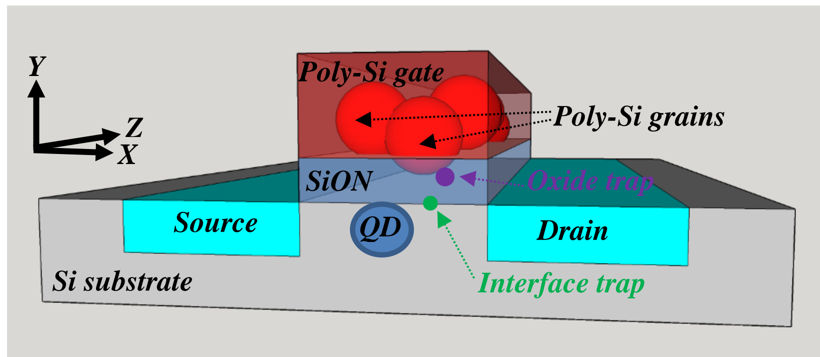


Figure 3.31 The physical model to explain the original source to form the RTS1 and RTS2.

The quantum dot in the channel, which was generated by Poly-Si grains in the gate, mainly controlled the drain current in the certain gate voltage range at low temperature. The trap that was responsible for RTS1 was located in the $SiON$ layer. The trap corresponding to RTS2 was located on the surface between Si layer and $SiON$ layer.

For RTS1, the charged oxide trap could make an additional energy state in the potential barrier for the quantum dot and then change the transparency of the potential barrier [130-132]. The oxide trap was supposed to be a hole trap, due to the fabrication process for the pMOS transistor. The gate, source and drain electrodes were high p-doped. During the annealing process, the dopant

could diffuse to the oxide layer, which would form a hole trap [120]. If a hole was captured by the trap, the trap would be in the neutral state. There was no effect on the potential barrier in the channel. However, if the hole was emitted from the trap, the trap would be negative. The negative charge in the oxide layer could form a small potential well in the channel. As the transistor was in the single hole transistor situation, if this potential well were located in one side of the quantum dot barrier, it could increase the transparency of this potential barrier and then increase the current, which is shown in Figure 3.32.

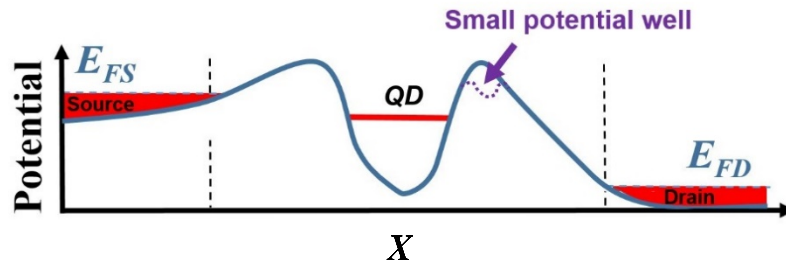


Figure 3.32 The schematic diagram of the potential in the channel cross-section from source to drain.

In this case, the low and high states of the current could respond to the trapping and de-trapping of a hole in the trap. Therefore, the probability distribution of I_D ($P(I_D)$) in high and low states could describe the probability of a hole to occupy the trap. The trap could be also treated as a small quantum dot. The occupation of the quantum dot can be described by the wave function (φ_{QD}). So that the relationship could be shown as:

$$P(I_D) = |\varphi_{QD}|^2 \tag{3.8}$$

As the occupation of the quantum dot was a true stochastic process, the I_D in high or low state would be also a true stochastic process which was the main reason why the RTS can be used to generate truly random numbers. Then the next question was whether the probability could be well controlled or not.

In order to investigate the properties of the oxide trap, we extracted the probability P_H that the I_D was found at the high current state and P_L at the low current state for RTS1. The probability distribution in each high or low state in Figure 3.19 (h-n) and Figure 3.22 (g-l) were found to be in coincidence with Gaussian distribution functions. The probability corresponded to the area formed by the probability distribution function. So that the probability of each state was calculated by integrating the probability distribution function over the corresponding current. In this case, P_H and P_L were extracted as:

$$P_H = \frac{\int_{-\infty}^{\infty} P_H(I_D)\, dI_D}{\int_{-\infty}^{\infty} P_H(I_D)\, dI_D + \int_{-\infty}^{\infty} P_L(I_D)\, dI_D} \quad (3.9)$$

$$P_L = \frac{\int_{-\infty}^{\infty} P_L(I_D)\, dI_D}{\int_{-\infty}^{\infty} P_H(I_D)\, dI_D + \int_{-\infty}^{\infty} P_L(I_D)\, dI_D} \quad (3.10)$$

Using this mothed, P_H and P_L of RTS1 were calculated in the selected bias conditions to demonstrate the dependence on V_G and V_D (Figure 3.33-3.34)

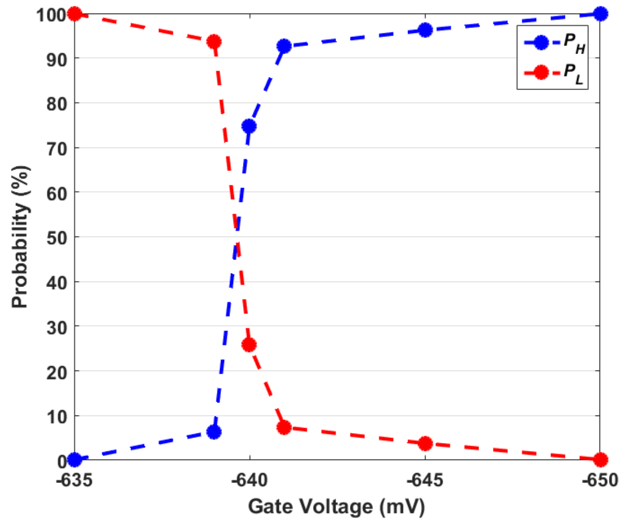


Figure 3.33 The V_G dependence of P_H and *P_L in RTS1. V_D bias was -13.5 mV.*

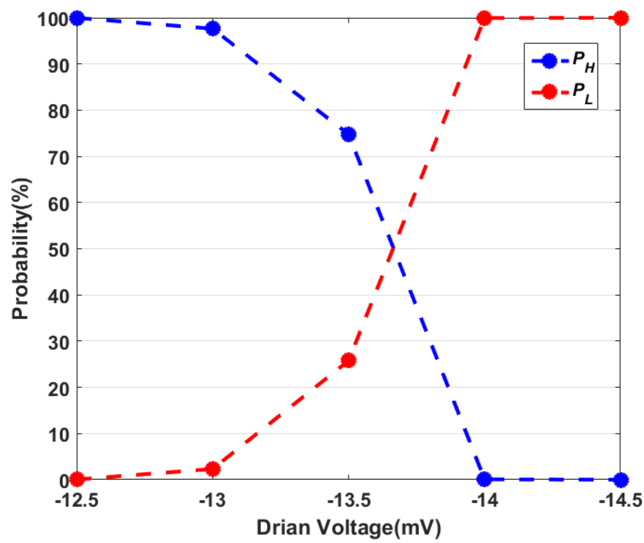


Figure 3.34 The V_D dependence of P_H and *P_L in RTS1. V_G bias was -640 mV.*

From Figure 3.33, the probability of I_D in high state rose with the increasing of $|V_G|$. The oxide trap was supposed to be a hole trap. On smaller $|V_G|$ bias, the potential difference between the hole in the trap and the gate electrode was small. The hole was more likely to be captured by the trap. In

this condition, I_D was more likely to be observed in the low state. While increasing $|V_G|$ bias, the potential difference will be increased, and the tunneling rate will also be increased [133]. In this condition, I_D was more likely to be observed in the high state. Therefore, the hole trap model (Figure 3.35) was in agreement with the experiment data in Figure 3.33. In contrast, the probability of I_D in high state was observed to decrease with the increasing of $|V_D|$ bias. This phenomenon could be explained by the short channel effect [120]. Due to the channel length was just 75 nm, the strong Drain-Induced barrier lowering (DIBL) effect could affect not only the potential in the channel but also the electrical field in the oxide layer. A larger $|V_D|$ bias would have the opposite effect with $|V_G|$ bias, which decreased the functionality of $|V_G|$ bias to dominate the tunneling rate. As we could see a strong dependence on $|V_D|$ bias, the oxide trap should be close to the drain electrode. As the short channel effect caused by larger $|V_D|$ bias could make the model complex, we would mainly focus on $|V_G|$ bias control to the probability of I_D.

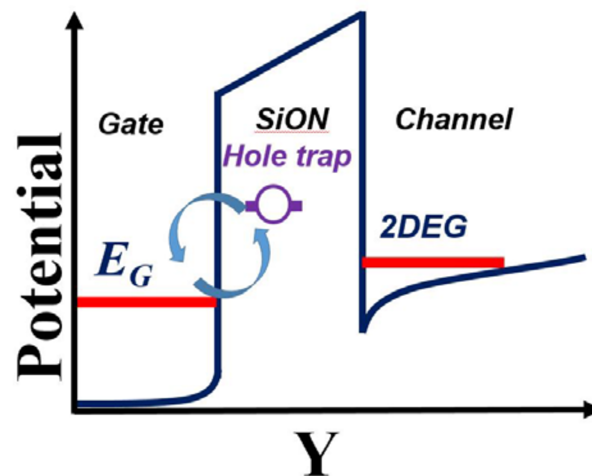


Figure 3.35 The potential diagram with the hole trapping and de-trapping from the oxide layer to the gate electrode.

For the RTS2, the charged interface traps could increase the lattice and Coulomb scattering and degrade the mobility [116]. In this case, if the interface trap was charged, the current would be decreased which can explain the current vibration. This interface trap could still be treated as a small quantum dot, which could trap and de-trap a hole with a small potential barrier [116]. Although the mechanism to affect I_D between the oxide trap and the interface trap was different, the same method for RTS1 could be used to investigate the properties of the interface trap. We extracted the probability P_H that the I_D was found at the high current state and P_L at the low current state for RTS2 on V_G and V_D dependence, shown in Figure 3.36 and Figure 3.37.

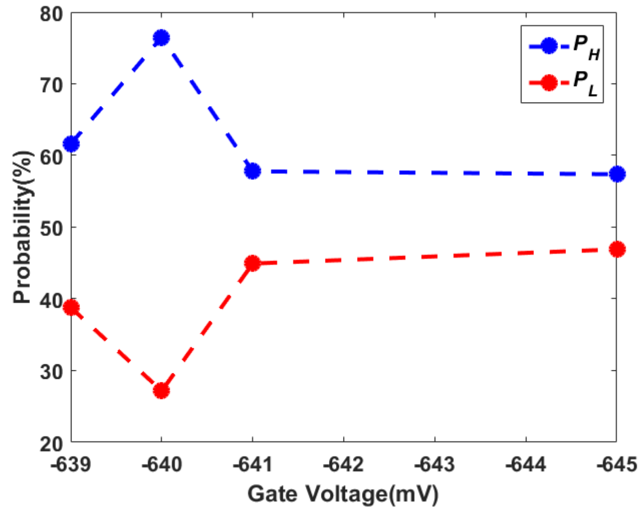


Figure 3.36 The V_G dependence of P_H and *P_L in RTS2. V_D bias was -13.5 mV.*

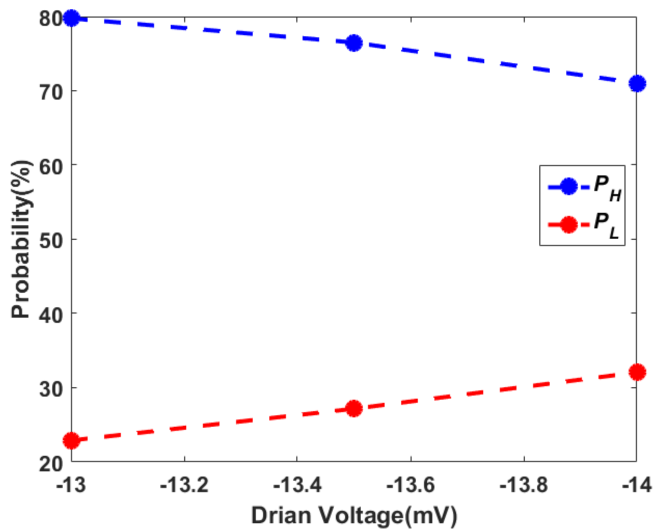


Figure 3.37 The V_D dependence of P_H and *P_L in RTS2. V_G bias was -640 mV.*

From V_D in Figure 3.36, V_G did not correspond to the probability distribution of I_D for RTS2. It implied that the gate bias did not affect the trapping and de-trapping process of the interface trap. As the trap was located in the interface of the channel, the quantum barrier for this trap was defined by the energy gap between the trap energy level and the valence band level [116]. The gate bias would increase or decrease them at the same time. In this case, the gate bias could not change the tunneling rate of the interface traps. That could explain why there was almost no relationship between V_G and the probability distribution of I_D for RTS2. Besides, V_D bias could directly influence the channel potential level by DIBL effect, which could explain the probability distribution of I_D for RTS2 had a weak dependence on V_D bias. In this condition, it seemed that RTS2 was not suitable to be used for random number generation, because the amplitude was small and the probability

distribution was unable to be well controlled. However, the tunneling rate was much faster than RTS1, which would be a key factor for the device to be used as a random number generator.

As a brief summary of this part, the model was established to explain why different kinds of RTSs could be found on the Coulomb diamond edge at the low temperature. This provided a method to investigate the nature of the RTSs. However, the device we measured was a 10 μm width pMOS. It increased the chance to find an oxide trap that could influence carrier transport in the single hole transistor. However, because gate surface roughness was everywhere, many quantum dots might be activated at the same time to form many parallel single-hole channels, which could be one reason why the current could be observed to be multi-levels in addition to the RTS1 and RTS2. In order to confirm the model, the narrow channel pMOS was measured in the next part. Besides, the system noise level was around 10-50 pA due to the poor wire connection and the unguarded "Break-out" box. That also caused a small strange current from "Bulk" terminal to "Source" terminal, which was identified to be generated from the "Break-out" box. Fortunately, it did not affect the device performance in the cryostat. Otherwise, we could not get the standard I_D-V_D and I_D-V_G characterization. However, we could not confirm the source current was opposite to the drain current. In this case, after finishing the above experiment, the system was upgraded to 100 fA noise level.

3.4 RTN in a narrow-channel MOSFET (W/L=60nm/60nm)

In order to confirm the single electron/hole model and RTSs model in the last part, narrow-channel MOSFETs were bonded and measured at low temperature. Also, the temperature dependence of the RTSs was investigated. Before doing the measurement, the insert probe was modified to low noise feedthrough, the wire connections between the insert probe and the Agilent B1500 were changed to SMA and triaxial cables.

3.4.1 Measurement system upgrade

The root of the noise problems for the original setup was that the signal was no guard and no shield in the "Break-out" box. Moreover, after the "Break-out" box, all the signals were mixed in a 4-meter cable with 12-pin Fischer connectors. Although a new "Break-out" box with 100 fA noise level was designed and made with triaxial connectors and Fischer connectors (Figure 3.38), large parasitic resistances and capacitances were found in the 4-meter cable, which could cause more than 50 pA charging current peak while changing the voltage. This was also confirmed by the researchers in the National Physical Laboratory (UK) with their equipment. Furthermore, the connected wires in

the insert probe were just coated wires. Although the insert probe could be a shield from outside, the signals between each other were poorly isolated and shielded.

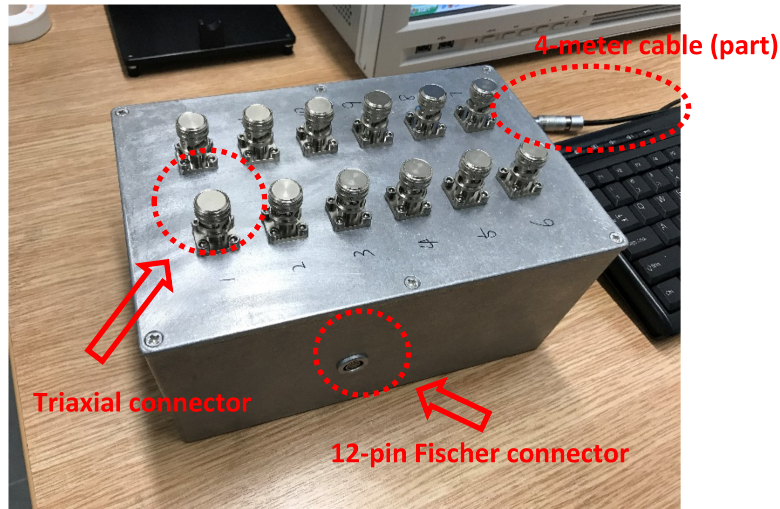


Figure 3.38 New "Break-out" box with full metal shield.

After identifying the noise source, we decided to re-design the total electrical feedthroughs from the Agilent B1500 to the sample carrier. The insert probe was totally modified to improve the signal transmission quality.

Firstly, the connection panel on the top of the insert probe was re-designed. The original 12-pin Fischer connector was replaced by 4-SMA connectors and 2-triaxial connectors (Figure 3.39-3.40).

Figure 3.39 The original connection panel layout with all singles mixed in one 12-pin Fisher connector[134].

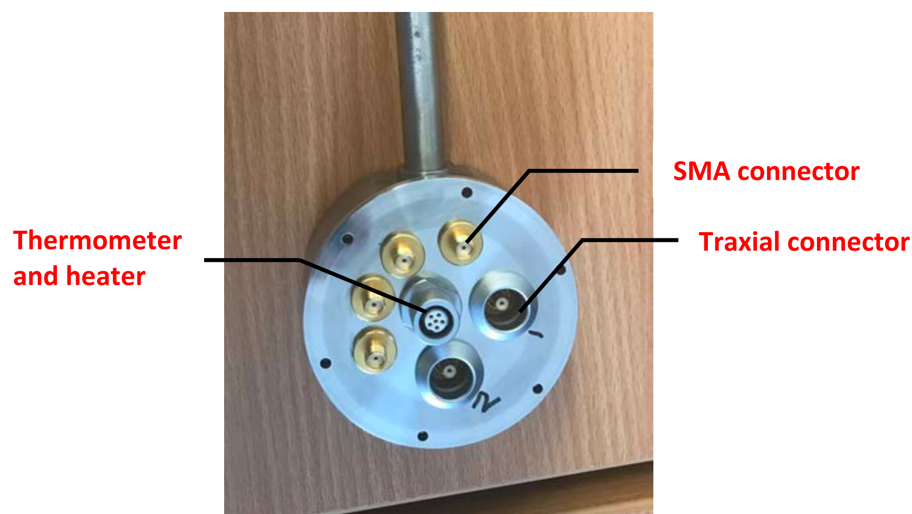


Figure 3.40 The modified connection panel layout with 4-SMA connectors and 2-triaxial connectors to separate the singles by independent guarding and shielding.

Secondly, the green tube on the bottom of the insert probe was re-designed to replace the coated wires by high vacuum SMA cables and triaxial cables (Figure 3.41). The chip carrier was directly connected to the feedthrough connectors from the connection panel on the top of the insert probe.

Figure 3.41 (Left) The original chip carrier mounted tube, which could mount 2 chip carriers at the same time. (Right) The re-designed chip carrier mounted tube, which only had 2 chip carrier position, but low noise level.

From the connection panel, the low noise SMA cables and triaxial cables were used to establish the connection to Agilent B1500 by suitable adaptors (Figure 3.42).

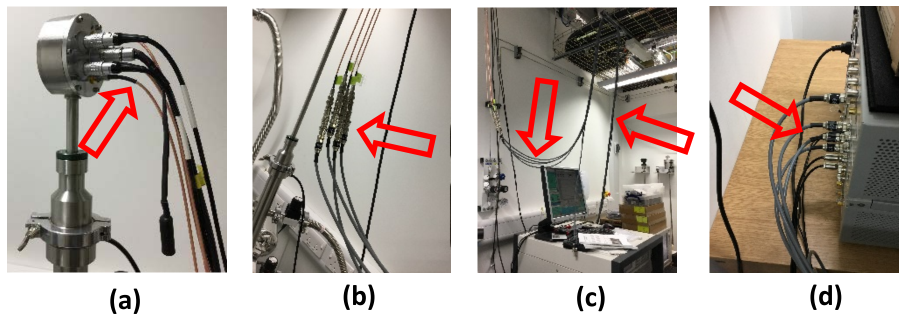


Figure 3.42 (a) 4 SMA cables and 2 triaxial cables making the electrical connection between the insert probe and outside equipment. (b) SMA cables connecting to triaxial cables by guard disconnected adaptors. (c) The long triaxial cables through the ceiling to reach the Agilent B1500. (d) Triaxial cables connecting with Agilent B1500 SMU ports.

After modifying the insert probe and rebuilding the electrical connections, a test sample was loaded into the cryostat using this new insert probe to test the lowest temperature and noise level. The lowest temperature was increased to 4 K slightly, which was mainly caused by the 4 SMA cables. The 4 SMA cables increased the thermal coupling between the outside and the VTI and broke the thermal equilibrium. As far as the temperature could be stable at 4 K, it indicated that the long insert prob performed in good functionality as a thermal anchor. As the 2 triaxial cables were special cryogenic cables, they did not contribute to the thermal leak too much. Whatever 4 K was enough to find the Coulomb blockade phenomenon in a MOSFET based on the previous result from our group [121]. What's more, the noise level was more important than the temperature change. Fortunately, the noise level was well controlled around 100 fA (+/-5V bias) in the 4 SMA channels. On the contrary, the 2 channels with the triaxial cable were a little bit higher, which were around 200 fA (+/-5V bias). It could be found that the special cryogenic triaxial cables performed better in the thermal leakage and the SMA cables performed better in the noise level. Overall, the system noise was well controlled below 1 pA, and the base temperature could be stable at 4 K. In this condition, we could confirm the system upgrading was successful.

3.4.2 Measurements and discussion

With the well-upgraded system, a narrow-channel nMOS with 60 nm length and 60 nm width was measured. This device was the smallest device in total samples. Based on the experience from the previous session, the device was directly cooled down to the lowest temperature, then heated up to the selected temperature to investigate the dependence of RTSs on the temperature. We measured the Coulomb blockade effect at 5 K, 10 K, 20 K, 40 K, and 60 K separately to identify that to what extent the quantum dot could block the drain current against the thermal energy at different temperatures. Experimental results are shown in Figure 3.43-3.47.

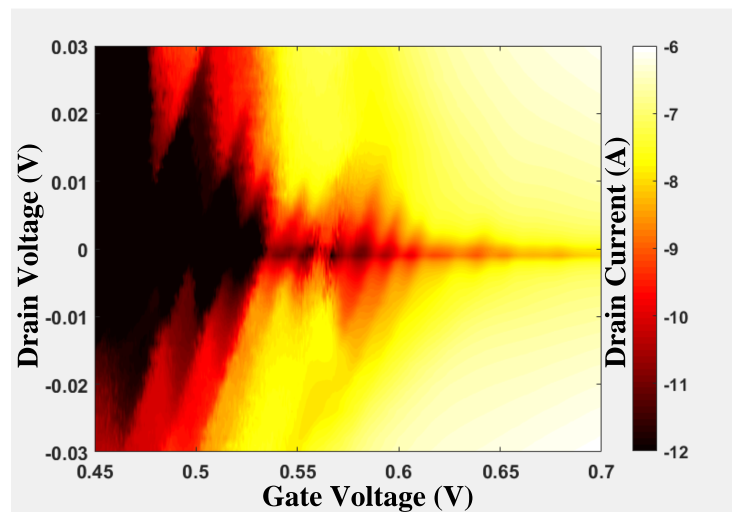


Figure 3.43 The contour plot for I_D versus V_G and V_D of the nMOS at 5K.

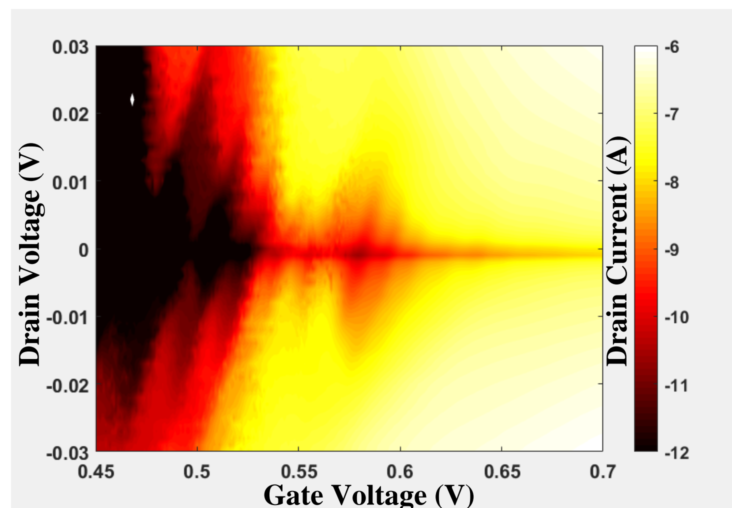


Figure 3.44 The contour plot for I_D versus V_G and V_D of the nMOS at 10K.

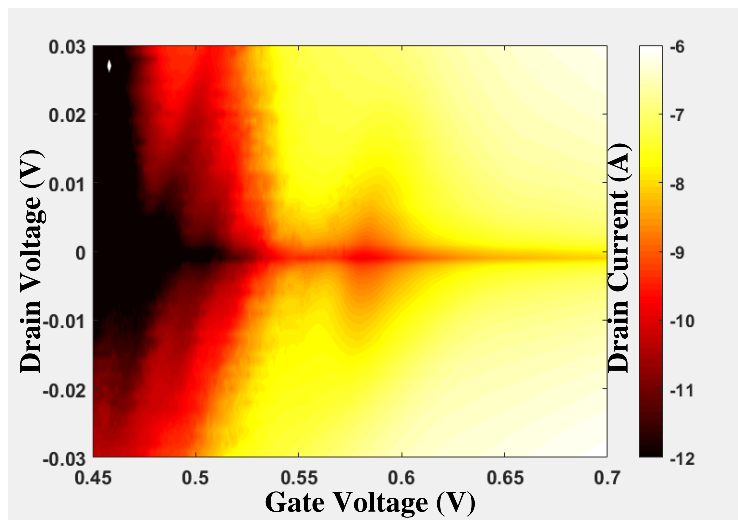


Figure 3.45 The contour plot for I_D versus V_G and V_D of the nMOS at 20K.

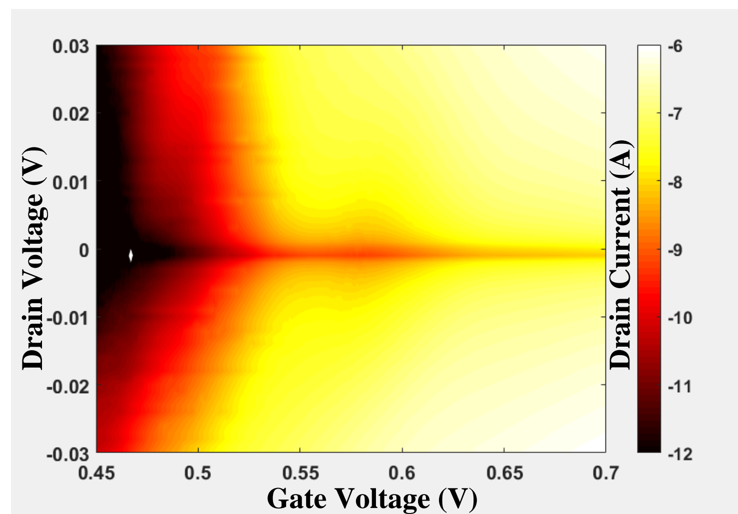


Figure 3.46 The contour plot for I_D versus V_G and V_D of the nMOS at 40K.

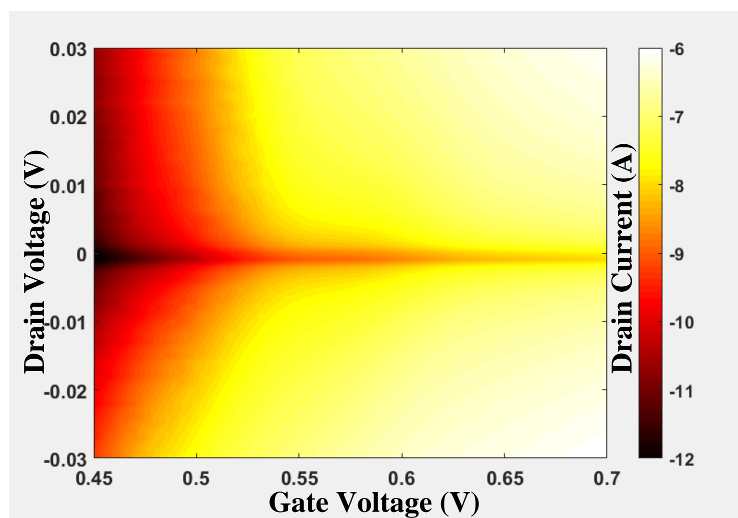


Figure 3.47 The contour plot for I_D versus V_G and V_D of the nMOS at 60K.

As the channel area was nearly equivalent to the Poly-Si grain size, the Coulomb diamond was much clear to be observed as we expected. The limited number of quantum dots have dominated the channel. From these five contour plot pictures, we could clearly find that the Coulomb diamond gradually disappeared. The thermal energy was rising along with the increase in temperature. Based on our single-electron transistor model, the potential barrier caused by the remote surface roughness was limited by the physical size of the Poly-Si grain. At the higher temperature, the thermal energy could overcome the potential barrier or even larger than the potential barrier. In this case, the Coulomb blockade effect was degraded. At 60 K, the Coulomb diamonds almost disappeared which provided a way to estimate the potential barrier height caused by the remote surface roughness. At room temperature (300K), the thermal energy could be approximately 26 meV. The Coulomb diamonds could be observed clearly at 5 K and disappeared at 60 K. The thermal energy was written as $k_B T$, where k_B was the Boltzmann's constant and T was the absolute temperature. Then the diffidence of the thermal energy between 5 K and 60 K could be calculated as 4.77 meV. It implied that the potential barrier height was approximately 4.77 mV on lower gate bias where the Coulomb diamonds could be observed at 5 K. Higher gate bias would decrease the potential barrier as we explained in the previous session. Once the thermal energy was larger than the potential barrier height, the Coulomb diamonds could not be observed.

By confirming the Coulomb diamonds at low temperature, the RTSs were measured on the edges of one Coulomb diamond based on our model at 10 K, 40 K, 60 K, 100 K, 150 K, and 200 K, separately. As the system was upgraded, the measurement speed could be much faster at a low noise level. The sampling interval was set to 4 ms. The RTSs caused by interface traps could be investigated in more detail. On each voltage bias condition, 5001 points were measured.

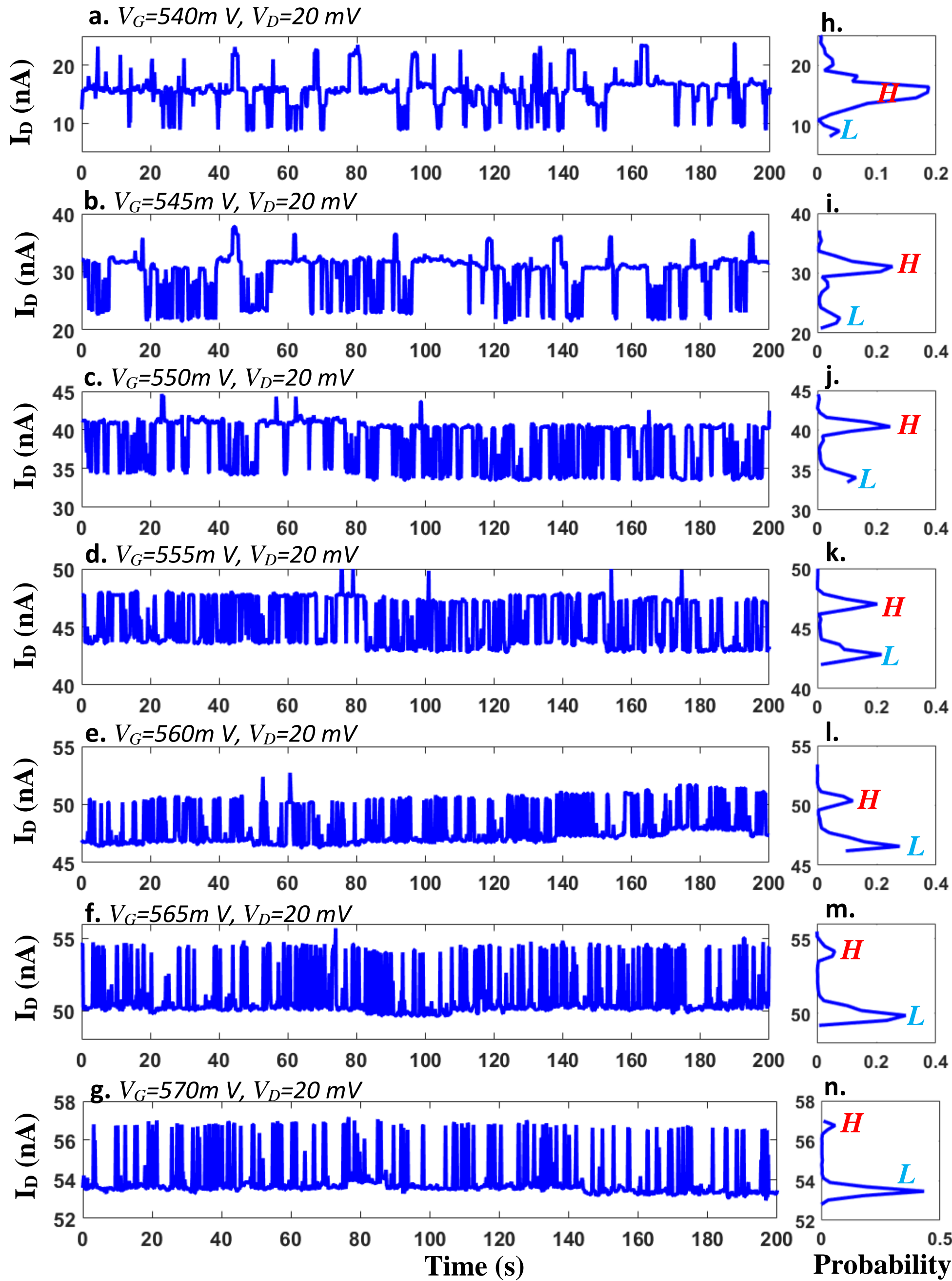


Figure 3.48 (a-g) The time-domain measurement results of the V_G dependence for the RTSs at 10 K. (h-n) The probability distribution of the current versus its valve. The high state of RTS was marked as red "H"; the low state of RTS was marked as blue "L".

In Figure 3.48, RTSs were observed at 10 K at the edge of one Coulomb diamond. As we expected, it was less likely to observe the RTSs with long switching time, which caused by the oxide traps in such a small device. One reason was that the channel area was too small to get the chance to find an oxide trap. The other reason was that the gate bias in the selected window was not enough to activate the deep oxide trap. However, the RTS in Figure 3.47 still showed a gate bias dependence, in which the probability of I_D in high state was decreased while increasing the V_G, which is shown in Figure 3.49.

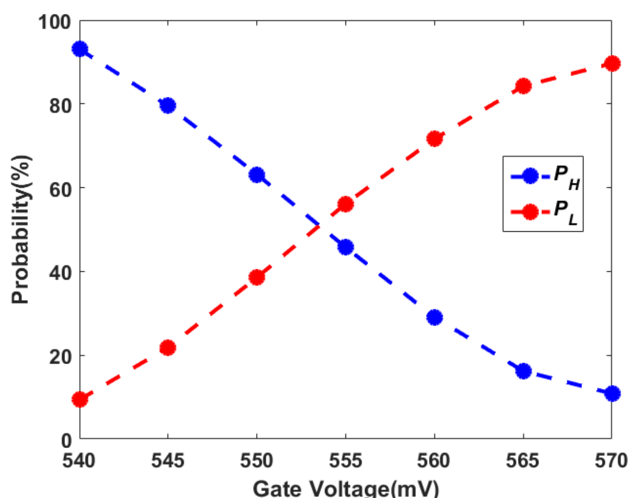


Figure 3.49 The V_G dependence of P_H and P_L in RTSs. V_D bias was 20 mV.

This RTS was quite different from the RTS1 and RTS2 in the previous session. It had a large amplitude and short switching time. Both of the features were suitable for random number generation. In order to investigate more details of the nature of this RTS, we extracted the average high-sate time (τ_H) and the average low-state time (τ_L) of the I_D (Figure 3.50). As τ_H and τ_L were both in millisecond level, the RTS was considered to be caused by the interface trap [116]. In the previous session, the RTS2 caused by the interface trap did not show a clear dependence on the gate bias. One reason could be that the bias window for the previous device was suitable for characterizing the RTS1 caused by the oxide trap. The probability distribution of I_D was reversed in a very narrow region about 10 mV (-635 mV to -645 mV). The V_G dependence of RTS2 might not be obvious. The other reason could be that the oxide trap might be close enough to affect the interface trap as they were both located in the same channel of one single hole transistor. In that case, the gate bias dependence would be degraded. The last but most important reason could be that the noise level of the system was too high to extract the RTS2 data precisely. In this case, by solving these problems, we could measure the pure RTS2 signals in a wide gate bias window from 540 mV to 570 mV. In addition, we increased the measurement speed from 1 s to 4 µs for the time interval. That was why we could measure much smaller capture and emission time for the interface trap. The measurement temperature 10 K should also be considered for comparing with the previous one at 2 K. With a higher thermal energy, capture and emission time was also decreased to some extent [5].

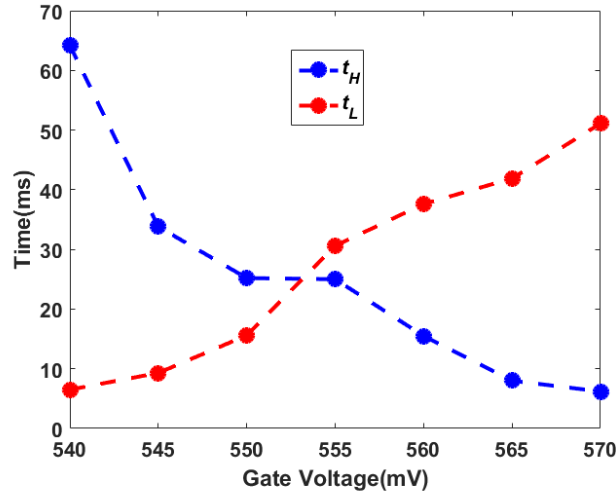


Figure 3.50 The τ_H and the τ_L as a function of gate voltage for the RTS.

As the amplitude of this RTS was quite large (Figure 3.51), even larger than RTS1 in the previous device, we supposed that the interface trap was just located under the remote Poly-Si grain that caused the quantum dot. In this case, the interface trap created an additional energy level in the quantum dot, which provided an additional resonant tunneling channel to increase the current in large amplitude [135].

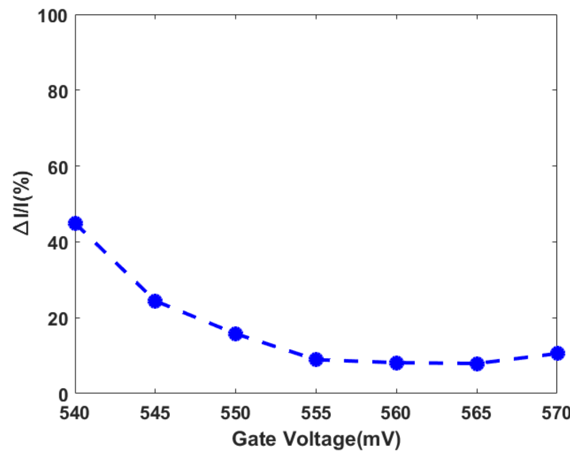


Figure 3.51 The $\Delta I/I$ as a function of gate voltage for the RTS.

The potential diagram across the channel was shown in Figure 3.52. If the interface trap was not occupied by the electron, the additional tunneling channel was opened. From Source to Drian, it became more transparent. As a result, I_D would be increased. If the interface trap were occupied by the electron, the trap would become neutral and the additional tunneling channel would disappear. The current could only be dominated by the quantum dots generated by the Poly-Si grain. In this case, I_D would be in the low current state. In the low gate bias condition, the Fermi level in the quantum dot was low. The electrons were less likely to occupy the charge trap. So that we could observe I_D was more likely in the high state (Figure 3.48(a)). When the gate bias was increased, the

Fermi level in the quantum dot was also increased. The higher energy electrons were more likely to occupy the charge trap. So that we began to observe that I_D was more likely in the low state (Figure 3.48(b,c,d,e,f,g)). The experiment data made an agreement with our model.

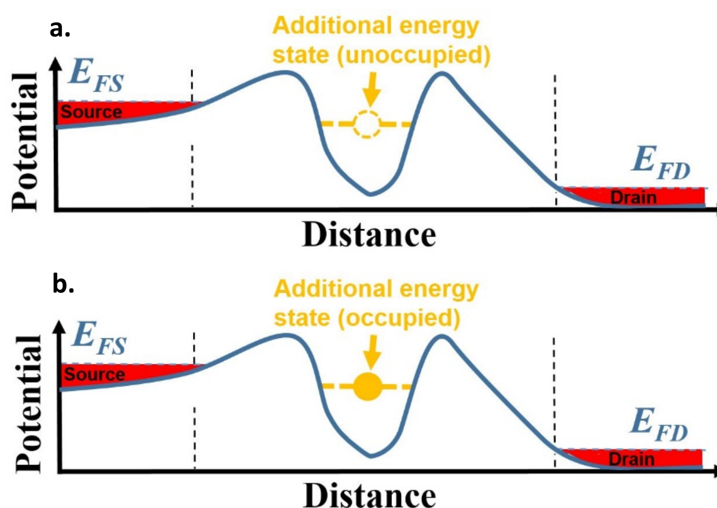


Figure 3.52 The physical model of the RTS. (a) The unoccupied state of the additional energy state. (b) The occupied state of the additional energy state.

In order to investigate the temperature dependence of the RTS, we did the time-domain measurements in the same gate bias window at 40 K, 60 K, 100 K, 150 K, and 200 K, separately. At 40 K (Figure 3.53), the RTSs were still observed a similar amplitude at 10 K. The probability distribution of I_D also showed the same dependence on V_G. However, the RTSs at 40 K had much long capture and emission time than 10 K obviously. If they were caused by the same interface trap, the capture and emission time should be lower with the higher thermal active energy. In this case, the RTSs at 40 K was supposed to be caused by an oxide trap due to the long capture and emission time [116]. In addition, more than one traps should be activated, because more than two levels of the current could be observed.

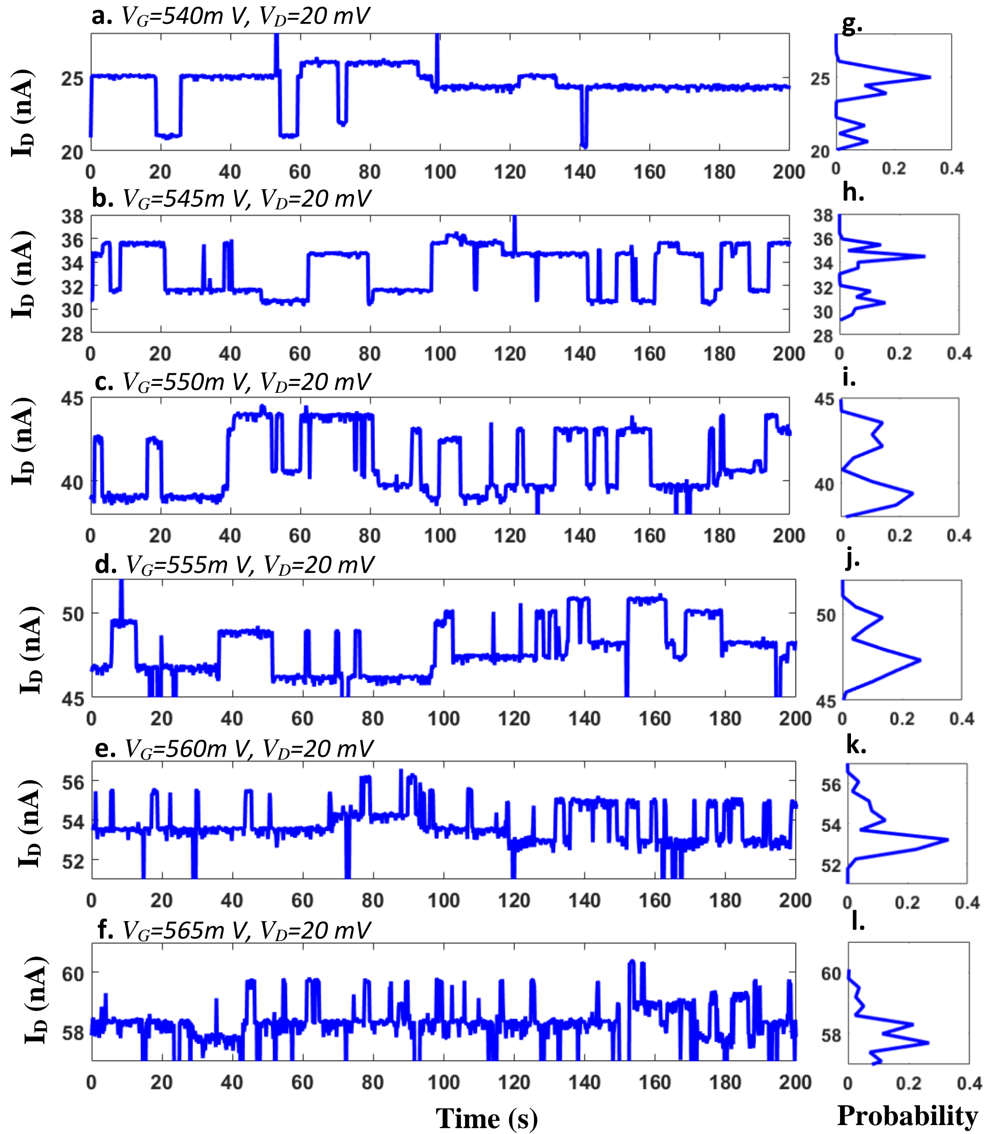


Figure 3.53 (a-f) The time-domain measurement results of the V_G dependence for the RTSs at 40 K. (g-l) The probability distribution of the current versus its valve. The high state of RTS was marked as red "H"; the low state of RTS was marked as blue "L".

When the temperature was increased to 60 K, it could be found that a deep oxide trap dominated the I_D in the V_G dependence measurement (Figure 3.54). At the same time, small RTSs could be observed clearly at both the high and low current states. Considering the short capture and emission time, the small RTSs were suggested to be caused by the interface trap. By the interference from the larger RTSs, the V_G dependence of the small RTSs could not be observed.

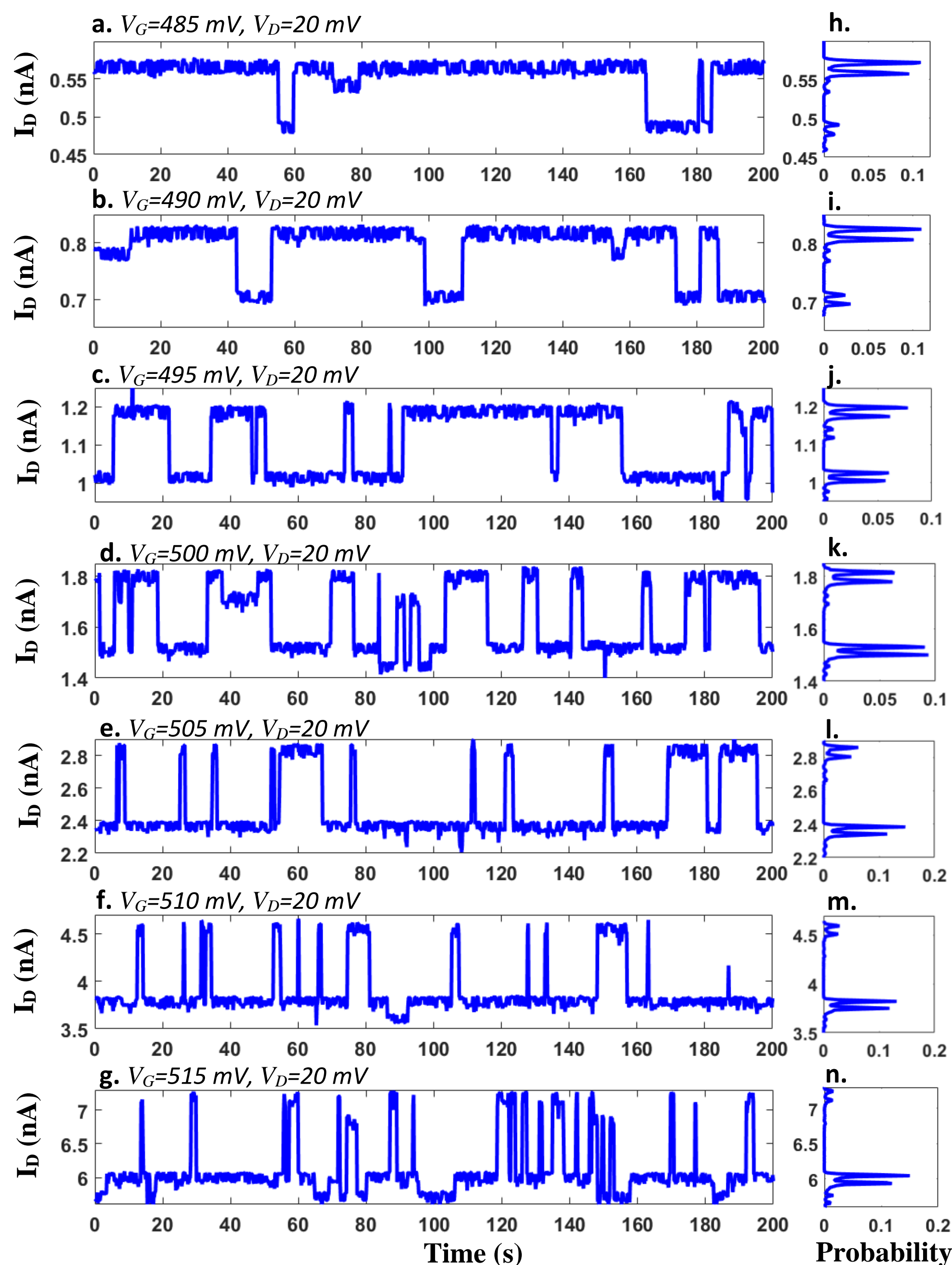


Figure 3.54 (a-g) The time-domain measurement results of the V_G dependence for the RTSs at 60 K. (h-n) The probability distribution of the current versus its valve. The high state of RTS was marked as red "H"; the low state of RTS was marked as blue "L".

From the measurement results at 40 K and 60K, it was found that the deep oxide traps started to be activated. Once the oxide traps were active, it would affect the interface traps to lose the V_G dependence. We should also consider that more interface traps would also be active in a higher temperature in which the Fermi level was higher. This point was confirmed by the measurement result at 100 K, 150 K and 200 K (Figure 3.55-3.57). At 100 K and 150 K, many kinds of small RTSs caused by different interface traps were mixed together. In this case, we could not observe the typical two-level small RTSs in the low or high state of the big RTSs. At least we could still observe that one oxide trap with a long lifetime mainly dominated the channel. At 200 K, more parts of the channel were active, many oxide traps and interface traps had the complex interference between each other. The amplitude was decreased and two levels of the current states merged gradually. As this device was a high-performance transistor, the channel was highly doping to increase the ON current. If the temperature continued to increase, the channel was almost switched on at this bias range. The carrier concentration increased by the oxide traps would be negligible. In this condition, we would be unable to see the typical RTSs at higher temperatures, especially in the large channel area. This was consistent with room temperature results in Figure 3.6.

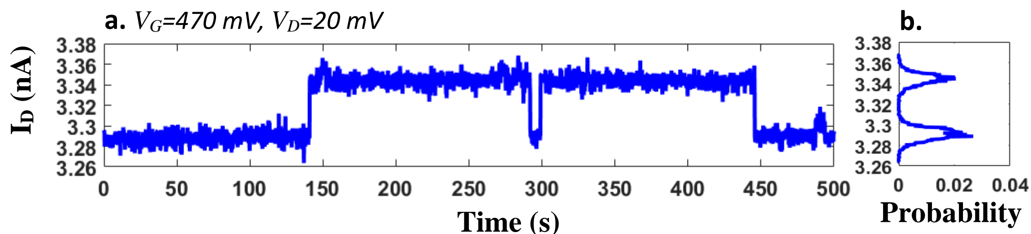


Figure 3.55 (a) The time-domain measurement results of the V_G dependence for the RTSs at 100 K. (b) The probability distribution of the current versus its valve. The high state of RTS was marked as red "H"; the low state of RTS was marked as blue "L".

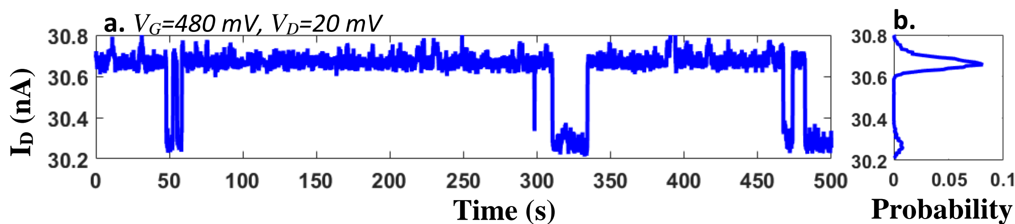


Figure 3.56 (a) The time-domain measurement results of the V_G dependence for the RTSs at 150 K. (b) The probability distribution of the current versus its valve. The high state of RTS was marked as red "H"; the low state of RTS was marked as blue "L".

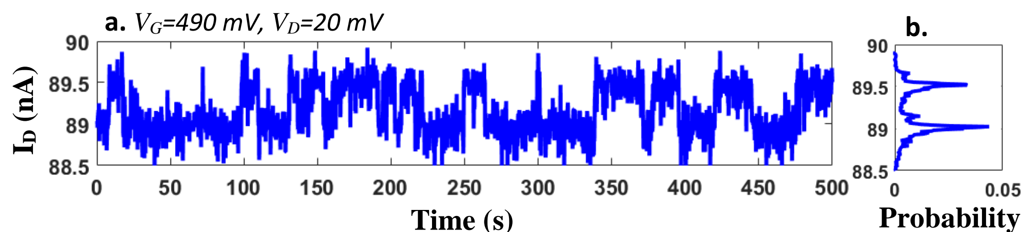


Figure 3.57 (a) The time-domain measurement results of the V_G dependence for the RTSs at 200 K. (b) The probability distribution of the current versus its valve. The high state of RTS was marked as red "H"; the low state of RTS was marked as blue "L".

In addition, by upgrading the cryostat system, we could confirm the source current was always consistent with the drain current (Figure 3.58). In this case, we could confirm that the RTSs were not caused by the gate junction leakage [5]. This point was quite important to identify the nature of the RTSs towards the random number generation.

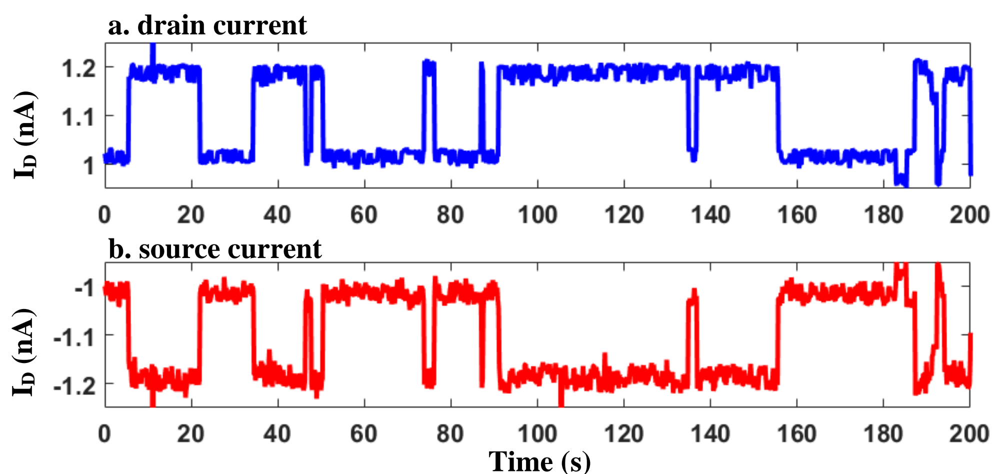


Figure 3.58 The comparison of the I_D and I_S of the RTS at the same time scale. (a) The I_D time-domain measurement results of the V_G dependence for the RTS (b) The I_S time-domain measurement results of the V_G dependence for the RTS.

3.5 Establish the relationship between the Coulomb diamond, RTSs and QRNG

So far, the reason why the RTSs could be found based on the Coulomb diamond at low temperature had been explained. Based on the bias conditions at Coulomb diamond edges, a one-channel single hole/electron transistor model could be achieved. Although the device could be a wide-channel device, the other channels could be suppressed at a certain bias condition. In this case, the effect of the selected trap could be clearly observed as typical RTSs. On the contrary, the bias conditions far away from the Coulomb diamond would activate more parts of the total channels even in the narrow channel devices, the RTSs from many traps would be mixed. Then the typical two-level RTS phenomenon could not be observed.

The reason why the wide-channel device was chosen at the beginning was that it increased the probability of finding the oxide traps. In order to see the typical RTS caused by an oxide trap, the quantum dot was also needed to form the single electron/hole transistor. So that the horizontal peaks could not be observed in all of the Coulomb diamonds' edges in Figure 3.16. As the Poly-Si grains were everywhere in the gate electrode, a wide-channel device definitely had a higher probability to observe the typical RTS caused by an oxide trap. However, the wide-channel device might cause multiple quantum dot coupling at the same bias condition, which could be also observed in Figure 3.25. Between the six selected Coulomb diamonds, many small diamonds could be also recognized. Then the current was not dominated by one quantum dot, and multiple RTSs would be mixed to degrade the typical RTS two-level system. Later in the narrow-channel device, we did not observe the RTSs caused by the oxide traps at low temperature, the RTSs caused by the interface traps were observed. It confirmed that the probability of finding an oxide traps whose energy level was just aligned to the quantum dot in the certain gate bias was quite low at low temperatures. However, it did not mean that there were no oxide traps. After increasing the temperature to raise the thermal energy, the oxide traps in high energy levels were activated, which could be confirmed in the high-temperature measurement result. In addition, as the energy level of the interface traps were much closer to the channel [116], they were more likely to be activated with the quantum dots. In this case, it was more likely to observe the pure RTSs caused by interface trap in the narrow channel device at low temperature. This provided a way to eliminate the RTSs caused by the oxide traps.

Considering the final target was to use the RTSs as a random number generator in the silicon devices, two kinds of RTSs could be selected now. As we discussed in the last part, both of these two kinds of RTSs were generated by the quantum tunneling events, which were the true stochastic processes. From the measurement results, the RTSs generated by the oxide trap had a larger amplitude, which

was much easier to measure. And the probability distribution of the current could be more sensitive to the gate bias. However, the oxide trap was quite randomly distributed in the oxide layer. This uncertainty was useless in the engineering point of view. Moreover, the capture and emission time was too long. It could be more than hundreds of seconds. The capture and emission time for an oxide trap definitely limited the speed of the number generator. In contrast, the RTSs generated by the interface trap had a higher switching rate. Although the probability distribution of the current could be affected by the RTSs caused by the oxide traps and the amplitude was smaller, one way was found to observe only the RTSs caused by the interface trap. In addition, the wind-window gate bias dependence was suitable to control the probability distribution more precisely. Moreover, if the energy level of the interface trap was aligned with the main quantum dot, the large amplitude of RTSs could be observed due to the resonance tunneling current. The interface trap could be treated as a small quantum dot in the channel. In the devices that had been measured, the gate bias controlled both the energy level of the quantum dot and the barrier height. At higher temperature, the barrier height was significantly decreased comparing to the thermal energy. That was why we lost the gate bias control to the RTSs caused by interface traps at higher temperatures. Due to the short channel effect, drain bias control was not a good idea. It provided a method to design a QRNG based on the channel quantum dot. If both the energy level of the quantum dot and the barrier height could be well controlled, the probability distribution of the capture and emission would be manipulated accurately. In this case, if the multi-gate structure were used to form the quantum dot and raise the potential barrier respectively in the channel of the silicon transistor, we could build a quantum dot in the channel which had a similar effect with the interface traps. Therefore, we could be able to observe similar RTSs and control the tunneling rate by manipulating the gate bias and barrier height even in the higher temperature. Then the RTSs generated from this kind of multi-gate transistors could be possible for the application as a random number generation source.

3.6 summary

In summary, we successfully measured two conventional MOSFETs in the low temperatures with the upgraded cryostat system. The mechanism of two types of RTSs was clarified. It showed us a clear RTSs comparison between the oxide trap and interface trap, which helped us build the theoretical basis to design our multi-gate transistor devices for the QRNG source.

Chapter 4 Fabrication of the multi-gate silicon nanowire transistor devices

4.1 Introduction

In this chapter, the multi-gate silicon nanowire transistors were successfully fabricated by collaborating with the senior researcher in our group, using the Southampton Nanofabrication centre. The fabrication process flow was demonstrated. Strong quantum confinement was observed in the initial electrical measurement.

4.2 Fabrication process flow

The fabrication started from 100 nm standard <100> SOI wafer with 145 nm-thick buried-oxide (BOX). The main steps included patterning silicon nanowire, gate oxide formation, dopant diffusion, patterning First gate, patterning Top gate, and patterning metal contact. The flow chart was shown in Figure 4.1.

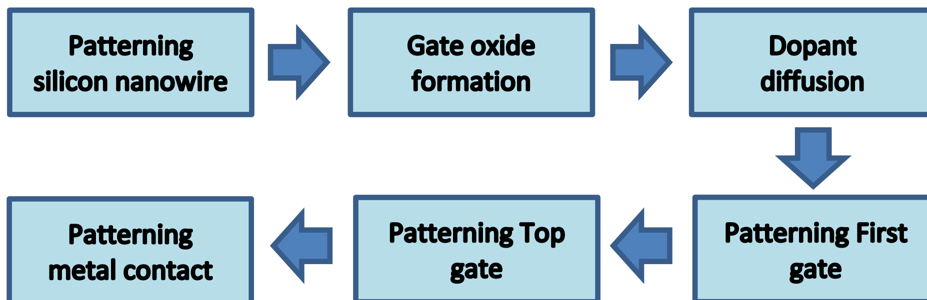


Figure 4.1 The fabrication process flow chart

4.2.1 Patterning silicon nanowire

The typical bottom-up processes were used to define the silicon nanowire [136]. The 100 nm Si layer was thinned down to 24 nm firstly. The thin-down process was combined with dry oxidation in the furnace at 1000 °C and HF (1:7) etching to remove the SiO_2 for several cycles. After the last time furnace oxidation, 24 nm silicon layer remained with 21 nm SiO_2 layer on the top. The SiO_2 layer formed the hard mask to pattern the silicon nanowire. In order to overcome the limitation of the patterning size by E-beam lithography, a “double mask” method was used to reduce the silicon nanowire patterning size. Firstly, we coated the hard mask layer with 30 nm thickness Hydrogen Silsesquioxane (HSQ). Then the HSQ was patterned by E-beam lithography, shown in Figure 4.2.

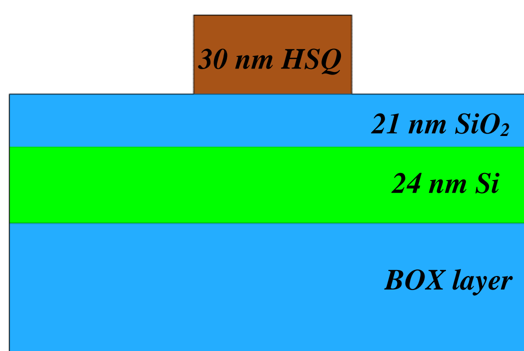


Figure 4.2 Schematic diagram after E-beam lithography

Then Reactive ion etching (RIE) was used to etch the SiO_2 hard mask layer. The etching rate of the HSQ to SiO_2 was 2:1. In this case, 3.6 nm SiO_2 layer was remained after RIE, shown in Figure 4.3.

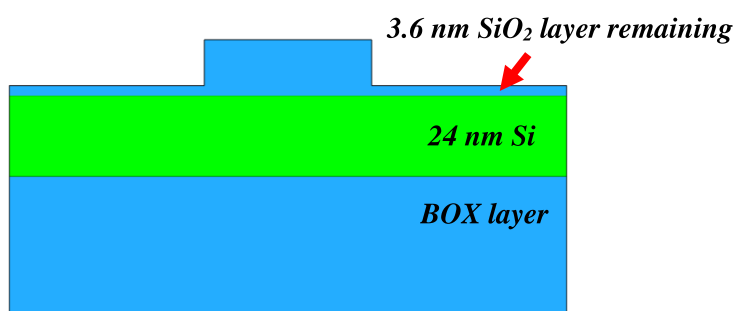


Figure 4.3 Schematic diagram after RIE

To remove the remaining SiO_2 layer and shrink the hard mask patterning, low concentration HF (1:200) etching was proceeded to form the final hard mask for the nanowire region. The hard mask was estimated as 5 nm thickness, shown in Figure 4.4.

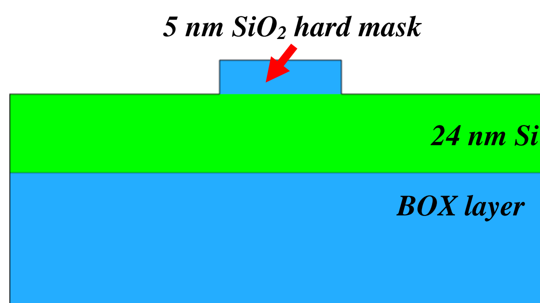


Figure 4.4 Schematic diagram after HF etching

After cleaning by fuming nitric acid (FNA), the Tetramethylammonium Hydroxide (TMAH) etching was carried out to form the triangle or trapezoid nanowires, shown in Figure 4.5. 25% TMAH etching with Isopropyl alcohol (IPA) was used to do 1500% over etching. Long-time over etching was benefit to get the atomically flat interfaces on the silicon nanowire. Due to the long-time over etching, the

undercut was formed under the SiO_2 hard mask, which also could be seen in the scanning electron microscope (SEM) images (Figure 4.6).

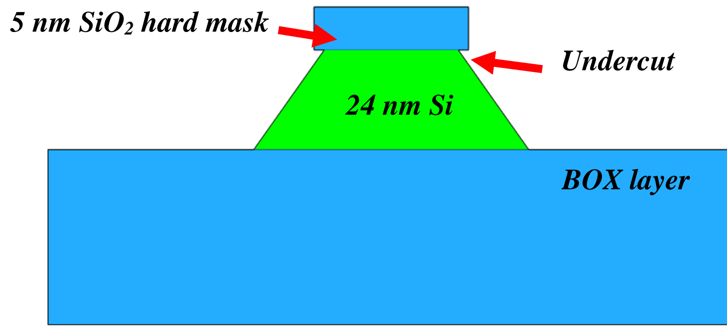


Figure 4.5 Schematic diagram after TMAH etching and undercut part

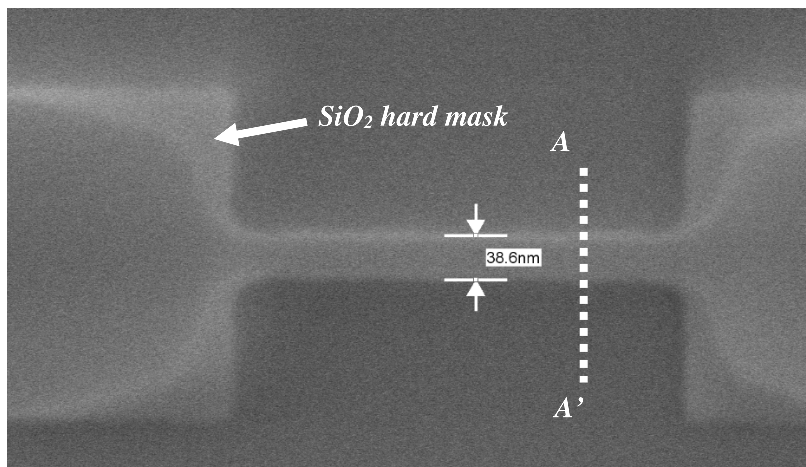


Figure 4.6 SEM view of the nanowire after TMAH etching

The initial HSQ patterning width in Figure 4.2 was 50 nm. From the SEM image, the nanowire width was significantly reduced by the "double mask" method. The silicon nanowire should be a trapezoid in the AA' cross-section in Figure 4.6. With the hard mask, it was difficult to see the edges of the nanowire clearly. The hard mask was removed before gate oxide formation, in order to avoid affecting the uniformity of the gate oxide layer. HF etching was used to remove the hard mask. Without the hard mask, we measured the nanowire width in the SEM (Figure 4.7). Base on the theoretical estimation of the TMAH, the top base of the trapezoid should be around 22 nm and the bottom base should be around 48 nm. It was found that the actual value was quite close to the theoretical value in Figure 4.7.

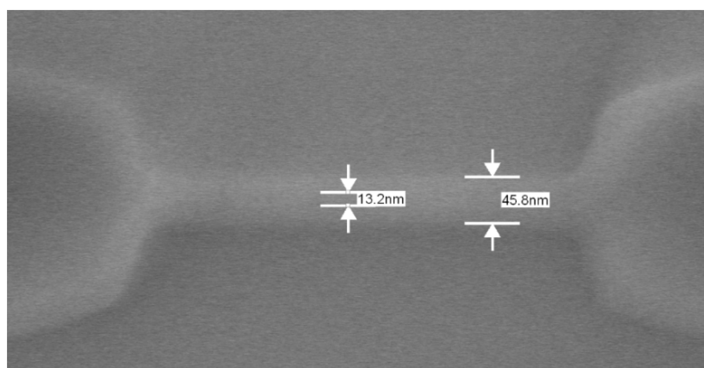


Figure 4.7 SEM view of the nanowire after removing the hard mask

4.2.2 Gate oxide formation

The bare silicon nanowires were oxidized in the furnace at 1000°C to form the 20 nm gate oxide layer. During this process, 9 nm silicon was oxidized from <100> direction. After oxidation, it was unable to measure the nanowire width in the SEM. The TCAD software was used to simulate the process and estimate the geometry. In Figure 4.8-4.10, the nanowire dimensions that had the initial HSQ mask width at 30 nm, 50 nm, and 75 nm were extracted from the simulation results.

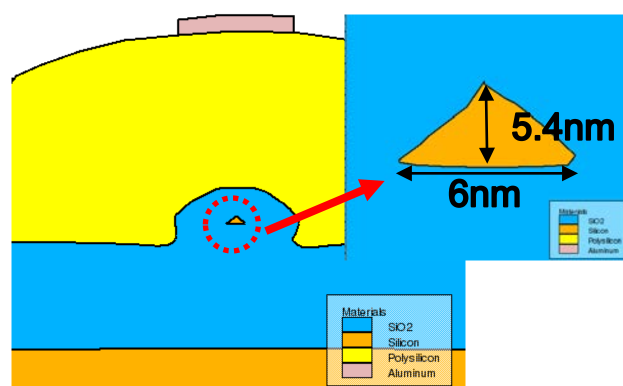


Figure 4.8 The nanowire dimensions from the TCAD simulation (HSQ mask: 30 nm)

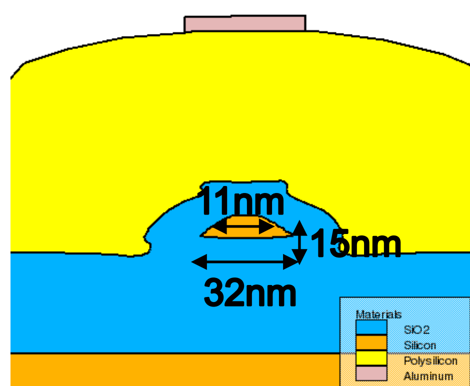


Figure 4.9 The nanowire dimensions from the TCAD simulation (HSQ mask: 50 nm)

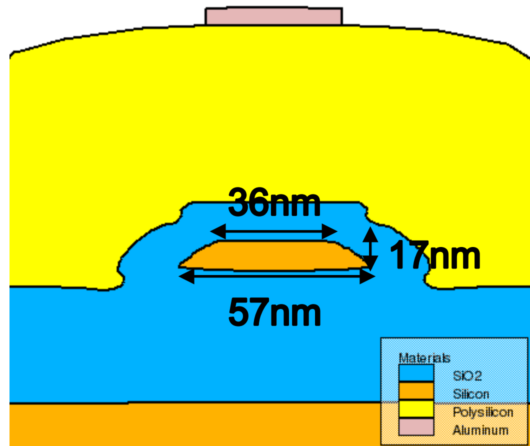


Figure 4.10 The nanowire dimensions from the TCAD simulation (HSQ mask: 75 nm)

4.2.3 Dopant diffusion

The raised source/drain technique was used to pattern the source, drain and First gate and the same time. As all the silicon surface was covered by SiO_2 layer, dopant diffusion windows should be open at first. The PMMA950 resist was used to pattern the diffusion windows by E-beam lithography. Then RIE was proceeded to remove the SiO_2 layer. In this process, the etching rate and etching selectivity were well controlled to reduce the damage to the silicon layer underneath. The diffusion windows could be clearly seen in the SEM image (Figure 4.11).

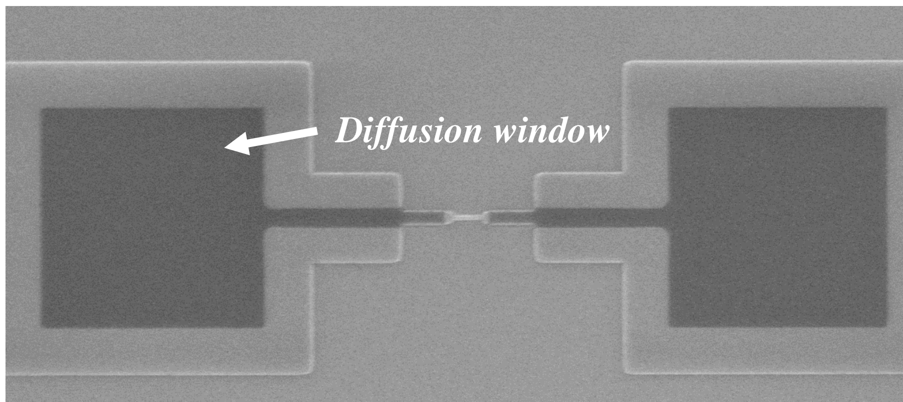


Figure 4.11 SEM image for the diffusion windows location

We did not dope the silicon directly after open the diffusion windows. The Poly-Si layer was deposited by the low-pressure chemical vapour deposition immediately (LPCVD) after using the diluted HF to remove the surface native oxide, shown in Figure 4.12.

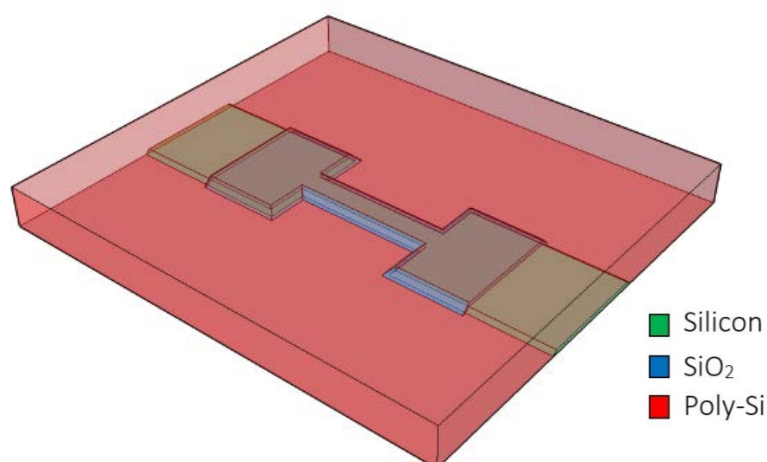


Figure 4.12 Schematic diagram after depositing the Poly-Si layer

Then the Poly-Si layer was doped and the dopant was activated in the Rapid thermal anneal (RTA) process for 1 minute at 950 ℃ with N_2. The Poly-Si was fully metalized and the dopants diffused to the source and drain.

4.2.4 Patterning the First gate.

In the raised source and drain method, the First gates were also patterned at the same time. 100 nm thickness HSQ was used to pattern the source, drain and the First gate by E-beam lithography. The thick HSQ layer became the hard mask in the following Inductively Coupled Plasma Etching (ICP) etching to define the source, drain and the First gate, shown in Figure 4.13. The patterns were also confirmed in the SEM image (Figure 4.14).

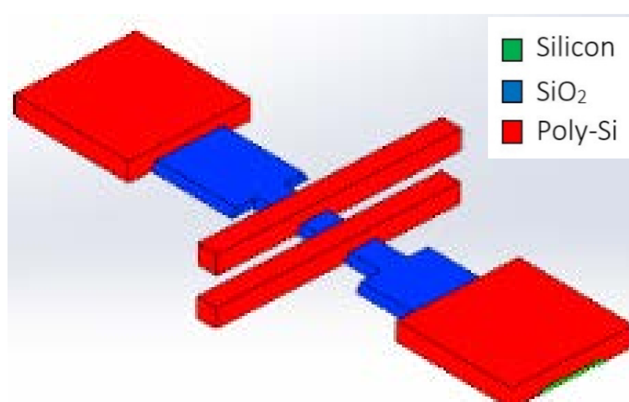


Figure 4.13 Schematic diagram after ICP etching

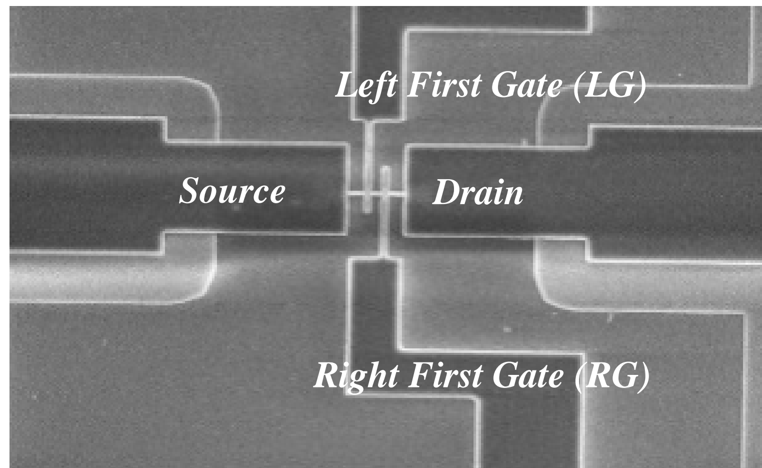


Figure 4.14 SEM image after ICP etching

In order to insulate the Top gate layer to the source, drain and First gate layer, the Poly-Si surface was oxidized by RTA for 3 minutes at 950℃ with O_2 after removing the surface thin oxide layer by dilute HF. 9 nm thick thermal oxide was grown.

4.2.5 Patterning Top gate

Another Poly-Si layer was deposited by LPCVD after oxidizing the First gate surface. With the same process in the last step, the Poly-Si layer was doped and annealed by RTA. After removing the remained dopant by HF, the Top gate was patterned using 100 nm thickness HSQ by E-beam lithography. The high selectivity ICP etching was applied to define the Top gate, shown in Figure 4.15. The Top gate layer was also confirmed by an optical microscope in Figure 4.16.

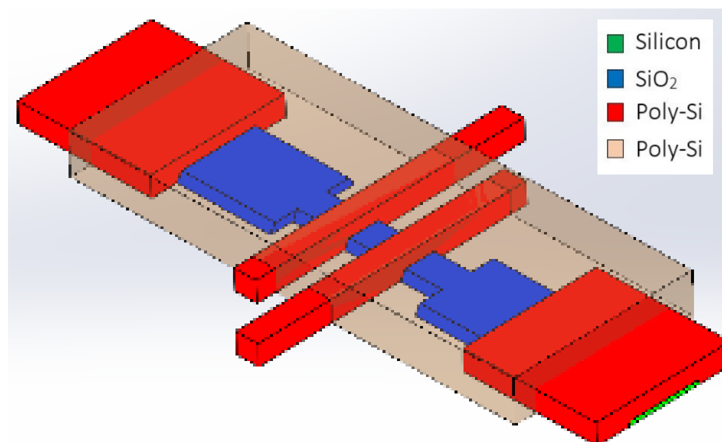


Figure 4.15 Schematic diagram after patterning the Top gate

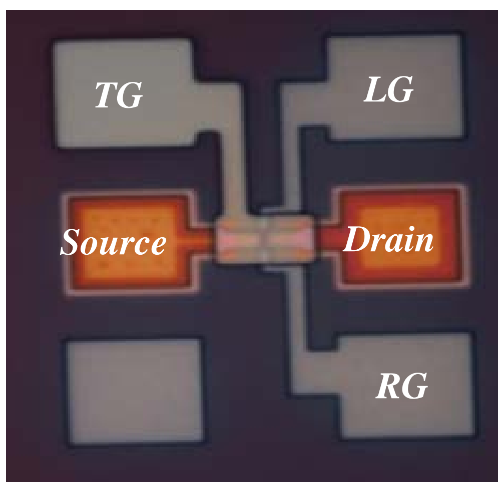


Figure 4.16 Optical image after patterning the Top gate

To protect the device region and avoid contamination, 250 nm oxide layer was deposited to cover all the surfaces by Plasma-enhanced chemical vapour deposition (PECVD).

4.2.6 Patterning metal contact

In order to get the metal contact with the source, drain, First gate, and Top gate electrodes, we need to open the windows on the 250 nm PECVD oxide layer. The Zeon electron-beam positive-tone (ZEP) resist was applied to pattern the windows on the PECVD oxide layer by E-beam lithography. Then the well-controlled HF wet etching was used to remove the oxide layer and open the windows. The optical image in Figure 4.17 showed that the contact windows were opened successfully.

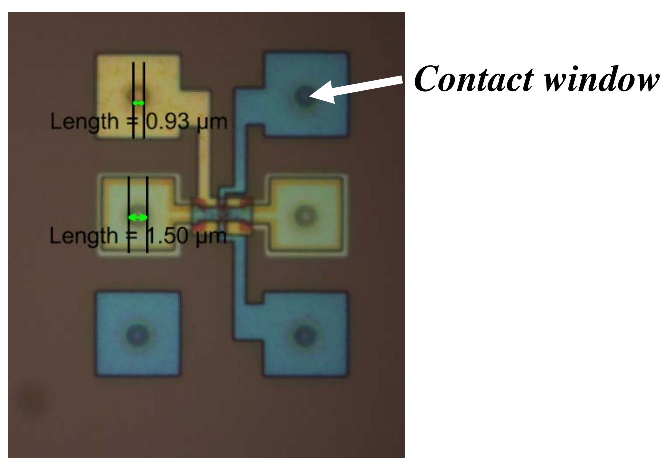


Figure 4.17 Optical image for the contact windows

Then the lift-off process was applied to form the metal layer. The 300 nm thickness PMMA/MMA double-layer resist was used to pattern the metal layer by E-beam lithography. After removing the native oxide on the bare Poly-Si surface by HF etching, the E-beam evaporation was utilized to deposit 20 nm Titanium and 180 nm Aluminium. Subsequently, the chip was put into the N-Methyl-2-Pyrrolidone (NMP) with the ultrasonic bath to accelerate the lift-off process. The last step was annealing in the furnace for 7 minutes at 450°C in H_2/N_2 to increase the contact quality. The optical images in Figure 4.18 showed the final device.

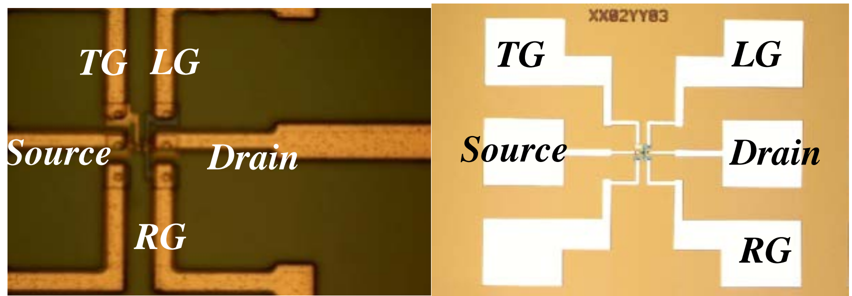


Figure 4.18 (Left) Optical image for the metal contact. (Right) Optical image for one device

Some part of the metal layer was damaged after the lift-off process, shown in Figure 4.19. One possible reason was the ultrasonic wave. Another possible reason was that the resist height in the lift-off process might not be high enough. And this problem reduced the yield of the device.

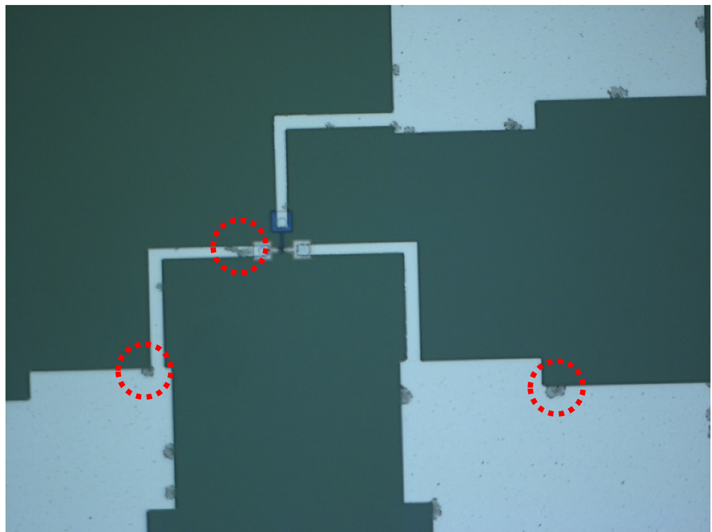


Figure 4.19 Optical image for the damaged metal layer.

4.3 Quantum confinement in the silicon nanowires

Before going to the main devices, the accessory part in the same chip was measured firstly. This accessory part included silicon nanowire transistors with one First gate and no Top gate. The schematic diagram was shown in Figure 4.20. The optical microscope image was shown in Figure 4.21 with metal contact.

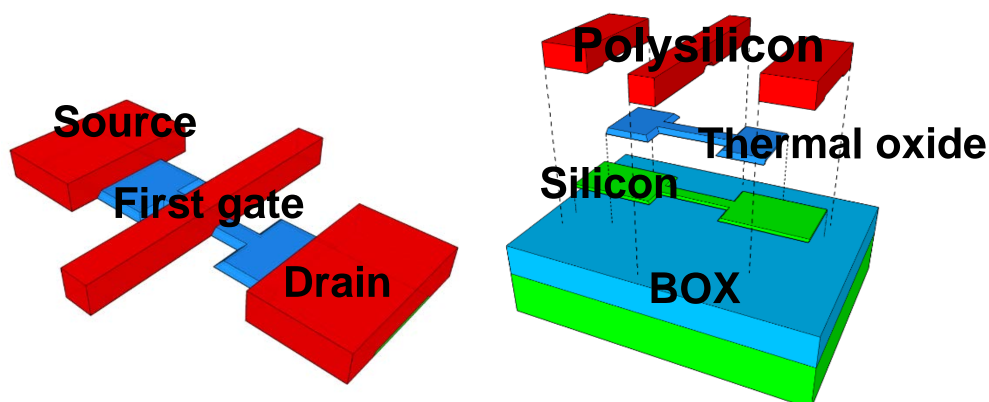


Figure 4.20 The schematic diagram of the silicon nanowire transistor and the layer structure.

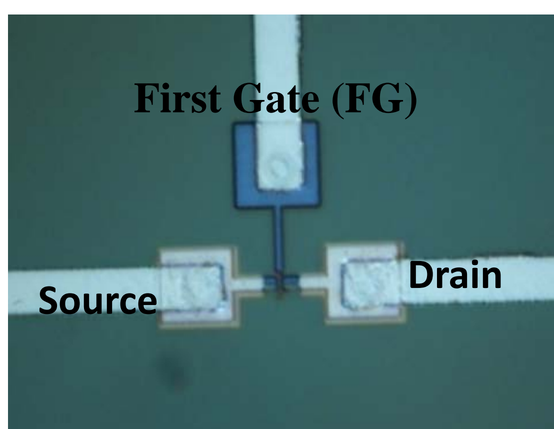


Figure 4.21 The optical image of the silicon nanowire transistor.

The details of the device design and structure were shown in Figure 4.22. The first green layer was single-crystal silicon. The initial design widths (W) for the HSQ patterning varied from 10 nm to 200 nm. The First gate length (L) was fixed to 100 nm in these transistors. A thermal oxide layer covered the silicon layer. The purple layer shows the dopant diffusion windows the source and drain. The red layer was the polysilicon layer over the silicon nanowire. The grey layer was the metal layer. As these transistors had the same processes as the main devices, the source and drain were n-doped. We expected that the dopant diffusion distance could reach 300 nm. If not, the channel would be difficult to open by the First gate.

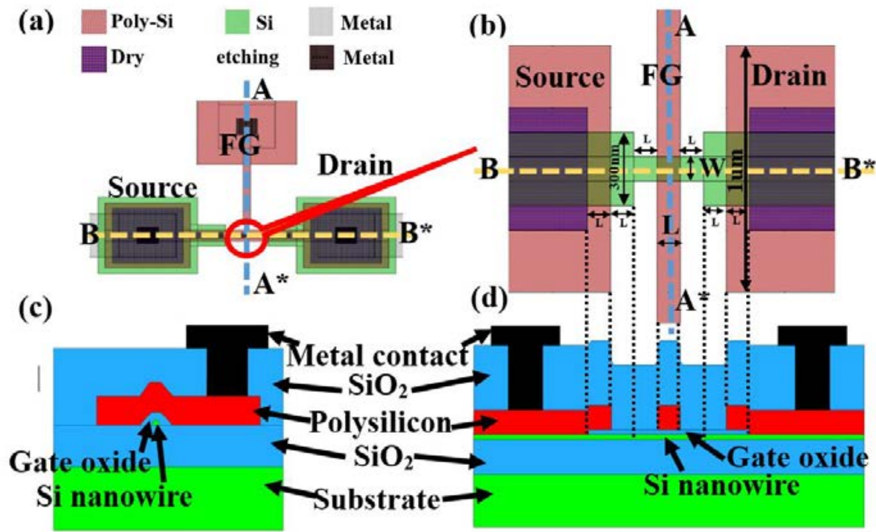


Figure 4.22 (a) The planar view of the device region. (b) The planar view of nanowire region. (c) The cross-sectional view in AA direction (blue dash line). (d) The cross-sectional view in BB* direction (yellow dash line).*

4.3.1 Measurement and discussion

In order to investigate the transport properties of these narrow-channel silicon nanowire transistors with atomically flat interfaces, the devices were measured at room temperature by using Agilent B1500 and Cascade M150 probe station at room temperature. The measurement system has been calibrated with the background noise level below 100fA. The V_{FG} was swept from -2 V to 1V while the V_D was applied at 50 mV and 1V for each of the devices respectively. Due to the significant shrinking in the process, the difference between the HSQ pattern width (W) and the actual nanowire dimensions was considerable. As the cross-section of the nanowire should be a triangle or trapezoid, we used the width of final lower base (W_L) to describe the width dimension of the nanowires. The W_L was extracted from the process simulation results by TCAD as we showed in the process part. From the conventional drain current (I_D) versus First gate voltage (V_{FG}) measurement results, when the W were less than 30 nm, there was no current from source to drain. Considering the fabrication process, the silicon nanowires with W=20 nm should have been fully oxidized and no silicon channel should exist. It made the agreement with the simulation result by TCAD. The W_L was calculated as from 181 nm down to 6 nm. The conventional drain current (I_D) versus First gate voltage (V_{FG}) characteristics were observed from the measurement results (Figure 4.23).

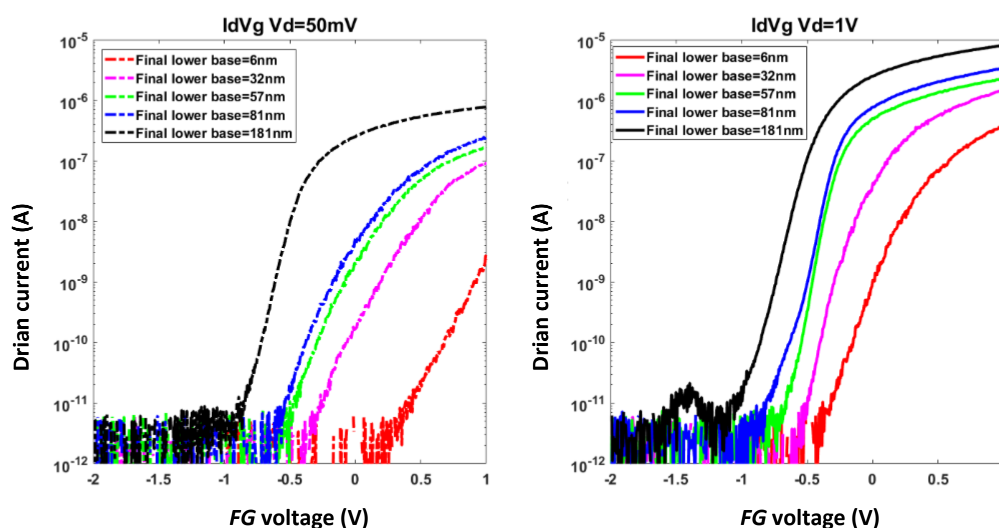


Figure 4.23 Drain current versus First gate voltage characteristics on drain voltage at 50mV and 1V with different channel width.

From the measurement results in Figure 4.23, the threshold voltage (V_{th}) was significantly shifted to the right both on V_D=50 mV and V_D=1 V. In order to see the quantitative shift, we extracted the V_{th} from the raw data (Figure 4.24). The V_{th} of these silicon nanowire FETs was defined as the voltage at which the drain current is larger than 1e-10 A [137]. The increment of the V_{th} at low drain bias was found to be much larger than that at the high drain bias.

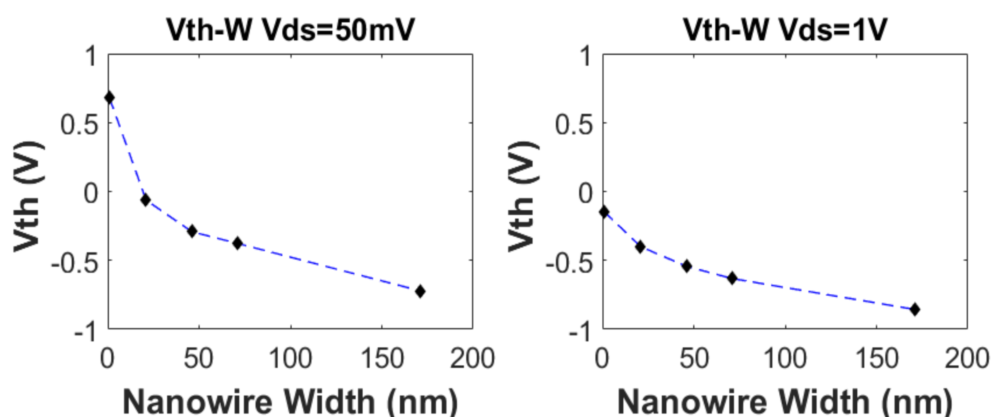


Figure 4.24 Threshold voltage of the silicon nanowire transistors with different width.

The reason why V_{th} was increased by scaling the nanowire diameter was considered to the quantum confinement effect in the ultra-thin and ultra-narrow silicon devices [120, 138]. In the classic planar FETs with a wide channel, the formation of the channel occurred in a large enough space, in which the carriers would not be confined in the channel scale. Classically, the V_{th} was independent of the width of the channel. The short channel effect was mainly degraded the performance of the transistors. However, in the scaled silicon nanowire devices, the dimension of the channel cross-section was equivalent to the mean free path of the electrons in silicon. It also implied that the channel dimensions were approaching the Bohr radius in silicon. The electrons could be confined

in the scaled channel. The mean free path was the average travel distance by a particle between successive collisions. The electrons mean free path in the silicon could be calculated as:

$$\lambda = v_{average}\tau_c \quad (4.1)$$

The $v_{average}$ was the average thermal velocity. The typical thermal velocity at room temperature of electrons in silicon was around 10^7 cm/s, which was over the drift velocity in silicon devices [120]. The τ_c was the collision time, which could be calculated by the mobility equation:

$$\mu = \frac{q\tau_c}{m^*} \quad (4.2)$$

The electron mobility μ in silicon was roughly considered to be 1400 $cm^2V^{-1}s^{-1}$ [120]. The electron effective mass in silicon was about $0.2m_0$, where $m_0 = 9.11 * 10^{-31}kg$ (the free electron rest mass). Then, the electron mean free path in the silicon at room temperature could be estimated as:

$$\lambda = v_{average}\tau_c = v_{average} * \frac{qm^*}{\tau_c} \approx 16 \text{ nm} \quad (4.3)$$

According to the nanowire dimensions from the TCAD simulation results, both of the thickness and the width in the cross-section was approaching to the electron mean free path or even smaller. So that the electrons were confined in the scaled direction. As the silicon nanowire was scaled into ultra-thin and ultra-narrow, it became a quantum well or even a quantum wire. For a quantum well, the electron was confined in $\hat{z}$ direction and was free to move in $\hat{x}$-$\hat{y}$ plane. For the quantum wire, the electron was confined in in $\hat{x}$-$\hat{y}$ plane and was free to move in $\hat{z}$ direction. To describe the confined carriers in such structures, the three-dimensional effective mass equation should be solved [138]:

$$-\frac{\hbar^2}{2m^*}\nabla^2 F(r) + E_{C0}(r)F(r) = EF(r) \quad (4.4)$$

The $F(r)$ was the envelop function, E_{C0} was the original energy of the bottom of the conduction band. E was the total energy which should be calculated. For the quantum well, carriers could move freely in $\hat{x}$ - $\hat{y}$ plane. So that the solution of the form could be:

$$F(r) = \phi(z)\frac{e^{ik_x x}e^{ik_y y}}{\sqrt{L_x L_y}} \quad (4.5)$$

After substituting equation 4.5 into equation 4.4, the equation for $\phi(z)$ could be found:

$$\frac{d^2\phi(z)}{dz^2} + k_z^2\phi(z) = 0 \tag{4.6}$$

As,

$$k_z^2 = \frac{2m^*}{\hbar^2}[\varepsilon - E_{C0}(z)] \tag{4.7}$$

and,

$$\varepsilon = E - \frac{\hbar^2}{2m^*}\left(k_x^2 + k_y^2\right) \tag{4.8}$$

ε should be the energy associated with the confinement in $\hat{z}$ direction due to the kinetic energy $\left[\frac{\hbar^2}{2m^*}\left(k_x^2 + k_y^2\right)\right]$ was associated with motion in $\hat{x}$-$\hat{y}$ plane. If the quantum well was deep enough, the $\phi(z)$ could be given by the infinite well solution as:

$$\phi(z) = \sqrt{\frac{2}{W}}\sin k_z z \tag{4.9}$$

Where

$$k_z = \frac{n\pi}{W} \tag{4.10}$$

So that if the electron was moving in the $\hat{x}$-$\hat{y}$ plane, its total energy is:

$$E = E_{C0} + \frac{\hbar^2 k_z^2}{2m^*} + \frac{\hbar^2}{2m^*}\left(k_x^2 + k_y^2\right) \tag{4.11}$$

Or

$$E = E_{C0} + \varepsilon_n + \frac{\hbar^2}{2m^*}\left(k_x^2 + k_y^2\right) \tag{4.12}$$

Where ε_n is the energy that was quantized according to

$$\varepsilon_n = E(k_n) = \frac{\hbar^2 k_n^2}{2m_0} = \frac{\hbar^2 n^2 \pi^2}{2m_0 W^2} \quad n = 1,2,\ldots \tag{4.13}$$

Carriers in the quantum wires could be solved in a similar way. Instead of equation 4.4, the wave function could be written as:

$$F(r) = \phi(x,y)\frac{e^{ik_z z}}{\sqrt{L_z}} \tag{4.14}$$

Supposed that the confinement potential was infinite,

$$\phi(x,y) = \sqrt{\frac{2}{W_x}}\sin k_x x\sqrt{\frac{2}{W_y}}\sin k_y y \tag{4.15}$$

with,

$$k_x = \frac{n_x \pi}{W_x}, k_y = \frac{n_y \pi}{W_y} \tag{4.16}$$

The energy level of the bottom of the sub-bands in the quantum wire could be:

$$\varepsilon_{n_x,n_y} = \frac{\hbar^2\pi^2}{2m^*}\left(\frac{{n_x}^2}{{W_x}^2} + \frac{{n_y}^2}{{W_y}^2}\right) \tag{4.17}$$

So, if the electron transported in $\hat{z}$ direction, the total energy was:

$$E = E_{C0} + \varepsilon_{n_x,n_y} + \frac{\hbar^2 k_z^2}{2m^*} \tag{4.18}$$

According to results from the equation in 4.12 and 4.18, the electrons would be in a higher energy level when moving in the quantum well or quantum wire. It also implied that the bandgap in the scaled nanowire channel was expanded theoretically. Moreover, the bandgap should have a larger expansion in the smaller dimension, shown in Figure 4.25.

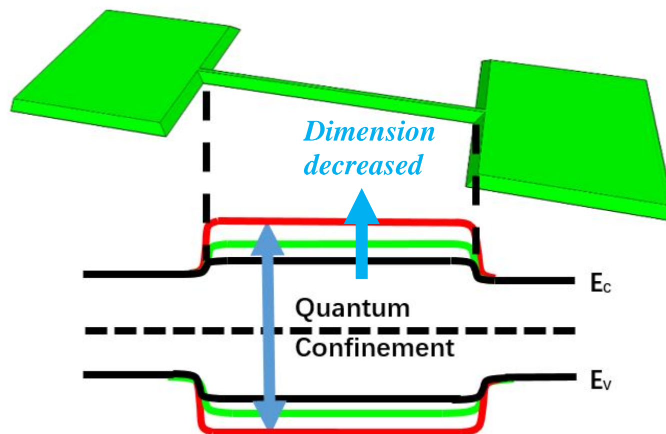


Figure 4.25 Quantum confinement in the smaller dimensional silicon devices.

We summarised both the simulation and experiment results from the literature [137-141] and extracted the relationship between the nanowire diameter and the nanowire energy bandgap experimentally (Figure 4.26).

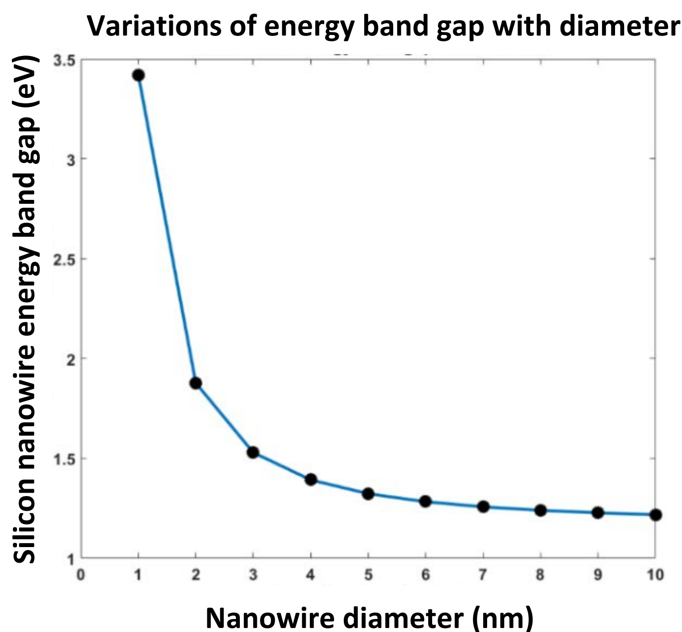


Figure 4.26 The diameter dependence of the energy bandgap in a silicon nanowire by fitting the results from the literature.

We supposed that the quantum confinement mainly came from the thickness direction of the channel in our silicon nanowire transistors. As the shape of cross-sections should be triangle or trapezoid, we estimated an average thickness of each nanowire based on the TCAD dimension simulation results. By using the bandgap expansion data from Figure 4.26, we built a 2D transistor model in TCAD to investigate the relationship between the bandgap increase and the V_{th} shift qualitatively (Figure 4.27).

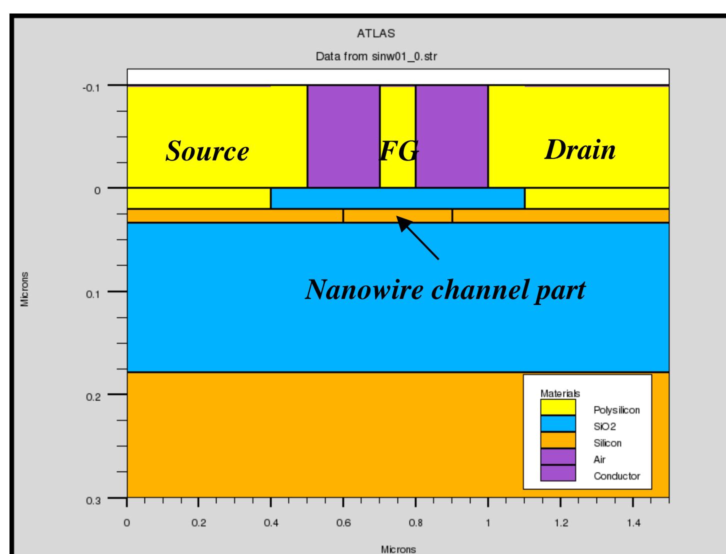


Figure 4.27 2D transistor model in TCAD.

The nanowire region was in the middle of the 2D transistor model in Figure 4.27, where the bandgap of the affinity parameters was modified according to the average thickness of each device. The dimensions for the other parts were exactly the same as our real derives. The simulation results were shown in Figure 4.28.

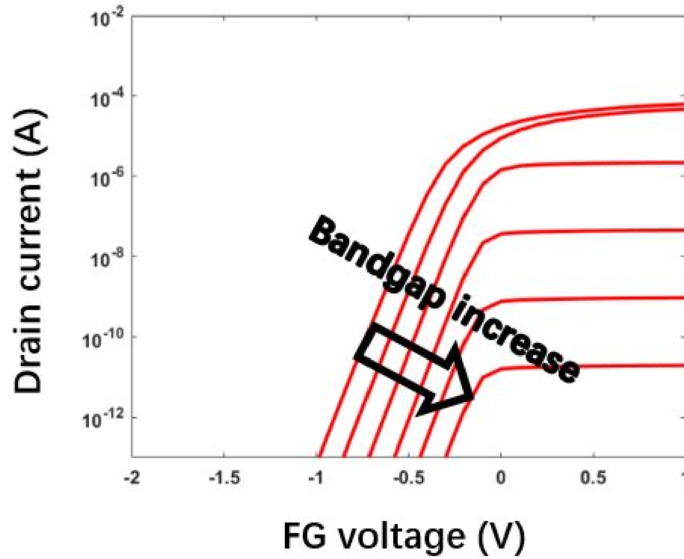


Figure 4.28 I_D versus V_{FG} characteristics on V_D=50 mV with different bandgaps and affinity.

From the simulation results, it could be confirmed that the V_{th} shift was mainly caused by the quantum confinement induced bandgap increase. However, the subthreshold slope (*SS*) was degraded from wide nanowire to narrow nanowire and this phenomenon happened neither in our simulation nor in the literature [140]. We supposed that the *SS* degradation came from the actual geometry of the nanowires. In our simulation model, the thickness was defined by an estimated average thickness. However, when the nanowire came narrower, the actual quantum confinement could not be estimated by the average thickness. The triangle regions on the left and right side of the nanowire would have the equivalent current contribution to the top surface channel. As the thickness in the triangle regions was decreased into zero, larger bandgap expansion was expected (Figure 4.29). In this point of view, multi-channels in different energy levels would be opened in sequence and contribute to the total current together. In order to investigate which part of the channel would be inversed by the gate voltage, we simulated the electric field distribution with certain gate bias in the cross-section of the channel. The results were shown in Figure 4.30-4.32.

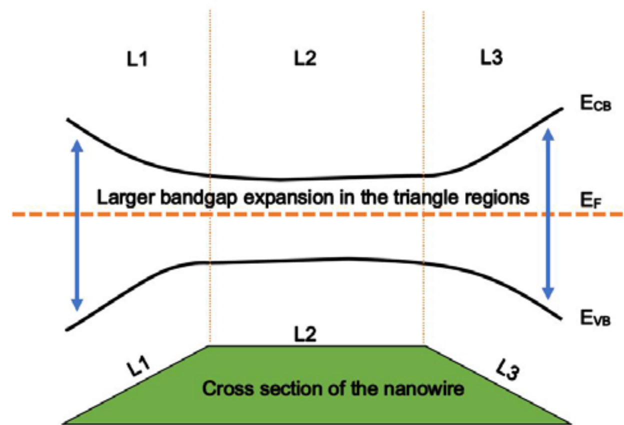


Figure 4.29 The schematic diagram to show the bandgap expansion in different part of the nanowire channel from the cross section of view

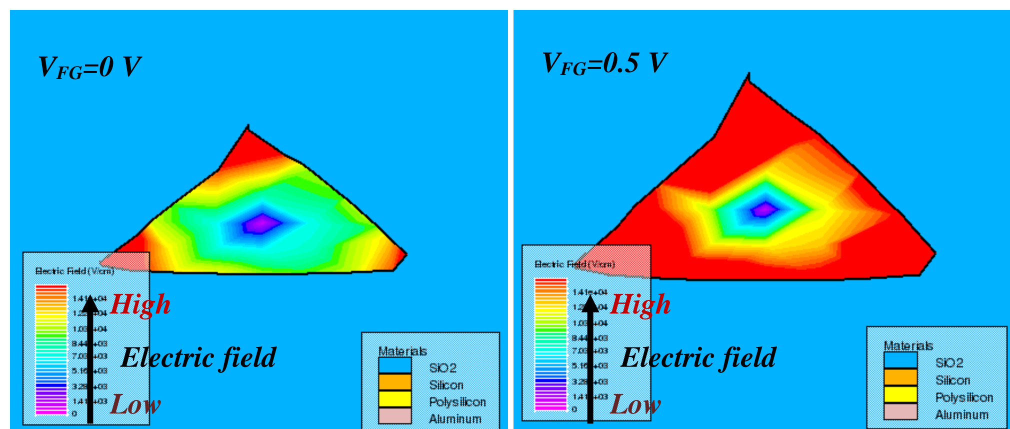


Figure 4.30 The electric field distribution in the cross section of the nanowire from the TCAD simulation (HSQ mask: 30 nm). In the left, V_{FG}=0 V. In the right, V_{FG}=0.5 V

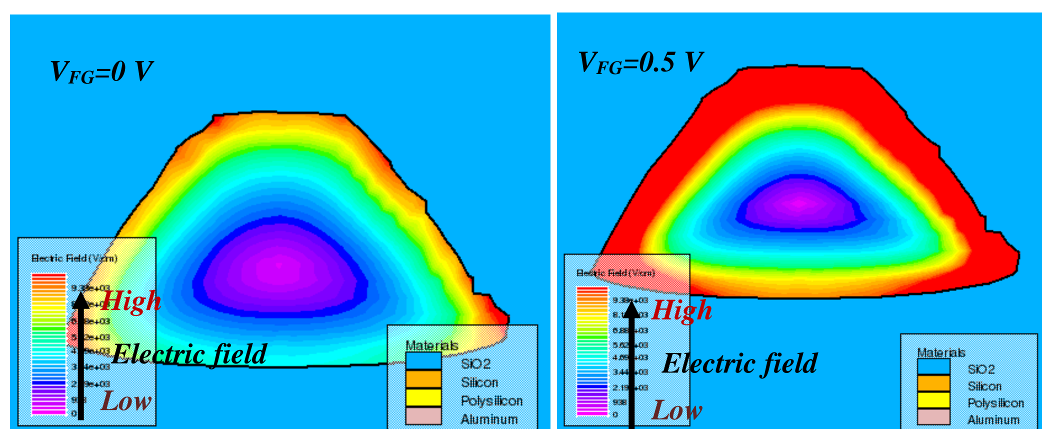


Figure 4.31 The electric field distribution in the cross section of the nanowire from the TCAD simulation (HSQ mask: 50 nm). In the left, V_{FG}=0 V. In the right, V_{FG}=0.5 V

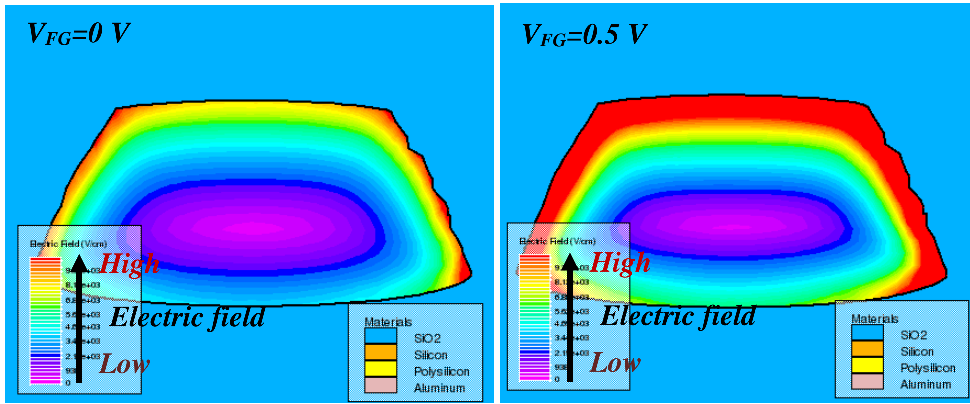


Figure 4.32 The electric field distribution in the cross-section of the nanowire from the TCAD simulation (HSQ mask: 75 nm). In the left, V_{FG}=0 V. In the right, V_{FG}=0.5 V

From the simulation results, it could be found that the electric field was higher in the corners and two side surfaces than the top surface at the same gate bias (V_{FG}=0 V). At V_{FG}=0 V, the channel was still in the subthreshold region. Although the bandgap at the corner and the side surfaces were much larger due to the smaller dimension, the channel in the corner or side surfaces that connected source and drain might be opened much earlier due to the strong electric field. In this case, the multi-channels in different energy levels were not opened in sequence based on the bandgap increase. It was also determined by the cross-section geometry, which affected the electric field distribution. In the smaller geometry part like in the corner, the V_{th} could be decreased by the strong electric field distribution. It was quite difficult to compare that in what extent that the strong electric field could shift the V_{th} to the left, and to what extent the bandgap increase could shift the V_{th} to the right in the 2D simulations. If it were supposed that channels in the corner and the side surfaces would be opened first due to the strong electric field, we could build a multi-channel transport model to explain *SS* degradation in the silicon nanowires, shown in Figure 4.33.

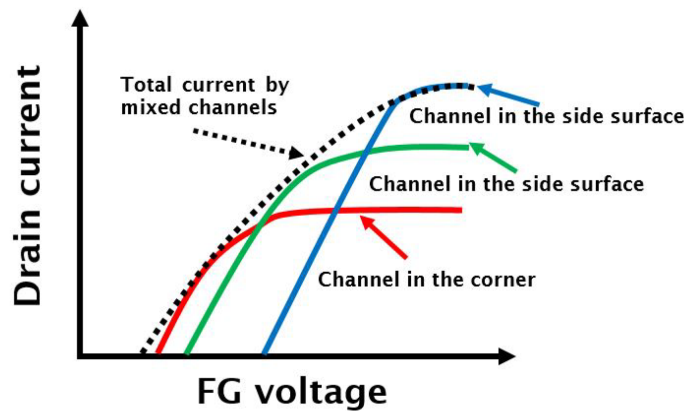


Figure 4.33 The multi-channel transport model to explain the SS degradation

In Figure 4.33, the channel that had a larger bandgap in the corner (the red solid line) was opened firstly. Because of the large bandgap increase, the ON current of this channel was limited, which

could be confirmed from the simulation results in Figure 4.28. During the V_{FG} was increased, the channel that a medium-large bandgap in the side surface (the green solid line) started to open. The ON current of this channel was larger than the previous one due to the smaller bandgap increase. Then the green line would be over the red line and dominated the total channel current. Finally, the late opened top surface channel (the blue solid line) would be over the green solid line again. For simplifying the model, we divided the total nanowire channel into three parts. Actually, the bandgap changing in these three parts was continuous as shown in Figure 4.29. In this case, it could be considered as thousands of channels in between the red solid line, green solid line and blue solid line. Then the total current mixed with these channels would be the black dashed line in Figure 4.33. Unfortunately, we were unable to use the 2D TCAD to simulate the hypothesis. However, by comparing the V_{th} shift and *SS* degradation in low and high drain voltages from Figure 4.23, it could be found that the V_{th} shift and *SS* degradation were suppressed to some extent by the higher energy electrons. The higher energy electrons could use more channels that were in higher energy levels. The total channel was opened more efficiently. In this case, the *SS* degradation was recovered. In addition, the higher drain voltage could cause the DIBL effect to decrease the barrier height. So that the V_{th} shift caused by the quantum confinement would be suppressed. This phenomenon could confirm our multi-channel model from another point of view.

4.4 Summary

In summary, we successfully fabricated the scaled multi-gate silicon nanowire transistors with atomically flat interfaces based on the SOI substrate. The “double mask” method and the combining dry etching and wet etching method were applied to break through the patterning size limitation in the E-beam lithography. The raised source and drain electrodes were patterned together with the First gate electrode. The metal layer damage was found after lift-off process. By measuring the transistor performance of the nanowires in the same chip with the main devices, the quantum confinement effect was observed in these ultra-thin nanowires. The *SS* was also found and explained by a multi-channel transport model. The multi-channel transport model would affect the manipulation of the RTSs in the silicon nanowires in the next two chapters.

Chapter 5 Manipulate the RTSs in the artificial quantum dot with the multi-gate silicon nanowire transistor

5.1 Introduction

In this chapter, we used the multi-gate silicon nanowire transistor to form an artificial quantum dot in the channel by controlling the multi-gate voltages. After the fabrication, 11 devices were picked up according to the transistor performance test, in which the two First gates and the Top gate could turn on/off the channel individually. The First gates raised the potential barriers for the quantum dot. The Top gate weakly inverted the channel and controlled the energy level in the quantum dot. The typical RTSs were observed in a certain bias condition in all the devices, which was generated by the artificial quantum dot. What's more, the probability distribution of the RTSs could be well controlled by the multi-gate structure. Two typical devices were selected to show the RTSs details in chapter 5 and 6. The main results were published in a journal paper [142].

5.2 Device structure

The device 3D structure was shown in Figure 5.1. The two First gates named *LG* and *RG* were 75 nm wide. The Top gate (*TG*) was 125 nm wide. From the cross-section view of the trapezoid nanowire, the upper base was 36 nm and the lower base was 57 nm for the final dimension of the nanowire, which could be confirmed in Figure 4.8. The cross-section of view from source to drain (AA') was shown in Figure 5.2.

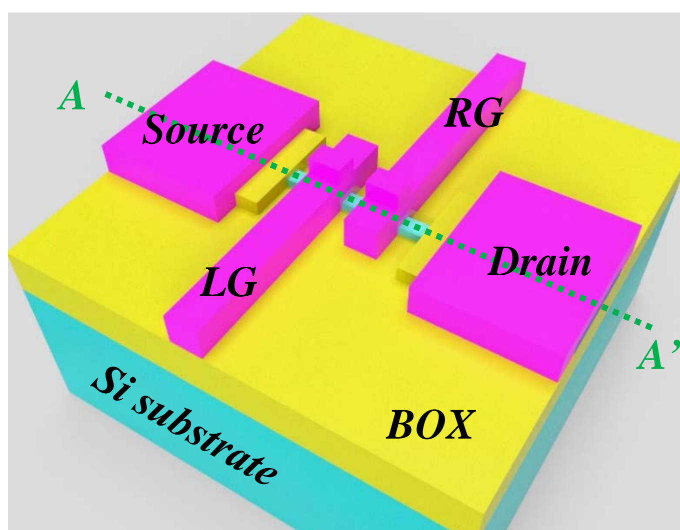


Figure 5.1 The 3D schematic structure of the device. TG was not shown [142].

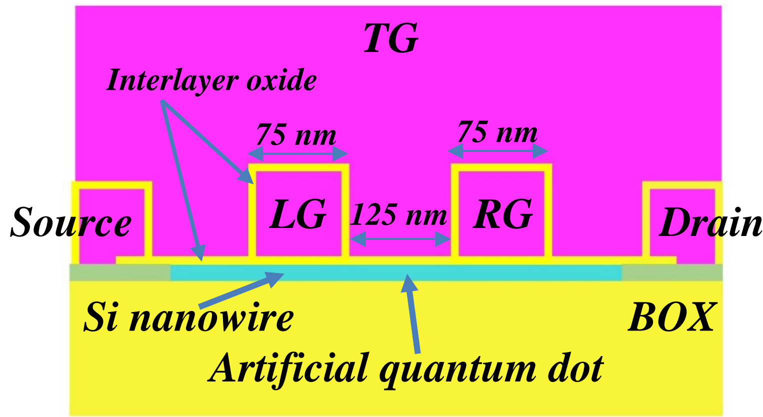


Figure 5.2 The cross section of view from AA' direction in Figure 5.1[142].

5.3 Measurement and results

The devices were measured at room temperature by using Agilent B1500 and Cascade M150 probe station at room temperature. The background noises were confirmed below 100 fA. We checked the two First gates respectively with V_{TG} =0V to confirm that the First gates could raise the potential barriers properly. The I_D versus V_{LG} and V_{RG} characteristics were shown in Figure 5.3 respectively.

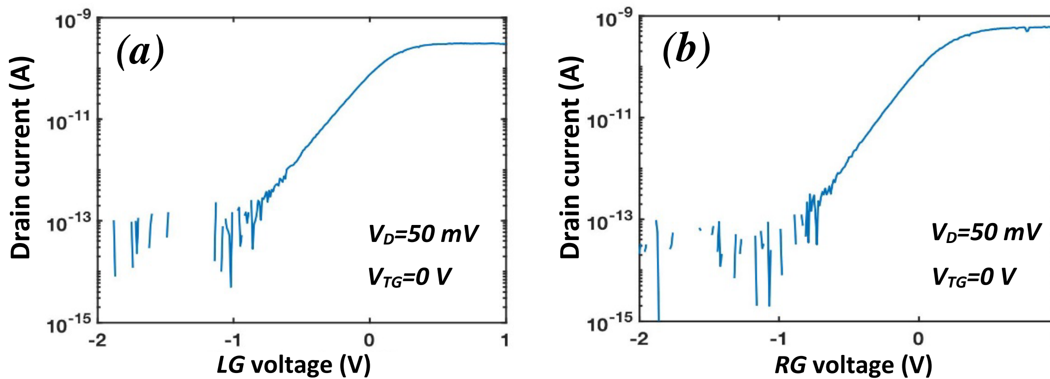


Figure 5.3 (a)The I_D versus V_{LG} characteristics with V_{TG}=0 V and V_{RG} floating at V_D=50 mV. (b) The I_D versus V_{RG} characteristics with V_{TG}=0 V and V_{LG} floating at V_D=50 mV [142].

The slight asymmetry was mainly for two reasons. One was that the *RG* was close to the drain. The other one was that the process variations in E-beam lithography and poor metal patterning could also degrade the symmetry of *LG* and *RG*. As the threshold voltage and subthreshold slope for both of them were almost the same, and the ON currents were in the same order of magnitude, the slight asymmetry could be negligible for the two First gates to raise the barrier.

In order to form the same height of the potential barriers on both sides of the artificial quantum dot, we applied the same voltage on *LG* and *RG* as V_{LG}= V_{RG}. The fast ramped gate voltage sweeping method was used to capture the overall characteristics of the RTSs. The I_D versus V_{TG} characteristics with different V_{LG}= V_{RG} were shown in Figure 5.4.

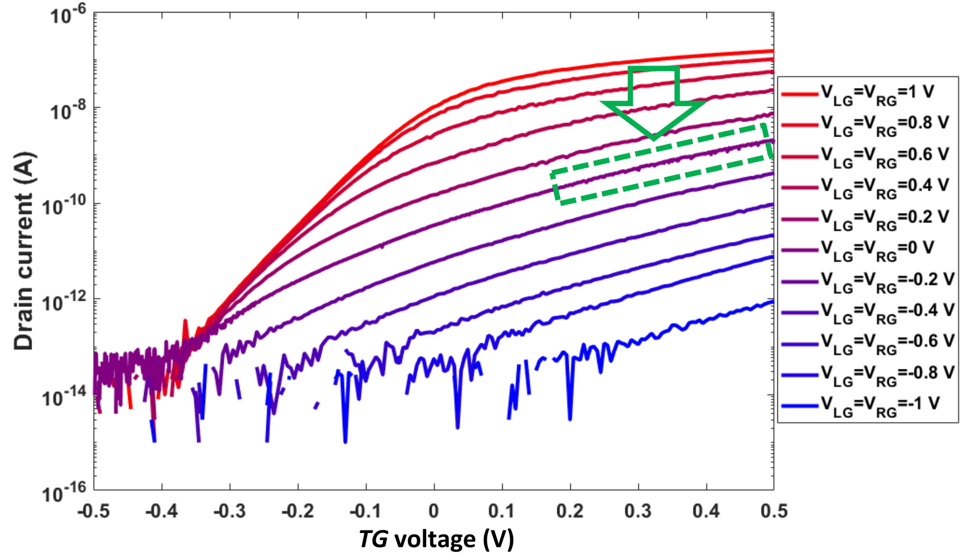


Figure 5.4 (a) The I_D versus V_{TG} characteristics with different $V_{LG}= V_{RG}$ at V_D=50 mV in log scale [142].

By decreasing the $V_{LG}=V_{RG}$ bias, the potential barriers in the channel were raised. Then the current flow was found to be limited as we expected. In addition, the RTSs were observed in certain areas. The marked area in Figure 5.4 was zoomed in to show the details in Figure 5.6. We extracted the threshold voltage shift (ΔV_{th}) which was about 5 mV. We supposed that it was caused by the single electron trapped and de-trapped in the artificial quantum dot. According to the equation [119],

$$C_{QD} = \frac{e}{\Delta V_{th}} \tag{5.1}$$

The e was the elementary charge. The C_{QD} was capacitance of the quantum dot, which could be calculated as,

$$C_{QD} = \frac{\varepsilon_0 \varepsilon_{ox} S}{t_{ox}} \tag{5.2}$$

The ε_0 was the dielectric constant of the vacuum. The ε_{ox}=3.9 was the relative dielectric constant of the oxide layer. The S was the quantum dot area. In our device, the surface area of the artificial quantum dot should include the top surface as well as the two side-surfaces as we discussed in chapter 4. Then the ΔV_{th} was calculated as 10 mV theoretically. Considering that the positive charges in the Poly-Si gate could screen the charge effect of the trapped electron, it leads a small effect to the ΔV_{th} [120]. The theoretical result (10 mV) was similar to the measurement result (5 mV). In this case, we could believe that the RTSs were generated by the electrons trapping and de-trapping from the artificial quantum dot in the multi-gate silicon nanowire transistor. The charging energy could be estimated by $E_c = e^2/2C_{QD} \approx 5\ meV$, which was much lower than the thermal energy (26 meV) [119].

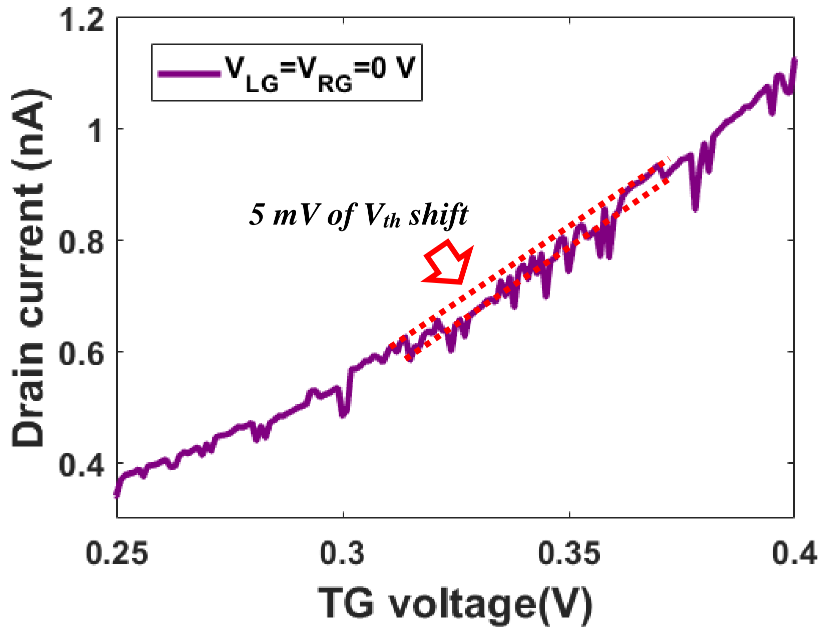
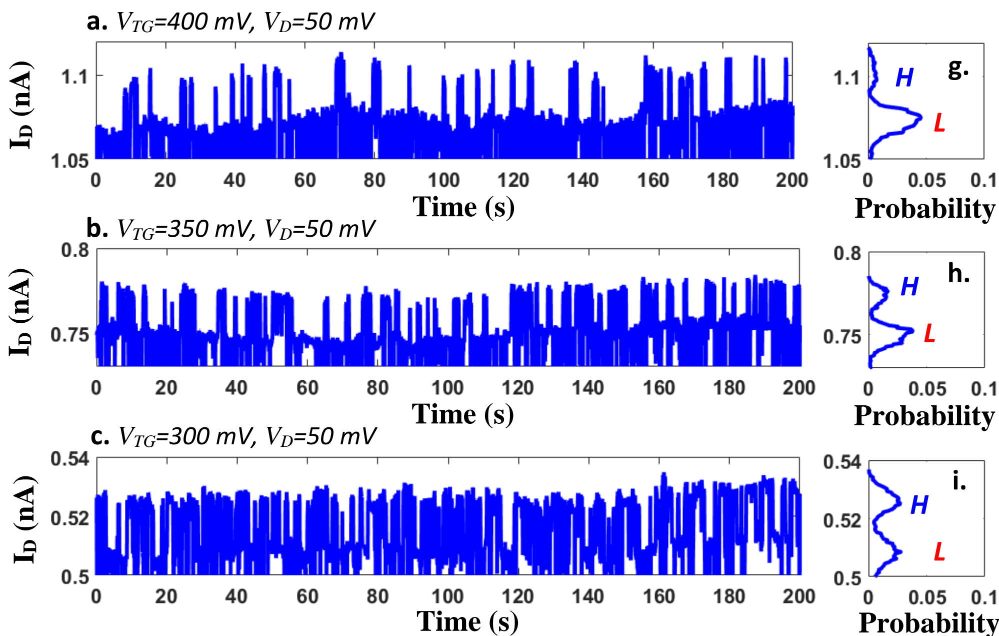


Figure 5.5 The marked area in Figure 5.4 where the RTSs could be observed [142].

After observing the RTSs in a certain area, the time-domain measurements of the I_D were implemented. The sampling time for the time domain measurement was set to 88 ms. The V_{LG}= V_{RG} was fixed to 0 V as a moderate barrier height to investigate the V_{TG} dependence with the same method in chapter 3. The measurement results of the probability distribution of the I_D was shown in Figure 5.6. By increasing the V_{TG} bias, the probability distribution of the I_D was observed to shift from mainly "High" state to mainly "Low" state. The occupation rate of the quantum dot was extracted in Figure 5.7(a) on the V_{TG} dependence using the same method in chapter 3.



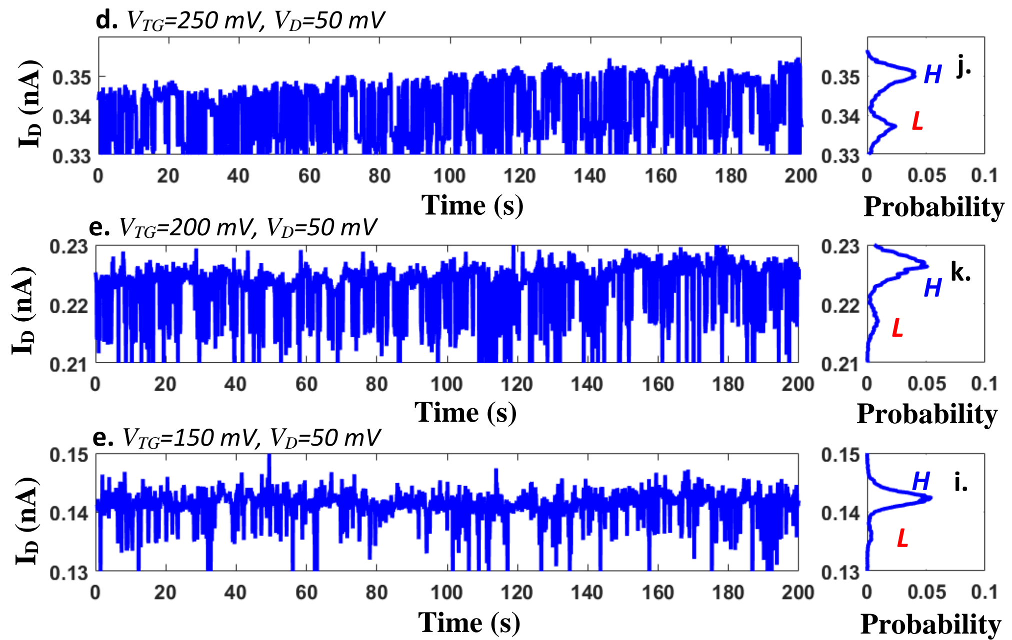


Figure 5.6 (a-f) The time-domain measurements on V_{TG} dependence with V_{LG}=V_{RG}=0 V and V_D=50 mV. (g-l) The corresponding probability distribution of the I_D [142].

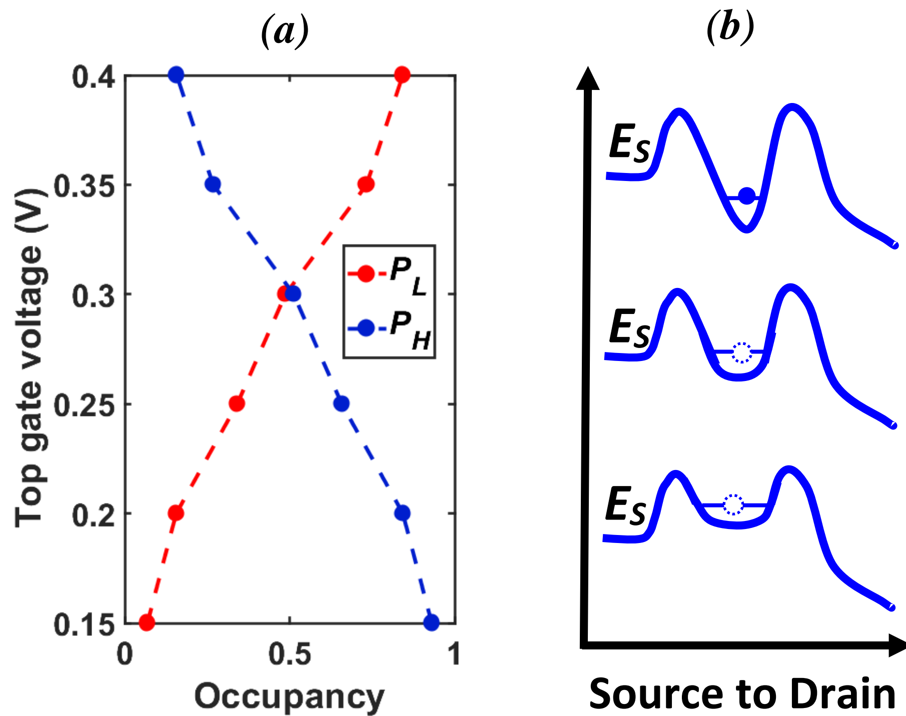


Figure 5.7 (a)The occupation of the I_D in "High" and "Low" states with different V_{TG}. (b)The schematic diagram of the energy level of the quantum dot on V_{TG} dependence[142].

At higher V_{TG} bias (V_{TG}=0.4 V), the energy level of the quantum dot controlled by the V_{TG} bias should be below the Fermi energy of the source. The electrons were more likely to occupy the quantum dot. In this condition, the channel was less transparency and the I_D was mainly in "Low" state. On the other hand, at lower V_{TG} bias (V_{TG}=0.15 V), the energy level of the quantum dot should be above the Fermi energy of the source. The quantum dot was more likely to be empty. In this condition, the channel was more transparency and the I_D was mainly in "High" state. The schematic diagram of the energy level of the quantum dot with changing V_{TG} bias was shown in Figure 5.7(b). It implied that the average time (τ_H) of the I_D in the "High" state was the emission time and the average time (τ_L) of I_D in the "Low" state was the capture time for the electrons. Then we extracted the τ_H and τ_L with the V_{TG} dependence, shown in Figure 5.8(a). And the ratio of the τ_H/τ_L was calculated in Figure 5.8(b) as well. It could be found that the V_{TG} bias also controlled the average lifetimes of the RTSs.

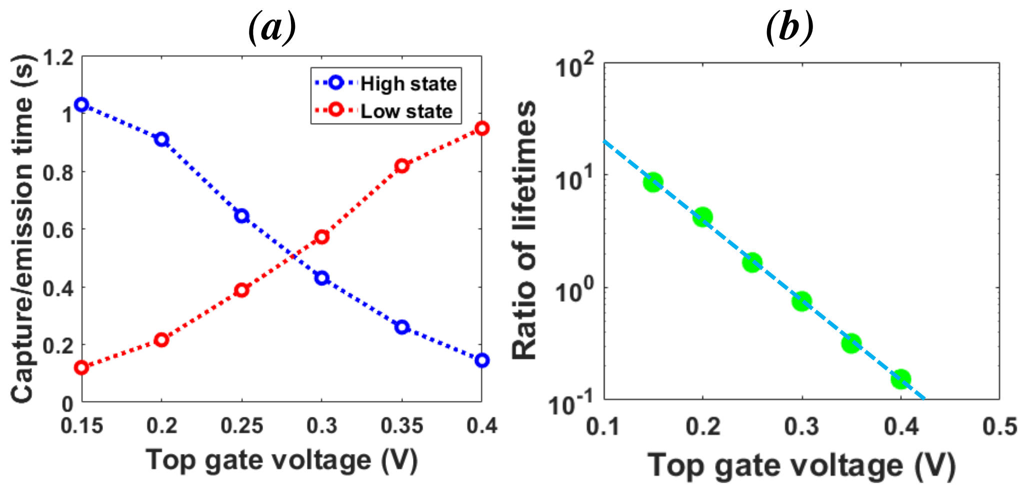


Figure 5.8 (a) The τ_H and τ_L with the V_{TG} dependence. (b) The ratio of the τ_H/τ_L with the V_{TG} dependence in the log scale by a linear fitting [142].

It was suggested that the energy level in the quantum dot could be estimated by the ratio of the lifetimes with respecting to the energy level of the source (E_s) [144]. In the Boltzmann distribution, the probability (P) for an electron surpassing the energy barrier (E) was,

$$P \propto exp\left(-\frac{E}{k_B T}\right) \tag{5.3}$$

The k_B was Boltzmann constant. The T was the absolute temperature. Therefore, the probability for the electron being trapped in the quantum dot was,

$$\frac{t}{\tau_L} \propto exp\left(-\frac{E - E_s - (\epsilon - E_s)}{k_B T}\right) \tag{5.4}$$

The probability for the electron being de-trapped in the quantum dot was,

$$\frac{t}{\tau_H} \propto exp\left(-\frac{E - E_s}{k_B T}\right) \tag{5.5}$$

The t was the real-time for the electron staying in a certain condition. $E - E_s$ was the energy barrier height. $\epsilon - E_s$ was the energy level of the quantum dot. Then we could derive the energy level of the quantum dot,

$$\epsilon - E_s = k_B T\left(\frac{\tau_H}{\tau_L}\right) \tag{5.6}$$

In this case, we could extract the energy level of the quantum dot from the ratio of the lifetimes with E_s=0 mV, shown in Figure 5.9(a). By referencing the occupancy distribution in Figure 5.7(a), it showed clearly that the quantum dot was mainly empty while the energy level of the quantum dot was above the E_s (Figure 5.9(b)). On the other side, the quantum dot was mainly occupied while the energy level of the quantum dot was below the E_s (Figure 5.9(d)). This confirmed that the V_{TG} bias manipulated the single electron trapped and de-trapped in the quantum dot by controlling the energy level of the quantum dot.

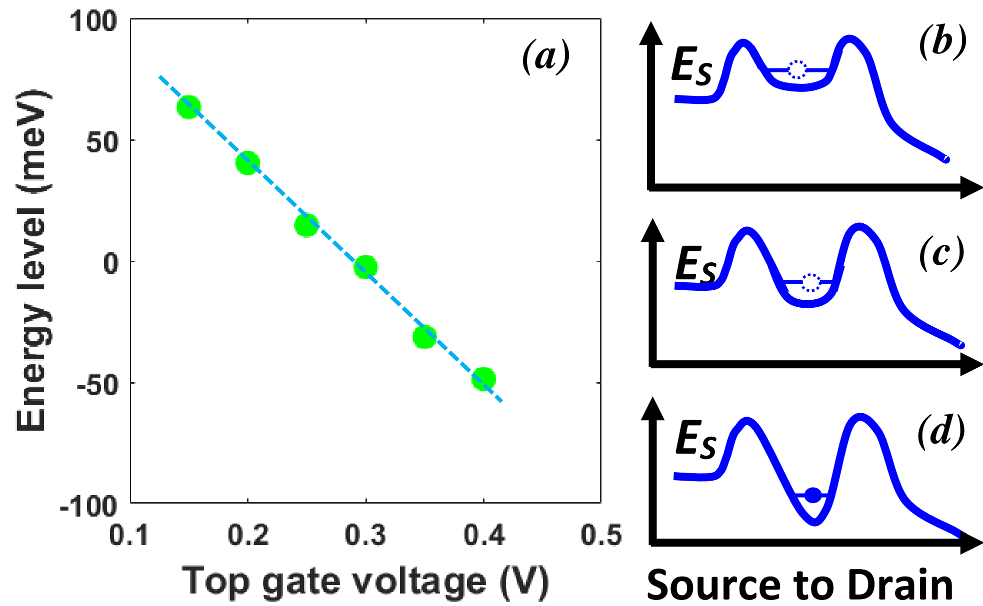


Figure 5.9 (a) The energy level of the quantum dot on the V_{TG}. The dash line was the linear fitting for the extracted points from the measurement results. (b-d) The schematic diagram of the energy level in the quantum with different V_{TG} bias [142].

Then in order to investigate the RTSs on the V_{LG}= V_{RG} dependence, the time-domain measurements were implemented with the condition that the V_{TG} was fixed at 50 mV and the V_{LG}= V_{RG} was increased from -1.6 V to 0.4 V respectively. The measurement results were shown in Figure 5.10.

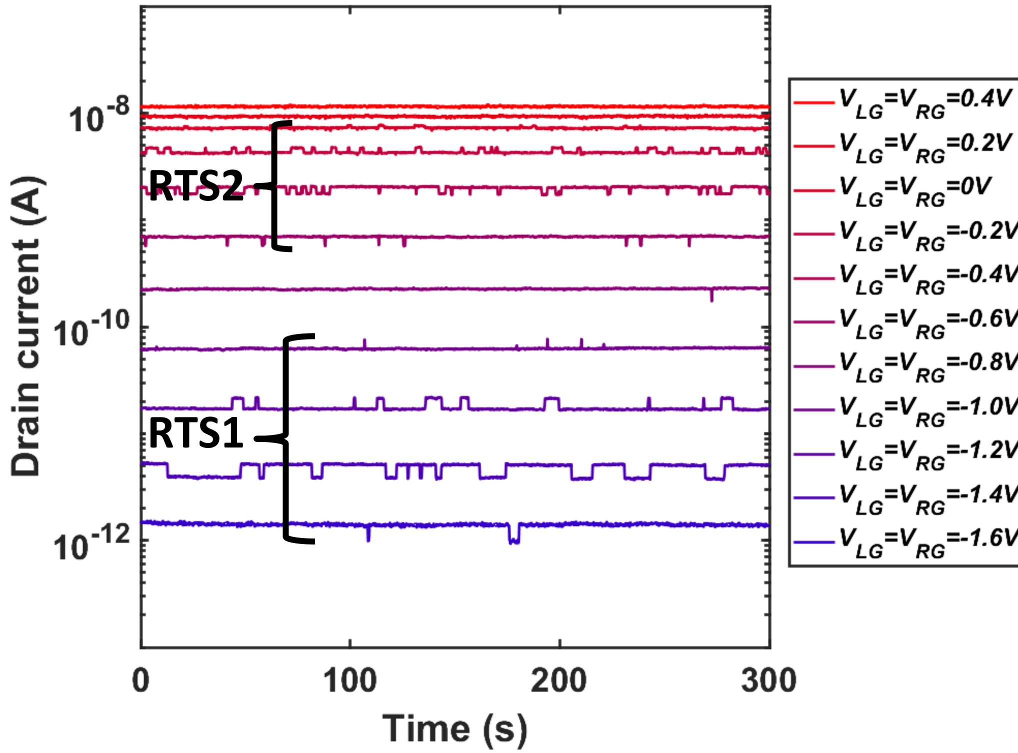


Figure 5.10 The time-domain measurement results on V_{LG}=V_{RG} dependence with V_{TG}=50 mV and V_D=50 mV [142].

From the results in Figure 5.10, one typical RTSs named RTS1 were observed at low V_{LG}= V_{RG} bias region with a long lifetime around 20 s. On the contrast, another typical RTSs named RTS2 were observed at higher V_{LG}= V_{RG} bias region with a short lifetime less than 10 s. The probability distribution of the I_D was extracted and typical ones were shown in Figure 5.11(a-d). Before V_{LG}= V_{RG} bias was increased to -0.8 V, the RTS1 was dominated the fluctuation of the I_D. Then the RTS1 was disappeared. By increasing the V_{LG}= V_{RG} bias further, the RTS2 started to dominate the fluctuation of the I_D. The reason why we could observe the V_{LG}= V_{RG} dependence of the RTSs was considered to be the coupling between the two First gates and the quantum dot. The schematic diagram of the equivalent circuit was shown in Figure 5.12.

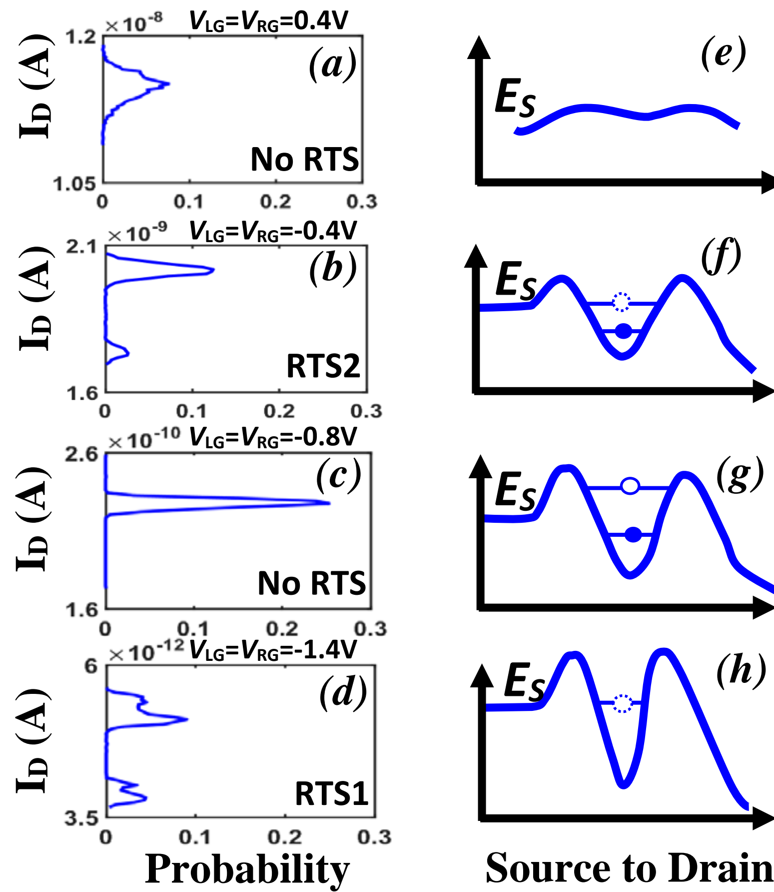


Figure 5.11 (a-d) The typical probability distribution of the I_D in the selected V_{LG}=V_{RG} bias condition. (e-f) The schematic diagram of the different energy levels in the quantum dot corresponding to (a-d) [142].

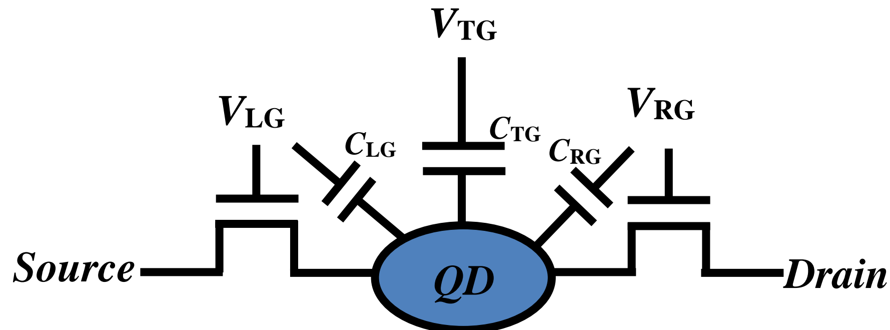


Figure 5.12 (a-d) The schematic diagram of the equivalent circuit for the multi-gate silicon nanowire transistor [142].

The two First gates were modeled as transistors with coupling capacitances (C_{LG} and C_{RG}) to the quantum dot. The Top gate was modeled as a capacitance (C_{TG}) coupled to the quantum dot. As we had confirmed the role of the V_{TG} bias was to control the energy level in the quantum dot in the previous discussion, the V_{LG}= V_{RG} bias was presumed to change the barrier height. Due to the potential barrier height for the quantum dot was finite, C_{LG} and C_{RG} could be changed by increasing the V_{LG}= V_{RG} bias. Therefore, the electron tunneling rate could be changed by the V_{LG}= V_{RG} bias. In this case, the occupancy of the quantum dot could be manipulated by the V_{LG}= V_{RG} bias. We could evaluate the capacitance coupling strength from the bias window that the RTSs were observed from mainly in "High" state to mainly in "low" state. From Figure 5.7(a), the bias window for V_{TG} dependence was about 0.25 V. From Figure 5.11, the bias window for the V_{LG}= V_{RG} dependence was about 0.8 V for RTS2. It implied that the coupling between the *TG* and the quantum dot was much stronger.

We could also use the same method to extract the energy level in the quantum dot in the different V_{LG}= V_{RG} conditions by using the ratio of the lifetime as the discussion for the V_{TG} dependence, shown in Figure 5.13(a).

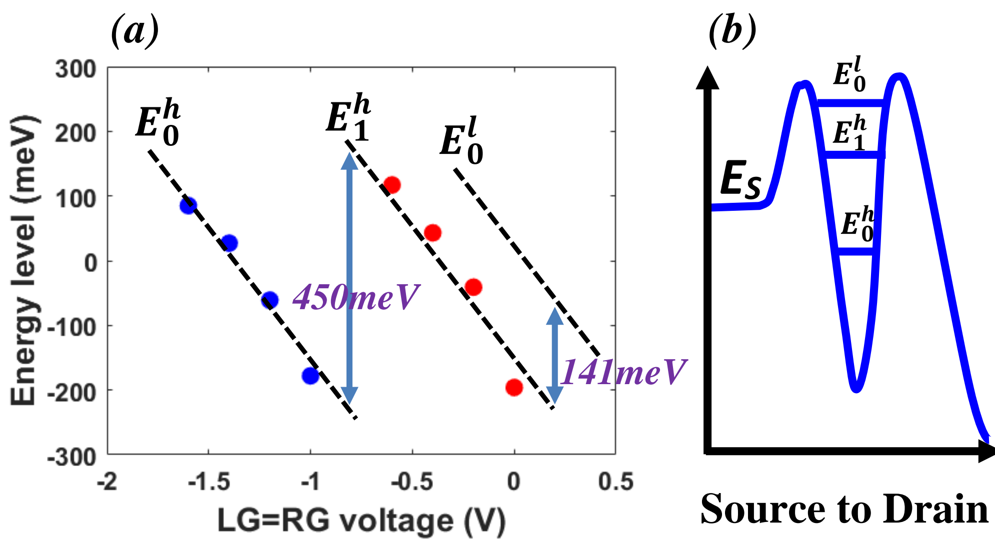


Figure 5.13 (a) The energy level of the quantum dot with different First gates voltage. The filled circles and squares were extracted from the measurement data for RTS1 and RTS2 respectively. The lines were the linear fitting for the different energy level (E_0^h, E_1^h and E_0^l) by the effective mass approximation [142].

The RTS1 corresponded to the lowest energy level in the artificial quantum dot with the low V_{LG}= V_{RG} bias (5.11(h)). By increasing the V_{LG}= V_{RG} bias, this energy level would be mainly occupied by the electron, so that the RTS1 could not be observed (Figure 5.11(g)). If the V_{LG}= V_{RG} bias were increased further, the higher energy level in the quantum dot was activated and RTS2 could be observed (Figure 5.11(f)). Until the potential barriers were almost plat, we could not see any RTSs generated from the artificial quantum dot (Figure 5.11(e)). It could be clearly seen the fitting lines for the

energy changing in the quantum dot for both RTS1 and RTS2 in Figure 5.13(a). The fitting lines also confirmed the coupling between the two First gates and the quantum dot.

In addition, the parallel fitting lines indicated the two quantum states in different energy levels in the same artificial quantum dot. The energy separation between the two quantum states could be roughly extracted from Figure 5.13(a). The energy splitting was induced by the quantum confinement in the direction perpendicular to the substrate surface [122, 124]. We could estimate the quantum confinement energy levels in the inversion layer on <100> Silicon surface by,

$$E_n^h = \frac{(n+1)^2\hbar^2}{2m_h}\left(\frac{\pi}{t_{Si}}\right)^2 = E_0^h(n+1)^2 \tag{5.7}$$

$$E_n^l = \frac{(n+1)^2\hbar^2}{2m_l}\left(\frac{\pi}{t_{Si}}\right)^2 = E_0^l(n+1)^2 \tag{5.8}$$

The $\hbar$ was the Planck constant divided by 2π. The $m_h = 0.980m_0$ was the effective mass for the heavy electrons in the conduction band valley. The $m_l = 0.198m_0$ was the effective mass for the light electrons in the conduction band valley. The m_0 was the mass of an electron in the vacuum. The t_{Si} was the effective thickness of the inversion layer. The $n \geq 0$ was an integer to indicate the number of nodes in the envelope wave function in the direction of depth. It could be found that the lowest energy level for RTS1 was corresponding to the E_0^h and the energy level for RTS2 could be corresponding to the E_1^h. The schematic diagram of the different energy levels in the artificial quantum dot was shown in Figure 5.12(b). Based on the estimation of the energy separation ($E_1^h - E_0^h \approx 446\,meV$) from Figure 5.12(a), we could extract the $t_{Si} = 1.6\,nm$ from the experimental data by assuming the linear band bending. This result was in agreement with the previous study in quantum confinement [122, 124]. Then the E_0^l could be estimated as 141 meV higher than E_1^h. It implied that the heavy electron mass dominated the conduction band valley in the inversion layer on <100> silicon surface [122, 124].

5.4 Summary

In summary, we could use the multi-gate silicon nanowire transistor to generate the RTSs. It was confirmed the RTSs could be manipulated by not only the Top gate bias but also the two First gates. The Top gate controlling was to change the energy level of the quantum dot. The two First gates controlling was to change the coupling between the First gates and the quantum dot and select a certain activated quantum state by adjusting the potential barrier. It confirmed the theoretical foundation to use the RTSs as a QRNG source with the multi-gate silicon nanowire transistors.

Chapter 6 Application of the RTSs for QRNG source

6.1 Introduction

Based on the manipulation of the RTSs in the silicon nanowire in the last chapter, we investigated the chance to use the RTSs as a QRNG source from the view of fundamental physics. In order to show that the RTSs manipulation was common in our multi-gate silicon nanowire transistors, we used a different device from the last chapter to refine the RTSs controlling.

6.2 Device structure

The device structure was shown in Figure 6.1(b). The length of the two First gate (*LG* and *RG*) was 100 nm. The Top gate (*TG*) was 150 nm. After TMAH etching process, the trapezoid cross-section feature of the nanowire could be found as 82 nm for the upper base and 117 nm for the lower base in the SEM image (Figure 6.2). After oxidation, the upper base of the trapezoid nanowire was 71 nm based on our TCAD simulation. The tripe-gates could control the potential barrier height and the energy level to form the quantum dot, shown in Figure 6.1(b).

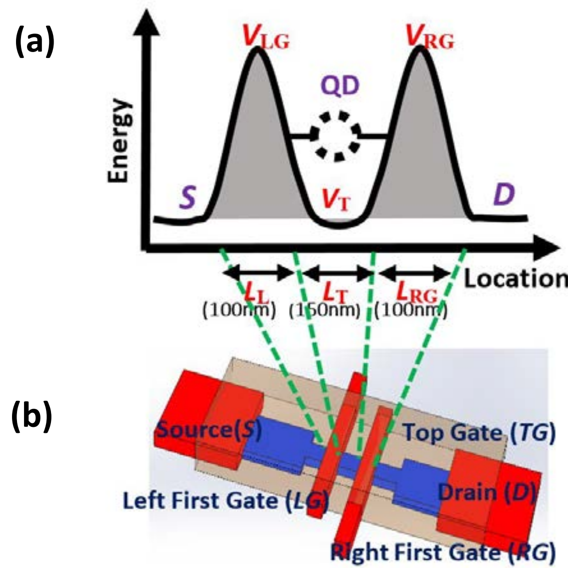


Figure 6.1 The structure of the triple gates silicon nanowire transistor with the atomically flat interface (a) The artificial quantum dot created by TG, LG and RG bias. (b) The schematic diagram of the device.

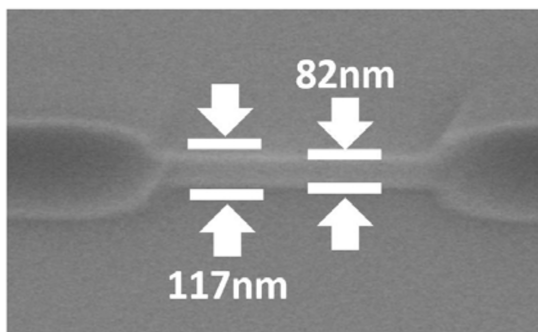


Figure 6.2 Dimension confirmation for the nanowire after TMAH in SEM image.

6.3 Measurement and results

The devices were measured at room temperature by using Agilent B1500 and Cascade M150 probe station at room temperature. The background noise was confirmed below 100 fA. With raised source and drain and self-aligned TG methods, we made a low-doped junction-less trapezoid Si nanowire transistor. The ON current was of the order of nA. It was consistent with the device in the last chapter. The standard I_D versus gate voltages characteristics were respectively carried out on TG, LG, and RG to confirm the successful transistor operations, shown in Figure 6.3-6.5. The *SS* for *TG* bias with V_{LG}= V_{RG}=0 V was 77 mV/decade at room temperature. The *SS* for *LG* bias with V_{TG} =0V and V_{RG} floating was 360 mV/decade. The *SS* for *RG* bias with V_{TG} =0V and V_{LG} floating and 260 mV/decade. The asymmetry would be coming from the application of the V_D, while the source was always grounded. As RG was much closer to the Drain, the current was more sensitive to the RG. TG not only controlled the energy level of the artificial quantum dot between LG and RG, but also affected the carrier concentration in the dopant diffusion region between doping diffusion windows to the First gate.

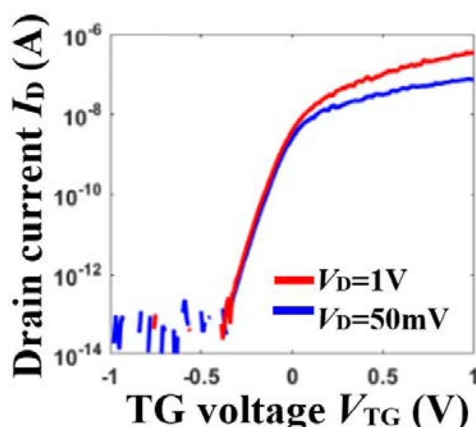


Figure 6.3 I_D-V_{TG} characteristics with V_{LG}= V_{RG}=0V at V_D=50 mV and 1 V

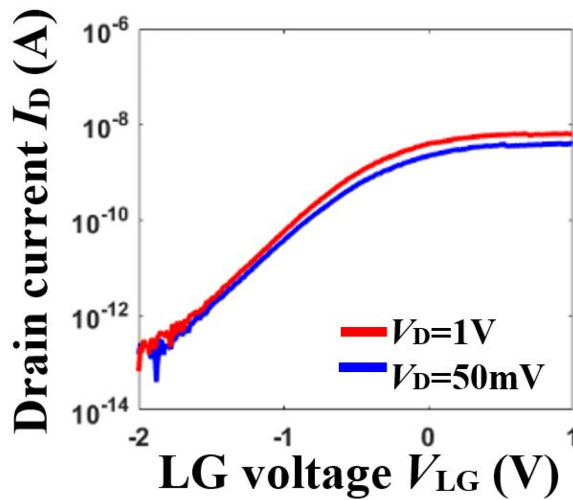


Figure 6.4 I_D-V_{LG} characteristics with V_{TG}=0 V and V_{RG} floating at V_D=50 mV and 1 V

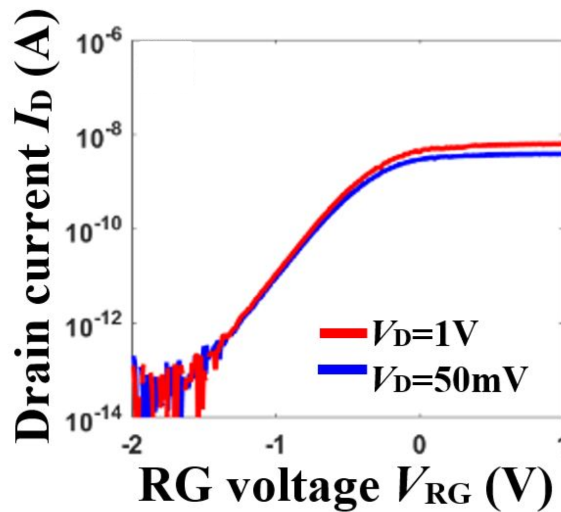


Figure 6.5 I_D-V_{RG} characteristics with V_{TG}=0 V and V_{LG} floating at V_D=50 mV and 1 V

By measuring the TG dependence of I_D under the different potential barriers formed by LG and RG, the typical RTSs were observed especially in the linear region of I_D versus V_{TG} characteristics, as shown in Figure 6.6(a). It was worth noting that the RTSs were observed mainly in 2 levels In the low V_{TG} (Figure 6.6(b)). From the medium to high V_{TG} region, the RTSs were observed in 3 levels (Figure 6.6(c)). In the high V_{TG} region, other RTSs with a large amplitude were observed. At the same time, the previous RTSs with small amplitude still could be observed both in high and low state of the large amplitude RTSs (Figure 6.6(d)). This implies that we had observed multiple quantum states in our artificial quantum dot during changing TG, LG, and RG voltages. It was supposed to be caused by the different bandgap increase in the different part of the channel due to the quantum confinement as we discussed in chapter 4.

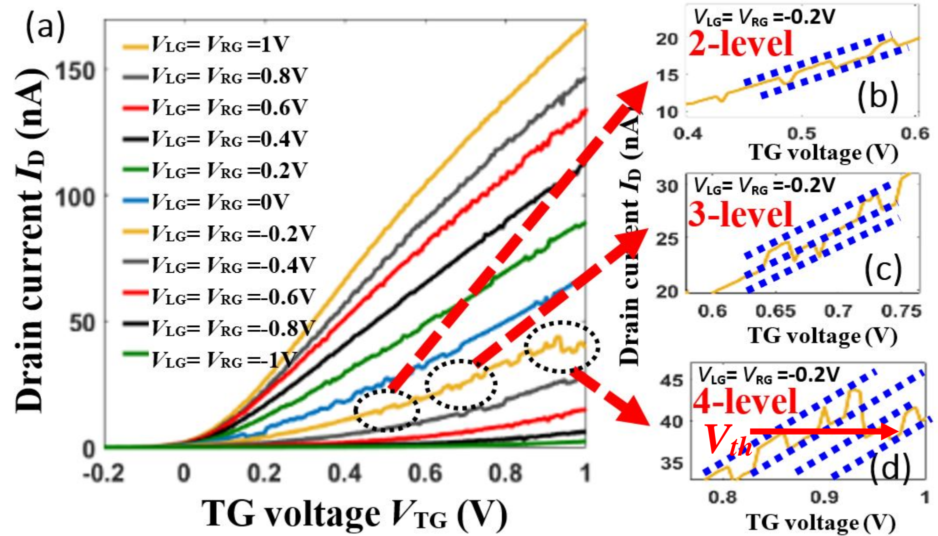


Figure 6.6 The Multi-level RTSs observed in low, medium, and high V_{TG} region and clearly shift the Vth in the linear region.

In order to utilize the RTSs from the artificial quantum dot as QRNG source, the probability ratio of "high" and "low" states should be the same to guarantee the equal probabilities to find each state [12]. As we discussed in the last chapter, the electron probability distribution (electron wave function) could be manipulated by changing V_{TG}, V_{RG} and V_{LG}. The V_{RG} and V_{LG} could control the energy barrier height for the artificial quantum dot and the V_{TG} could control the carrier concentration in the Si nanowire [143]. So that we did similar time dependence measurements on I_D with a constant V_D=50 mV and V_{RG}=V_{LG}=-50 mV. The V_{TG} was changed from 100 mV to 300 mV systematically. The measurement results were shown in Figure 6.7. The sampling time for the time domain measurement was 108 ms.

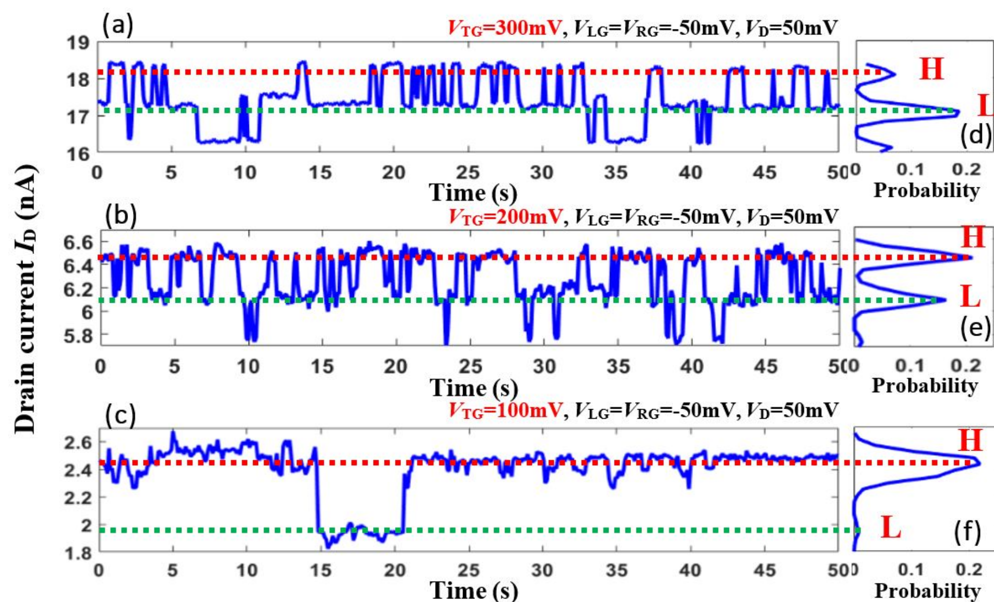


Figure 6.7 The time domain measurements with different V_{TG}.

During increasing V_{TG} from 100 mV to 300 mV, it could be found that the probability of I_D in "high" states decreased from 92% to 18% and the probability of I_D in "low" states increased from 8% to 82% with 5000 sampling points in Figure 6.7(d-f). As the V_{TG} controlled the energy level of the quantum state in the quantum dot, the energy level of the quantum state should be higher than the Fermi level in the source at V_{TG}=100 mV. There was no quantum state in the quantum dot. So that the electron could not be trapped in the quantum state. In this case, the I_D was mainly in "high" state. By increasing the V_{TG}, the energy level of the quantum became closer to the Fermi level in the source. So that the resonance tunneling start. The I_D was switched between "high" state and "low" state. While the energy level of the quantum state was lower than the Fermi level in the source at V_{TG}=300 mV. The electron was more likely to occupy the quantum state. In this case, the I_D was mainly in "low" state. The measurement results were consistent with that in the last chapter.

In order to investigate the correlation behavior of the I_D, we extracted the lag plot of the I_D with the time lag (Δt) at 108 ms and 100s in Figure 6.8.

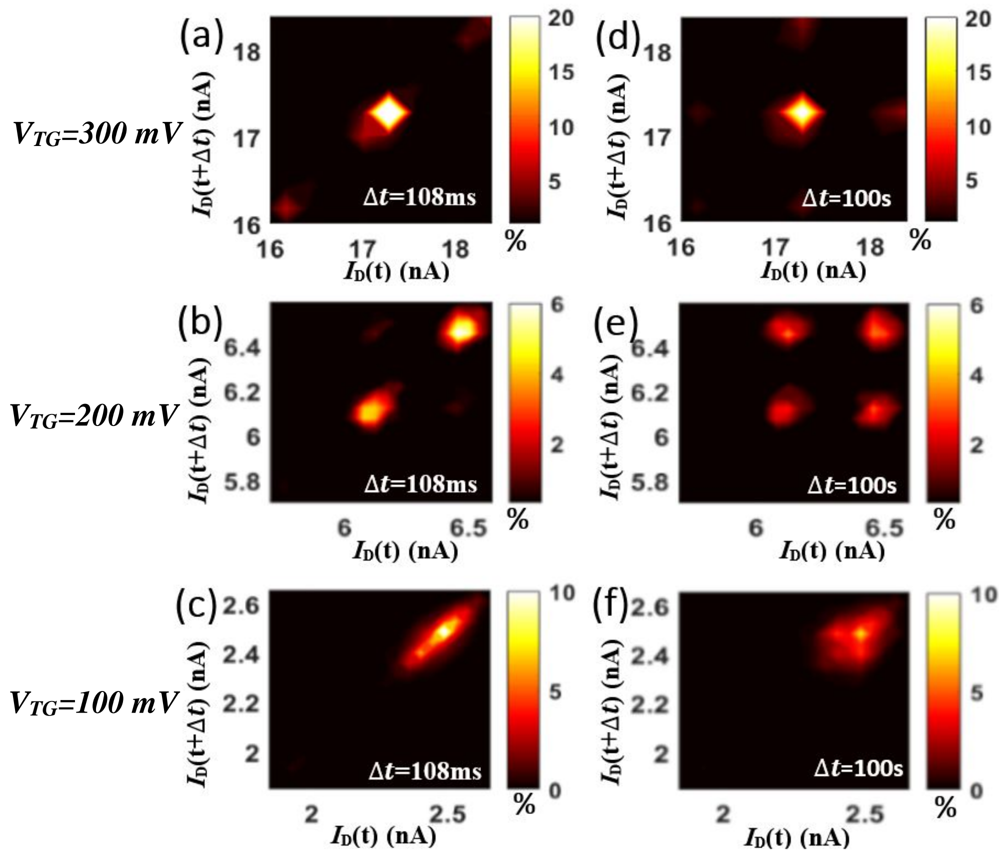


Figure 6.8 The lag plots with time lag (Δt) at 108 ms and 100 s to show the local correlation of the I_D with different V_{TG} bias.

In Figure 6.8(a-c), it could be observed that the correlations between two sampling points next to each other were in a linear correlation to some extent. However, after a long time lag (100s), the I_D was no linear correlation, shown in Figure 6.8(d-f). In this case, in order to use the RTSs to be a QRNG source, the sampling time should be determined by the relaxation time of the electron from the artificial quantum dot. No matter how It was clearly shown that we could find a certain V_{TG} between 100 mV and 300 mV to make the probability ratio of the “high” and “low” state equal (50 %: 50%).

In addition, the two First gate controlled the potential barrier height of the artificial quantum dot. The probability ratio of the I_D in “high” and “low” state was also the function of the first gate bias [p3]. We also did the time dependence measurement on I_D with a constant V_D=50 mV and V_{TG}=200 mV. The V_{RG}=V_{LG} was changed from -100 mV to 0 mV systematically. The step for V_{RG}=V_{LG} was much smaller than that in the last chapter because we wanted to investigate the more details about First gate dependence for one certain kind of RTSs. The measurement results were shown in Figure 6.9. The sampling time for the time domain measurement was also 108 ms.

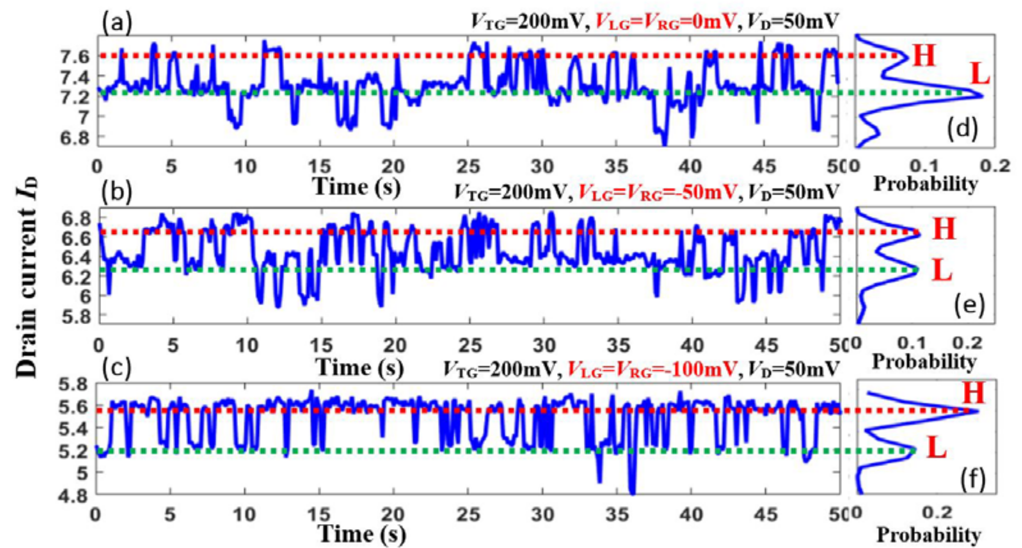


Figure 6.9 The time domain measurements with different V_{RG}=V_{LG}.

During decreasing the V_{RG}=V_{LG} from 0 V to -100 mV, the probability of the I_D in “high” states was increased and that in “low” state was decreased, shown in Figure 6.9(d-f). In the higher V_{RG}=V_{LG} bias, the potential barrier for the artificial quantum dot was low, the electrons were more likely to occupy the quantum state. While the V_{RG}=V_{LG} bias was decreased, the potential barrier was higher even than thermal energy. The electron was quite difficult to tunneling into the quantum dot. Then the quantum state was more likely to be unoccupied, which formed by a certain TG voltage in the quantum dot. In this case, the variation tendency of the probability distribution on I_D could be

explained and the schematic diagram of the energy level in the quantum dot was shown in Figure 6.10.

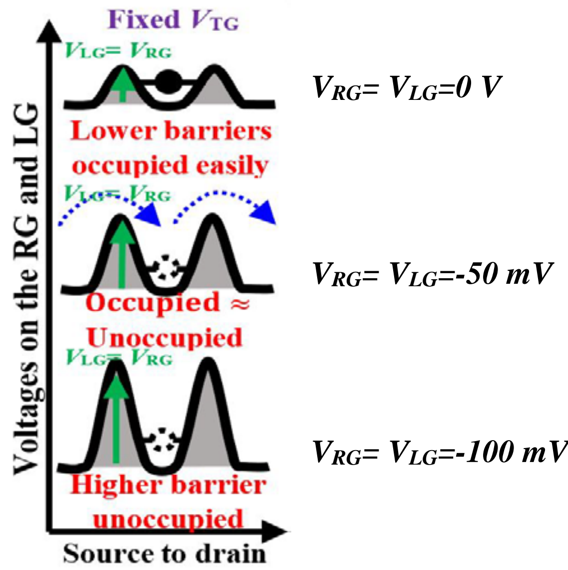


Figure 6.10 The schematic diagram of the energy level in the quantum dot.

It was also clearly shown that we could find a certain V_{RG}=V_{LG} bias between -100 mV and 0 mV to achieve the probability ratio of the “high” to “low” state equal (50 %: 50%). The lag plots of the I_D with the time lag (Δt) at 108 ms and 100s were also extracted from the raw data and shown in Figure 6.11.

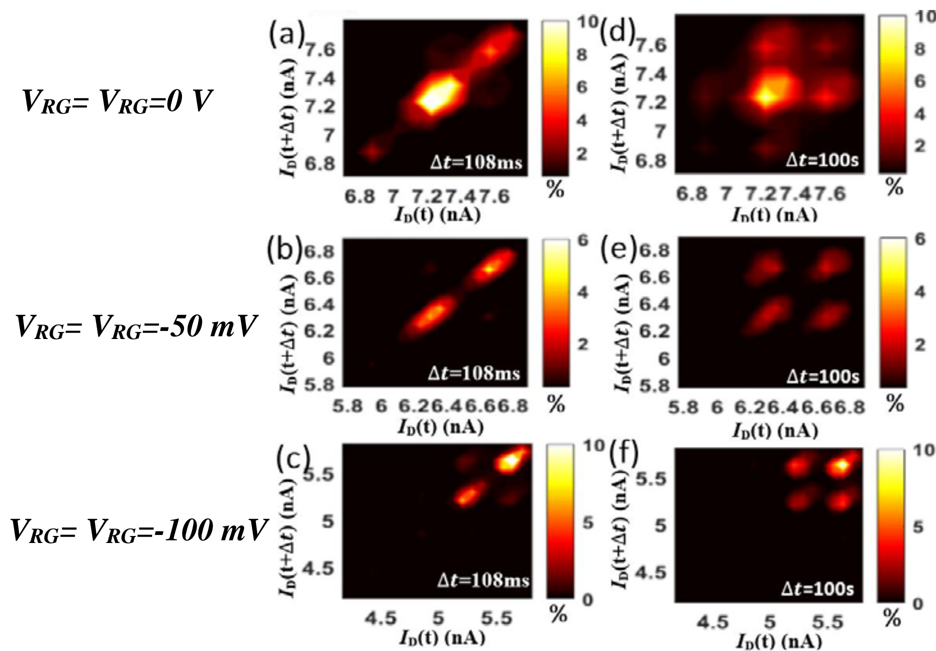


Figure 6.11 The lag plots with time lag (Δt) at 108 ms and 100 s to show the local correlation of the I_D with different V_{RG}=V_{LG} bias.

It could be found that a certain long sampling time was still needed to eliminate the local correlations. In addition, by comparing the measurement results in Figure 6.7(b, e) with that in Figure 6.9(b, e), it could be found that the probability distribution of the I_D was stable in the same voltage condition in the two times measurements.

By comparing the Top gate controlling and the First gates controlling results, both of them could adjust the probability distribution of the I_D. However, the Top gate could affect the I_D in a large range from 2 nA to 20 nA in Figure 6.7. The transport properties in the channel might be completely changed to make the model more complex. On the contrary, the First gates could be inversed the probability distribution in a limited I_D range from 5 nA to 8 nA in Figure 6.9. That was why only use Top gate was difficult to manipulate the RTSs from the quantum dots precisely. It was consistent with the experience in chapter 3. Therefore, the Top gate was selected to form not only the channel at a certain current level but also the certain energy level in the quantum dot. Then the First gates were used to adjust the potential barrier height to manipulate the RTSs. In order to achieve better performance to use the RTSs as a QRNG source, the higher I_D and larger amplitude ΔI_D and shorter capture/emission times were required of RTSs. In this case, the global sweeping on different V_{RG}=V_{LG} bias were proceeded from V_{TG}=0 V to V_{TG}=1 V to find a better combination of the Top gate bias and the First gate bias. The V_D bias was 50 mV all the time. We extracted the probability distribution of the I_D from the measurement data. The typical low V_{TG} region was shown in Figure 6.12 and the typical high V_{TG} region was shown in Figure 6.13. It was found that if the V_{TG} were very low (100 mV), the amplitude of the RTSs was quite small. Although we could still observe the highest part of probability distribution was moving to the higher current level with V_{RG}=V_{LG} bias increasing in Figure 6.12(a), it was quite difficult to clearly separate the “high” and “low” states area. With increasing the V_{TG} to 300 mV (Figure 6.12(c)), the “high” and “low” states were well separated by a larger amplitude RTSs. However, a much lower current state was observed and get a larger proportion during increasing V_{RG}=V_{LG} bias. According to the multi-channel transport model in chapter 4, there should be multiple energy levels of quantum states in our artificial quantum dot. With a higher V_{TG} bias, the high energy level quantum state in the side surface channel could be activated. By lowing the potential barrier of the quantum state, the electrons started to occupy in higher energy level quantum state. At the same time, the electrons had already occupied the previous quantum state that was at lower energy level. In that case, the I_D could be observed in a much lower state named “low2” state (l’).

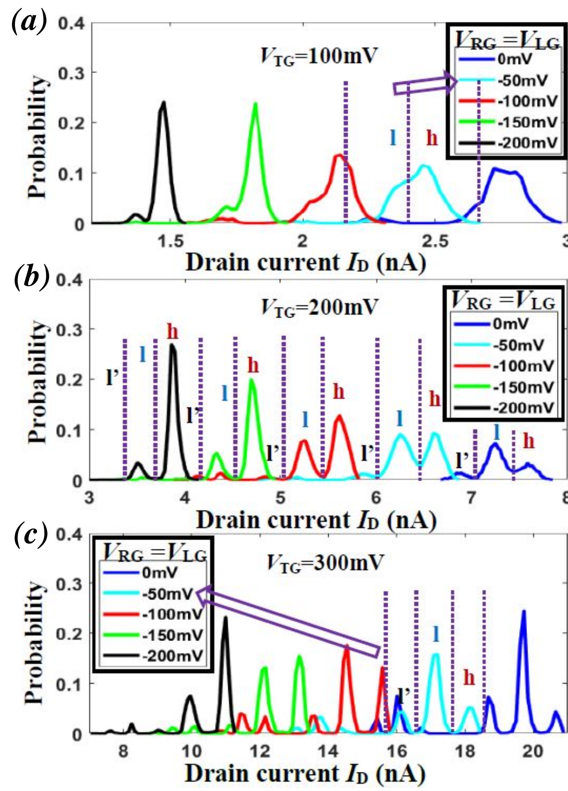


Figure 6.12 The variation tendency of probability distribution of I_D in "high (h)" and "low (l)" states with V_{LG}= V_{RG} from 50 mV to -200 mV in the low V_{TG} region.

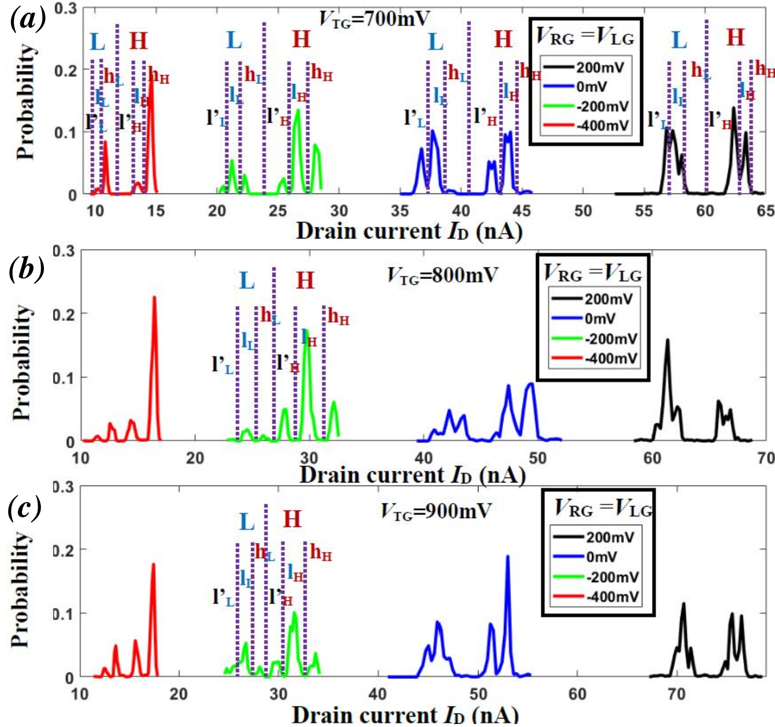


Figure 6.13 The variation tendency of the probability distribution of I_D in "High (H)" and "Low (L)" states with V_{LG}= V_{RG} from 200 mV to -400 mV in the high V_{TG} region.

We extracted the occupation rate of the "high" state (h), "low" state (l) and "low2" state (l') both in a certain V_{TG} bias to investigate the V_{RG}=V_{LG} dependence, and in a certain V_{RG}=V_{LG} bias to investigate the V_{TG} bias dependence respectively in the low V_{TG} region, shown in Figure 6.14.

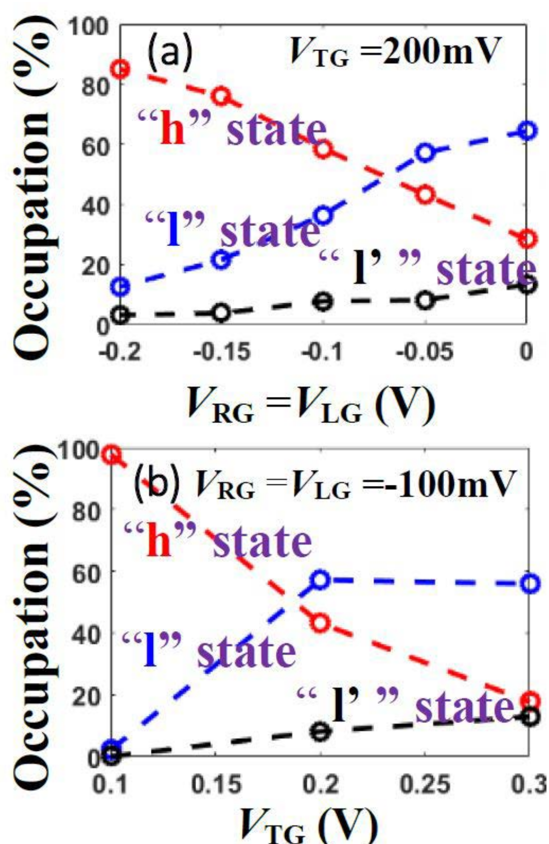


Figure 6.14 (a) The occupation rate of the three states on V_{RG}=V_{LG} dependence with V_{TG}=200 mV. (b)The occupation rate of the three states on V_{TG} dependence with V_{RG}= V_{LG}=-100 mV.

It could be found that the "low2" state would affect the occupation of the quantum states both in the V_{RG}=V_{LG} dependence and the V_{TG} dependence. Although the occupation rate was raised for the "low2" state in the V_{RG}=V_{LG} dependence, the occupation rates for "high" and "low" states could still keep in a typical inverse relationship at least (Figure 6.14(a)). However, the occupation rates for both "high" and "low" states were dropped while the "low2" state significantly raised in the V_{TG} dependence (Figure 6.14(b)). That was due to the V_{TG} bias controlled the energy level of the quantum states. By increasing V_{TG} bias, the lower energy quantum state was mainly occupied by the electrons and the effect for the current fluctuation was degraded. Meanwhile, the higher energy level quantum state was active to dominate the channel current fluctuation. Actually, it was two RTSs superimposed together. In this case, the First gate was more useful to generate purely one RTSs for a high-performance QRNG source application.

It was remarkable that with increasing the V_{TG} to the relatively high voltage region, another kind of RTSs was observed with the large amplitude that was about 5 nA in Figure 6.13. It was named RTS2 with "HIGH" and "LOW" states marked in Figure 6.13. The previous RTS including "high", "low" and "low2" states were named RTS1, which had a smaller amplitude about 500 pA. The typical time-domain measurement result was shown in Figure 6.15.

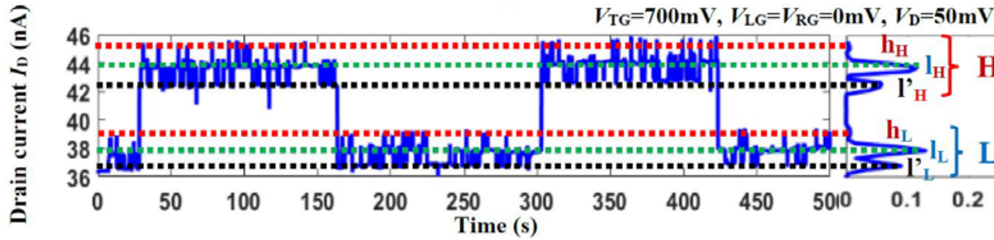


Figure 6.15 The time-domain measurement at V_{TG}=700 mV, V_{RG}= V_{LG}=0 V and V_D=50 mV. The RTS2 was marked as "H" and "L". The RTS1 was marked as "h", "l" and "l'".

It could be found that the capture and emission time was quite long for the RTS2. It was supposed that the RTS2 was formed by the oxide trap in the gate oxide layer as we had discussed in chapter 3 [5]. The RTS1 was observed in both "HIGH" and "LOW" states of the RTS2. That was consistent with the results in chapter 3. We extracted the occupation rate of the "HIGH" state (H), "LOW" state (L) both in a certain V_{TG} bias to investigate the V_{RG}=V_{LG} dependence, and in a certain V_{RG}=V_{LG} bias to investigate the V_{TG} bias dependence respectively in the high V_{TG} region, shown in Figure 6.16.

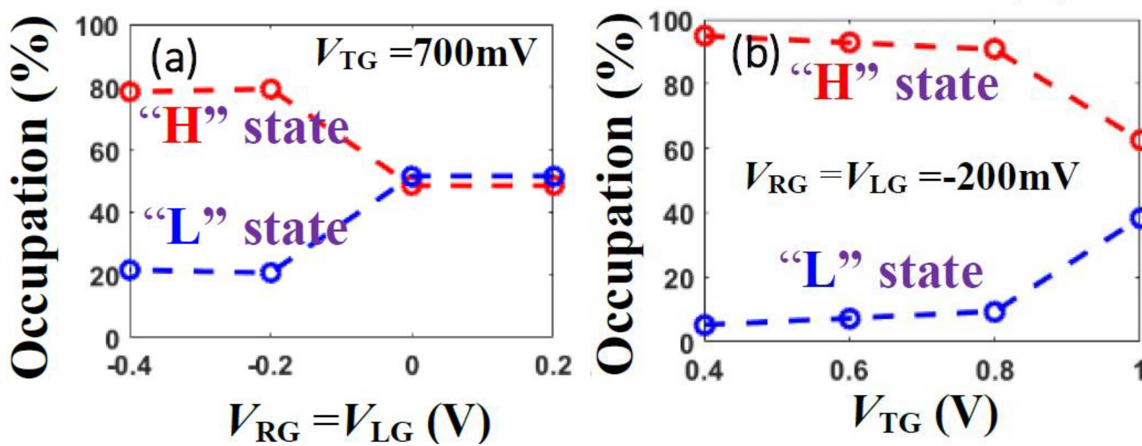


Figure 6.16 The occupation rate of the "HIGH" (H) and "LOW" (L) state in RTS2 (a) on V_{RG}=V_{LG} dependence with V_{TG}=700 mV and (b) on V_{TG} dependence with V_{RG}= V_{LG}=-200 mV.

The occupation of the oxide trap was shown the dependence neither on V_{RG}=V_{LG} bias nor on V_{TG} bias in Figure 6.16. That was mainly caused by the coupling between the Top gate and the First gates due to they were just separated by 9 nm thermal oxide layer. With a higher V_{TG} bias, the V_{RG}=V_{LG} bias could not raise the potential barriers efficiently. With a higher V_{RG}=V_{LG} bias, the V_{TG} bias could not adjust the energy level of the quantum states efficiently. In spite of this condition, the V_{TG} dependence could be still observed in Figure 6.16(b). With increasing the V_{TG} bias, the electrons were more likely tunneling into the oxide trap and making the charge trap neutral. So that the occupation rate in the "LOW" current state was raised according to the discussion in chapter 3. Due to the strong coupling between the Top gate and the First gates, actually, the changing of the occupation rate in Figure 6.16(a) also caused by the V_{TG} bias. With increasing the V_{RG}=V_{LG} bias, the actual V_{TG} bias applied on the oxide trap was increased. So that the tendency of V_{RG}=V_{LG} bias dependence will be the same as V_{TG} bias dependence. Once the V_{RG}=V_{LG}=0 V, the from the First gates bias to Top gate bias could be negligible. By a fixed the V_{TG} bias, the V_{RG}=V_{LG} bias lost control to the occupation rate, which could be confirmed from Figure 6.16(a).

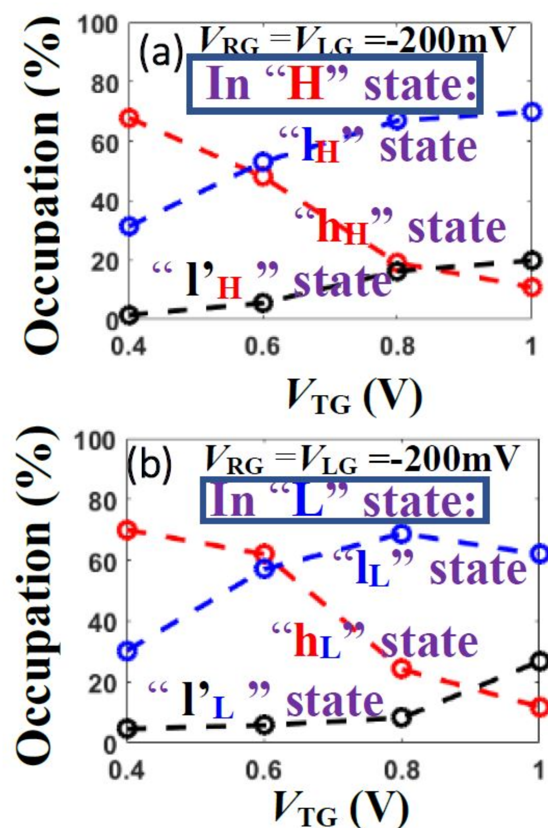


Figure 6.17 The occupation rate of the "high" (h), "low" (l) and "low2" (l') states of RTS1 (a) in the "HIGH" (H) state of RTS2 and (b) in the "LOW" (L) state of RTS2 on V_{TG} dependence with V_{RG}=V_{LG}=-200 mV.

In Figure 6.17, we extracted the occupation rate distribution of RTS1 in both the “HIGH” state and the “LOW” state of the RTS2 respectively. It implied that the RTS1 and RTS2 were almost independent. The RTS1 was still affected by the “low2” state.

As we discussed in chapter 3, the RTSs generated the oxide trap was not an ideal candidate to be used as a QRNG source, duo to its random formation process and long capture and emission time. From the global sweeping results, it could be found that the advantage of our multi-gate transistors was to manipulate the V_{TG} and V_{RG}=V_{LG} bias to select the activated quantum states at room temperature. We could decrease the V_{TG} bias to avoid any high energy level oxide traps to be activated. In this case, we could eliminate the RTS2. Then the V_{RG}=V_{LG} bias could control the potential barrier height and manipulate the occupancy of the quantum dot. In principle, we could achieve the situation that only the main RTS remained in the time-domain measurement and the occupancy to un-occupancy was 50% to 50%. Practically, the “low2” state could not be eliminated due to the physical structure of our trapezoid silicon nanowire. From the discussion of chapter 4, the bandgap changing was continuous from the top surface of the nanowire to the side surface of the nanowire. Therefore, the energy levels of the activated quantum states in artificial quantum dot were considered to be continuously changing. In this case, we could not eliminate the “low2” state, due to the energy level of the corresponding quantum state was quite close to the main quantum state. As far as the proportion of the “low2” was not that large compared to the “high” and “low” state especially in the V_{RG}=V_{LG} bias controlling, the “low2” state part could be assigned to the “low” state part in some extent. According to the global sweeping, we chose the V_{TG}=300 mV and did the refined time-domain measurements around V_{RG}=V_{LG}=-100 mV with the 1 mV step. We reached the point where the probability ratio of the “high” and “low” states was 50.85% to 49.15% on the bias condition of V_{TG}=300 mV, V_{RG}=V_{LG}=-106 mV and V_D=50 mV (Figure 6.18). 150,000 points were measured to prepare a symbolic randomness test. The sampling time was 24 ms.

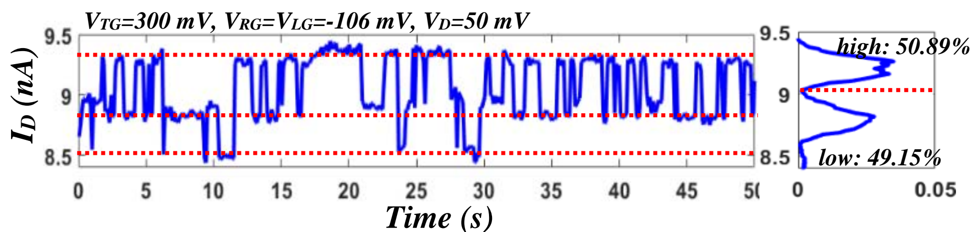


Figure 6.18 The time-domain measurement on V_{TG}=300 mV, V_{RG}=V_{LG}=-106 mV and V_D=50 mV.

6.4 Randomness test

As the research progress was still in the fundamental physics level to investigate the possibility of the RTSs generated by the multi-gate silicon nanowire to be a promising candidate as a QRNG source, the common RNG test from the engineering point of view was not implemented. Because from a QRNG source to a real QRNG product, it included many post-processing steps with the engineering technologies. In this part, we did a symbolic randomness test by comparing our random numbers extracted from the raw data with the random numbers generated by MATLAB.

The capture time was calculated as 0.5s as well as the emission time for the main quantum state. Therefore, the raw data were re-sampled with 1 s sampling time. The lag plots of the I_D with the time lag (Δt) at 24 ms and 1 s were extracted and shown in Figure 6.19.

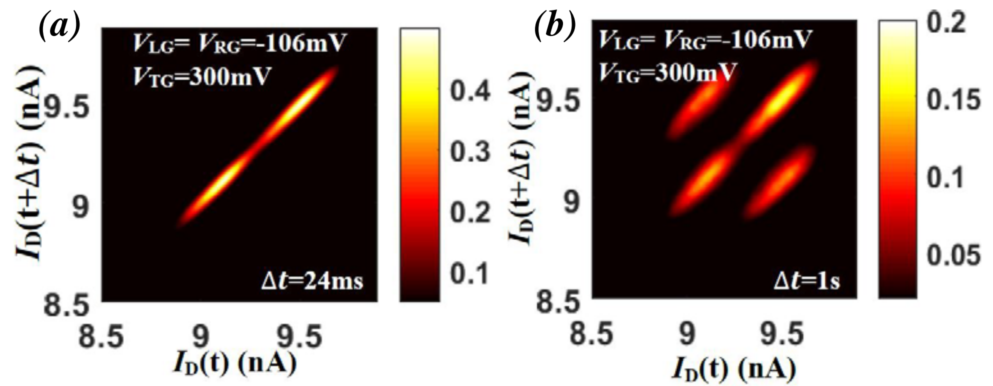


Figure 6.19 The lag plots with time lag (Δt) at (a) 24 ms and (b) 1 s to show the local correlation of the I_D with a certain voltage bias.

By comparing with Δt=24 ms in Figure 6.19(a), the local correlation was improved with Δt=24 ms in Figure 6.19(b). Although increasing Δt could eliminate the local correlation more efficiently, we preferred to keep Δt=1 s due to it was determined by the relaxation time of the quantum dot. Then we used these re-sampled data to generate 0/1 digital numbers and counted the numbers of 1 and 0, shown in Figure 6.20.

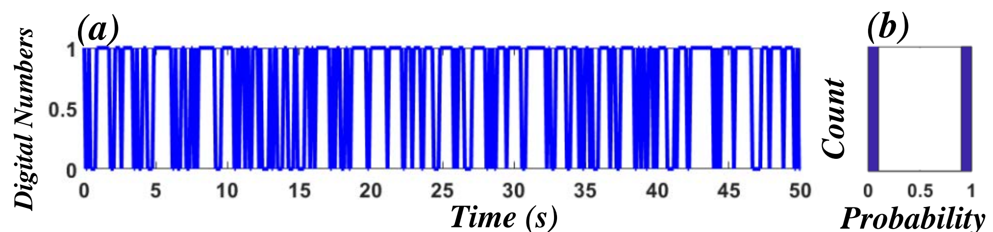


Figure 6.20 (a) The 0 and 1 digital numbers extracted from the re-sampling data. (b) The number count of 0 and 1.

No matter how the amount of 0 and 1 was almost equal. From these original 1-bit digital number sequences, we combined every 2 digits, 3 digits, 4 digits, and 8 digits to generate the 2-bit, 3-bit, 4-bit, and 8-bit random number sequences respectively. The same kinds of sequences were also generated from MATLAB. The Monte Carlo method was utilized to calculate π for the randomness comparison. The results of the MATLAB sequences and our sequences were shown in Table 6.1.

Table 6.1 π calculation by Monte Carlo method

π	2-bit	3-bit	4-bit	8-bit
Our device	2.60	2.69	2.84	3.03
MATLAB	2.71	2.80	3.01	3.08

From the results in Table 6.1, it could be found that the 8-bit random number sequences from our device could be comparable to the MATLAB. The self-correlation plot for the 8-bit sequence from our device was shown in Figure 6.21(a), comparing to that from MATLAB in Figure 6.21(b). The main reason was that the 8-bit sequence degraded the local correlation by longer sampling time. In this case, if the re-sampling time for the raw data was increased, the randomness would be improved.

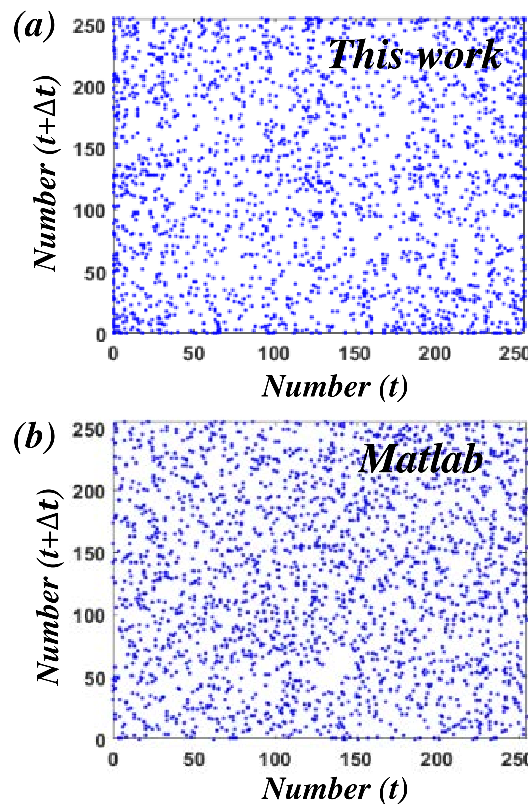


Figure 6.21 (a) Self-correlation plots for 8-bit random numbers generated by our device. (b) Self-correlation plots for 8-bit random numbers generated by MATLAB.

6.5 Summary

From the fundamental physics point of view, the RTSs generated by the multi-gate silicon nanowire transistors could a promising candidate to be a QRNG source. The Top gate could be used to select a certain quantum state to be active in the artificial quantum dot. The First gates could be used to raise the potential barrier and adjust the occupation of the quantum state. One main RTS could be generated with equal probability rates in “high” and “low” states by manipulating the multi-gate bias. Due to the physical structure in our silicon nanowire, the two quantum states that had close energy levels could not be easily separated. The fabrication process for the nanowire should be modified, especially the TMAH step. The channel with a uniform energy level distribution was required. In this case, the ideal one-dimensional transport silicon nanowire might have a better performance. In addition, the speed of our device to generate the numbers was quite slow, maxima 1 bit/s. However, the relaxation could affect by the width of the potential barrier [133]. In our device, the First gate width was 100 nm, which determined the width of the potential barrier. Nowadays, the commercial silicon devices had reached 7 nm node. It was possible to increase the speed by decreasing the First gate width with the help of the advanced fabrication technology in the future.

Chapter 7 Conclusions

In summary, we had successfully developed the RTSs source with an artificial quantum dot in the multi-gate silicon nanowire transistors. The devices were successfully fabricated and characterized in the University of Southampton. Comparing to the previous researches on the RTSs phenomenon, we ended the addiction to the randomly distributed defects for observing the RTs. By using the artificial quantum dot formed by the Top gate and the first gates, we could not only generate the RTSs at room temperature but also manipulate the probabilities of the “high” and “low” states. What's more, we could select different kinds of RTSs generated by different quantum states in the artificial quantum dot to dominate the drain current. More details of the achievements were classified from the perspectives of academic research and technical research.

From the perspective of academic research, the systematic method had been established to measure the typical RTSs in the conventional MOSFETs at different temperatures. In this case, we did not need to measure thousands of devices to find a poor one which showed the typical RTSs anymore. By characterizing different kinds of MOSFETs, two main kinds of typical RTSs were analyzed and explained by the proper physical models. The fundamental theory of the RTSs in silicon devices had been learned and applied to the design philosophy of the artificial quantum dot devices. With a similar method, the RTSs in the multi-gate silicon nanowire transistors were characterized. It was confirmed that the RTSs generated by the artificial quantum could be manipulated by the multi-gate biases. The Top gate voltage controlled the energy level in the quantum dot. The two first gates voltages controlled the potential barriers for the quantum dot and selected the different quantum states to generate different RTSs. By clarifying the operating principle of our multi-gate devices, we evaluated the RTSs quality in different bias conditions as a potential QRNGs source. Although there were still some issues like low speed and mixed RTSs, the RTSs from the artificial quantum dot certainly had the chance to be a promising QRNGs source with advanced CMOS technologies. In addition, the quantum confinement was investigated in our silicon nanowire with special cross-sections. The multi-channels with different energy level model was pointed out, which in some extent helped to explain the multi-RTSs from our artificial quantum dot.

From the perspective of technical research, the complete nano-fabrication process flow had been learned for CMOS technology. During the fabrication processes, my main jobs were to modify the design mask, follow the main steps as an assistant for the senior researches, and do the necessary electrical test after finishing the key steps. The basic processes could be done by myself, such as spin-on-dopant, photolithography, furnace oxidation, RTA, dry etching, wet etching, evaporation, and so on. In order to measure our devices at low temperatures, the cryostat system was redesigned and modified to a low noise level system. The mechanical fabrication part was done by

the mechanical workshop at the University of Southampton. The low-temperature measurement technique and low-noise measurement method had been clarified through these experiences. In addition, the device packaging processes were also developed for our group to measure the devices in different cryostats.

For future work, as our artificial RTSs source was still in the initial stage for the QRNG application, many issues should be solved to become a practical QRNG source. The first issue was the RTSs quality. In our devices, we could not completely eliminate other RTSs to effect the main RTSs due to the non-uniform shape of our silicon nanowire. The different quantum levels were introduced into the system by the quantum confinement in the silicon nanowire. Although it could be balanced in the post-processes, we preferred to improve the performance for the fundamental physics point of view. If the silicon nanowire could perform as a perfect one-dimensional transport device, this issue might be solved. In this case, the further low-temperature measurement might be expected to reach better performance or the nanowire definition process in the fabrication should be improved. In addition, by using advanced lithography technology, the width of the potential barriers could be decreased. That would increase the tunneling rate of the electrons or holes. Then the speed of the artificial RTSs source could be improved.

On the other hand, the main investigation for the QRNG source was from the physics point of view in this project. In order to develop this physics model to be real products, the assessment of the QRNGs should be done further. That is the standard way to evaluate the quality of the randomness of the random bits based on the probability theory and information theory [151]. The typical analyses usually include autocorrelation test, visual analysis, joint probability mass function test, mutual information test, entropy test, bias test, and statistical test. Some of the similar tests have been down initially in this work such as autocorrelation test and visual analysis. Obviously, certain amount of assessment work needs to do in the future. Especially for the statistical tests, it can be used to evaluate a sequence whether it is a turely random sequence or not. The two commonly used test suites were called NIST [152] (From National Institute of Standards and Technology) and TestU01 [153].

The outlook is that this kind of QRNGs device could be integrated into the CPU core and generate high-quality random bits as an independent component. In this case, it could be widely applied for the high-security cryptography and communications.

Appendix A Publication List

Journal Publication:

F. Liu, K. Ibukuro, M. Husain, Z. Li, J. Hillier, I. Tomita, Y. Tsuchiya, H. Rutt, and S. Saito. Manipulation of random telegraph signals in a silicon nanowire transistor with a triple gate. *Nanotechnology*. **29**(47):475201, 2018. https://doi.org/10.1088/1361-6528/aadfa6

Z. Li, M. Sotto, **F. Liu**, M. K. Husain, H. Yoshimoto, Y. Sasago, D. Hisamoto, I. Tomita, Y. Tsuchiya, and S. Saito. Random telegraph noise from resonant tunnelling at low temperatures. *Sci. Rep.*, **8**:250, 2018. https://doi.org/10.1038/s41598-017-18579-1.

S. Saito, Z. Li, H. Yoshimoto, I. Tomita, Y. Tsuchiya, Y. Sasago, H. Arimoto, **F. Liu**, M. K. Husain, D. Hisamoto, H. N. Rutt, and S. Kurihara. Quantum Dipole Effects in a Silicon Transistor under High Electric Fields. *J. Phys. Soc. Jpn.* 87(9):094801, 2018. https://doi.org/10.7566/JPSJ.87.094801.

J. Byers, K. Debnath, H. Arimoto, M. K. Husain, M. Sotto, Z. Li, **F. Liu**, K. Ibukuro, A. Khokhar, K. Kiang, S. A. Boden, D. J. Thomson, G. T. Reed, and S. Saito. Silicon slot fin waveguide on bonded double-SOI for a low-power accumulation modulator fabricated by an anisotropic wet etching technique. *Opt. Express*, **26**(25):33180-33191, 2018. https://doi.org/10.1364/OE.26.033180.

K. Ibukuro, M. K. Husain, Z. Li, J. Hillier, **F. Liu**, Y. Tsuchiya, H. N. Rutt, and S. Saito. Single electron memory effect using random telegraph signals at room temperature. *Front. Phys.* 7:152. 2019. https://doi.org/10.3389/fphy.2019.00152.

International Conference Publication:

F. Liu, M. K. Husain, Z. Li, M. S. H. Sotto, D. Burt, J. D. Fletcher, M. Kataoka, Y. Tsuchiya, and S. Saito. Transport properties in silicon nanowire transistors with atomically flat interfaces. *2017 IEEE Electron Devices Technology and Manufacturing Conference (EDTM)*, 193-195, 2017.

Z. Li, M. S. Sotto, **F. Liu**, M. K. Husain, I. Zeimpekis, H. Yoshimoto, K. Tani, Y. Sasago, D. Hisamoto, J. D. Fletcher, and M. Kataoka. Random-telegraphnoise by resonant tunnelling at low temperatures. *2017 IEEE Electron Devices Technology and Manufacturing Conference (EDTM)*, 172-174, 2017.

D. Burt, A. Z. Al-Attili, Z. Li, **F. Liu**, K. Oda, N. Higashitarumizu, Y. Ishikawa, O. M. Querin, F. Gardes, R. W. Kelsall, and S. Saito. Strain-engineering in Germanium membranes towards light sources on Silicon. *2017 IEEE Electron Devices Technology and Manufacturing Conference (EDTM)*, 92-94, 2017.

Bibliography

[1] G. E. Moore. Cramming more components onto integrated circuits. *Proceeding of the IEEE*, **86**(1):82-85, 1998.

[2] J-M Woo, H-H Park, H. S. Min, Y. J. Park, S-M. Hong, and C. H. Park, Statistical analysis of random telegraph noise in CMOS image sensors. *International Conference on Simulation of Semiconductor Processes and Devices 2008(SISPAD 2008)*, 77-80, 2008.

[3] H. Kurata, K. Otsuga, A. Kotabe, S. Kajiyama, T. Osabe, Y. Sasago, S. Narumi, K. Tokami, S. Kamohara, and O. Tsuchiya. Random telegraph signal in flash memory: Its impact on scaling of multilevel flash memory beyond the 90-nm node. *IEEE J. Solid-State Circuits*, **42:**1362-1369, 2007.

[4] M. Tanizawa, S. Ohbayashi, T. Okagaki, K. Sonoda, K. Eikyu, Y. Hirano, K. Ishikawa, O. Tsuchiya, and Y. Inoue. Application of a statistical compact model for random telegraph noise to scaled-SRAM Vmin analysis. *VLSI Technology 2010 Symposium on*, 95-96, 2010.

[5] E. Simoen and C. Claeys. *Random Telegraph Signals in Semiconductor Devices*. IOP Publishing, 2016.

[6] K. S. Ralls, W. J. Skocpol, L. D. Jackel, R.E. Howard, L.A. Fetter, R.W. Epworth, and D. M. Tennant. Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency ($\frac{1}{f}$?) Noise. *Phys. Rev. Lett.*, **52**:228, 1984.

[7] M. J. Kirton, M. J. Uren, S. Collins, M. Schulz, A. Karmann and K. Scheffer. Individual defects at the Si:SiO_2 interface. *Semicond. Sci. and Technol.*, **4**(12):1116, 1989.

[8] C. Leyris, F. Martinez, A. Hoffmann, M. Valenz, and J. C. Vildeuil. N-MOSFET oxide trap characterization induced by nitridation process using RTS noise analysis. *Microelectron. Reliab.*, **47**(1):41-45, 2007.

[9] S. Kawahito and N. Kawai. Column parallel signal processing techniques for reducing thermal and RTS noises in CMOS image sensors. *Proceeding of the IEEE*, 226-229, 2007.

[10] N. V. Amarasinghe, Z. Çelik-Butler, and P. Vasina. Characterization of oxide traps in 0.15 μm^2 MOSFETs using random telegraph signals. *Microelectron. Reliab.*, **40**(11):1875-1881, 2000.

[11] H. H. Mueller, D. Wörle, and M. Schulz. Evaluation of the Coulomb energy for single-electron interface trapping in sub-μm metal-oxide-semiconductor field-effect transistors. *J. Appl. Phys.*, **75**:2970, 1994.

[12] D. H. Cobden and M. J. Uren. Random telegraph signals from liquid helium to room temperature. *Microelectron. Eng.*, **22**(1-4):163-170, 1993.

[13] M. Herrero-Collantes and J. C. Garcia-Escartin. Quantum random number generators. *Rev. Mod. Phys.*, **89**:015004, 2017.

[14] R. R. Coveyou and R. D. MacPherson. Fourier Analysis of Uniform Random Number Generators. *Journal of the ACM*, **14**:100-119, 1967.

[15] H. A. Haus. *Electromagnetic noise and quantum optical measurements*. Springer Science & Business Media, 2012.

[16] L. I. Berger. *Semiconductor materials*. CRC Press, 1997.

[17] S. K. Tawfeeq. A random number generator based on single-photon avalanche photodiode dark counts. *J. Light. Technol.*, **27**(24):5665-5667, 2009.

[18] A. Uchida, K. Amano, M. Inoue, K. Hirano, S. Naito, H. Someya, I. Oowada, T. Kurashige, M. Shiki, S. Yoshimori, K. Yoshimura, and P. Davis. Fast physical random bit generation with chaotic semiconductor lasers. *Nat. Photonics.*, **2**:728-732, 2008.

[19] ID Quantique. QUANTIS random number generator. http://www.idquantique.com/random-number-generation. 2019

[20] R. Hughes and J. Nordholt. Strengthening the Security Foundation of Cryptography With Whitewood's Quantum-Powered Entropy Engine. 2016.

[21] Micro Photon Devices. Quantum random number generator. http://www.micro-photon-devices.com/. 2019

[22] Qutools. QuRNG—Quantum Random Number Generator. https://www.qutools.com/quRNG/. 2019

[23] PicoQuant. PQRNG Quantum Random Number Generator. http://www.picoquant.com. 2019.

[24] Quintessence Labs. Fast Quantum Random Number Generation. http://www.quintessencelabs.com. 2019

[25] M. Ghioni, A. Gulinatti, I. Rech, F. Zappa, and S. D. Cova. Progress in Silicon Single-Photon Avalanche Diodes," *IEEE J. Sel. Top. Quantum Electron.*, **13**:852–862, 2007.

[26] G. S. Buller and R. J. Collins. Single-photon generation and detection. *Meas. Sci. Technol.*, **21**:012002, 2010.

[27] M. D. Eisaman, J. Fan, A. L. Migdall, and S. V. Polyakov. Single-photon sources and detectors. *Rev. Sci. Instrum.*, **82**:071101, 2011.

[28] T. Jennewein, U. Achleitner, G. Weihs, H. Weinfurter, and A. Zeilinger. A Fast and Compact Quantum Random Number Generator. *Rev. Sci. Instrum.*, **71**:1675–1680, 2000.

[29] M. Wahl, M. Leifgen, M. Berlin, T. Röhlicke, H. Rahn, and O. Benson. An ultrafast quantum random number generator with provably bounded output bias based on photon arrival time measurements. *Appl. Phys. Lett.*, **98**:171105, 2011.

[30] M. Ren, E. Wu, Y. Liang, Y. Jian, G. Wu, and H. Zeng. Quantum random-number generator based on a photon-number-resolving detector. *Phys. Rev. A.* **83**:023820, 2011.

[31] W. Wei and H. Guo. Bias-free true random-number generator. *Opt. Lett.*, **34**:1876, 2009.

[32] N. Tega, H. Miki, T. Osabe, A. Kotabe, K. Otsuga, H. Kurata, S. Kamohara, K. Tokami, Y. Ikeda, and R. Yamada. Anomalously large threshold voltage fluctuation by complex random telegraph signal in floating gate flash memory. *IEDM Tech. Dig.*, 9506455, 2006.

[33] N. Tega, H. Miki, M. Yamaoka, H. Kume, T. Mine, T. Ishida, Y. Mori, R. Yamada, and K. Torii. Impact of threshold voltage fluctuation due to random telegraph noise on scaled-down SRAM. *Proc. 46th Annual Int. Reliability Physics Symp. (IRPS08),* 541–6, 2008.

[34] N. Tega, H. Miki, F. Pagette, D. Frank, A. Ray, M. Rooks, W. Haensch, and K. Torii. Increasing threshold voltage variation due to random telegraph noise in FETs as gate lengths scale to 20 nm. *Symp. on VLSI Technol. Dig. Tech.*, 10826235, 2009.

[35] 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. *Symp. on VLSI Technol. Dig. Tech.*, 10826237, 2009.

[36] M. Yamaoka, H. Miki, A. Bansal, S. Wu, D. Frank, E. Leobandung, and K. Torii. Evaluation methodology for random telegraph noise effects in SRAM arrays. *IEDM Tech. Dig.*, 12504333, 2011.

[37] M. J. Kirton and M. J. Uren. Capture and emission kinetics of individual Si:SiO_2 interface states. *Appl. Phys. Lett.*, **48**(19):1270, 1986.

[38] K. Nishiguchi, Y. Ono, and A. Fujiwara. Single-electron counting statistics of shot noise in nanowire Si metal-oxide semiconductor field-effect transistors. *Appl. Phys. Lett.*, **98**:193502, 2011.

[39] K. Nishiguchi, Y. Ono, and A. Fujiwara. Single-electron thermal noise. *Nanotechnology*, **25**:275201, 2014.

[40] K. Chida, K. Nishiguchi, G. Yamahata, H. Tanaka, and A. Fujiwara. Thermal-noise suppression in nano-scale Si field-effect transistors by feedback control based on single-electron detection. *Appl. Phys. Lett.*, **107**:073110, 2015.

[41] K. Chida, S. Desai, K. Nishiguchi, and A. Fujiwara. Power generator driven by Maxwell's demon. *Nat. Commun.*, **8**:15310, 2017.

[42] Y. Shi, H. M. Bu, X. L. Yuan, S. L. Gu, B. Shen, P. Han, R. Zhang, and Y. D. Zheng. Switching kinetics of interface states in deep submicrometre SOI n-MOSFETs. *Semicond. Sci. and Technol.*, **16**(1):21-25, 2001.

[43] J. H. Scofield, N. Borland, and D. M. Fleetwood. Temperature-independent switching rates for a random telegraph signal in a silicon metal–oxide–semiconductor field-effect transistor at low temperatures. *Appl. Phys. Lett.*, **76**:3248, 2000.

[44] M. J. Uren, D. J. Day, and M. J. Kirton. 1/f and random telegraph noise in silicon metal-oxide-semiconductor field-effect transistors. *Appl. Phys. Lett.*, **47**:1195, 1985.

[45] Z. Shi, J. P. Mieville, and M. Dutoit. Random telegraph signals in deep submicron n-MOSFET's. *IEEE Trans. Electron Dev.*, **41**(7):1161-1168, 1994.

[46] N. V. Amarasinghe and Z. Çelik-Butler. Complex random telegraph signals in 0.06 µm2 MDD n-MOSFETs. *Solid-State Electron.*, **44**(6):1013-1019, 2000.

[47] K. K. Hung, P. K. Ko, C. Hu, Y. C. Cheng. Random telegraph noise of deep-submicrometer MOSFETs. *IEEE Electron Dev. Lett.*, **11**(2):90-92, 1990.

[48] A. Godoy, F. Gámiz, A. Palma, J. A. Jiménez-Tejada, and J. E. Carceller. Random telegraph signal amplitude in submicron n-channel metal oxide semiconductor field effect transistors. *Appl. Phys. Lett.*, **70**:2153, 1997.

[49] E. Simoen, B. Dierickx, B. D. Canne, F. Thoma, C. Claeys. On the gate- and drain-voltage dependence of the RTS amplitude in submicron MOSTs. *Appl. Phys. A*, **58**(4):353-358, 1994.

[50] O. Roux dit Buisson, G. Ghibaudo, J. Brini. Model for drain current RTS amplitude in small-area MOS transistors. *Solid-State Electron.*, **35**(9):1273-1276, 1992.

[51] M. H. Tsai and T. P. Ma. The impact of device scaling on the current fluctuations in MOSFET's. *IEEE Trans. Electron Dev.*, **41**(11):2061-2068, 1994.

[52] H. Ishikuro, T. Saraya, T. Hiramoto, and T. Ikoma. Extremely large amplitude random telegraph signals in a very narrow split-gate MOSFET at low temperatures. *Jpn. J. Appl. Phys.*, **35**:858-60, 1996.

[53] H. M. Bu, Y. Shi, X. L. Yuan, Y. D. Zheng, S. H. Gu, H. Majima, H. Ishikuro, T. Hiramoto. Impact of the device scaling on the low-frequency noise in n-MOSFETs. *Appl. Phys. A*, **71**(2):133-136, 2000.

[54] H. M. Bu, Y. Shi, X. L. Yuan, J. Wu, S. L. Gu, and Y. D. Zheng, H. Majima, H. Ishikuro, and T. Hiramoto. Random telegraph signals and low-frequency noise in n-metal-oxide-semiconductor field-effect transistors with ultranarrow channels. *Appl. Phys. Lett.*, **76**(22):3259, 2000.

[55] A. Avellan, W. Krautschneider, and S. Schwantes. Observation and modeling of random telegraph signals in the gate and drain currents of tunneling metal–oxide–semiconductor field-effect transistors. *Appl. Phys. Lett.*, **78**(18):2790, 2001.

[56] D. Veksler, G. Bersuker, S. Rumyantsev, M. Shur, H. Park, C. Young, K. Y. Lim, W. Taylor, and R. Jammy. Understanding noise measurements in MOSFETs: the role of traps structural relaxation. *Proc. Int. Reliability Physics Symp. (IRPS10)*, 73-9, 2010.

[57] D. H. Cobden, M. J. Uren, and M. J. Kirton. Entropy measurements on slow Si/SiO2 interface states. *Appl. Phys. Lett.*, **56**(13):1245, 1990.

[58] M. Schulz, A. Pappas, and J. Vennemann. Single-electron trapping and telegraph noise in µm-sized MOSFETS. *AIP Conference Proceedings*, **282**(1):124, 1992.

[59] Z. Celik-Butler, P. Vasina, and N. V. Amarasinghe. A method for locating the position of oxide traps responsible for random telegraph signals in submicron MOSFETs. *IEEE Trans. on Electron Dev.*, **47**(3):646-648, 2000.

[60] N. B. Lukyanchikova, M. V. Petrichuk, N. P. Garbar, E. Simoen and C. Claeys. Impact of the free electron distribution on the random telegraph signal capture kinetics in submicron n-metal-oxide-semiconductor field-effect transistors. *Appl. Phys. Lett.*, **73**(17):2444, 1998.

[61] M. Schulz. Coulomb energy of traps in semiconductor space-charge regions. *J. Appl. Phys.*, **74**(4):3649, 1993.

[62] M. Schulz and A. Karmann. Individual, attractive defect centers in the SiO2-Si interface of µm-sized MOSFETs. *Appl. Phys. A*, **52**(2):104-111, 1991.

[63] M. Schulz and A. Karmann. Single, individual traps in MOSFETs. *Phys. Scr.*, **T35**:273, 1991.

[64] M.J. Kirton and M.J. Uren. Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/ƒ) noise. *Adv. Phys.*, **38**(4):367-468, 1989.

[65] J.P. Campbell, J. Qin, K. P. Cheung, L. C. Yu, J. S. Suehle, A. Oates, and K. Sheng. Random telegraph noise in highly scaled nMOSFETs. *Proc. 47th Annual Int. Reliability Physics Symp. (IRPS09)*, 382-8, 2009.

[66] D. H. Cobden, M. J. Uren, and M. Pepper. Dissipative tunneling in two-state systems at the Si/SiO_2 interface. *Phys. Rev. Lett.*, **71**(25):4230, 1993.

[67] D. H. Cobden and B. A. Muzykantskii. Finite-Temperature Fermi-Edge Singularity in Tunneling Studied Using Random Telegraph Signals. *Phys. Rev. Lett.*, **75**(23):4274, 1995.

[68] H. H. Mueller, U. Schirl, M. Schulz. Charge carrier fluctuations in sub-μm MOSFETs. *Proc. 3rd Int. Symp. On the Physics and Chemistry of SiO2 and the Si-SiO2 Interface.*, 1996.

[69] H. H. Mueller and M. Schulz. Conductance modulation of submicrometer metal-oxide semiconductor field-effect transistors by single-electron trapping. *J. Appl. Phys.*, **79**(8):4178, 1996.

[70] L. K. J. Vandamme, D. Sodini, and Z. Gingl. On the anomalous behavior of the relative amplitude of RTS noise. *Solid-State Electron.*, **42**(6):901-905, 1998.

[71] P. Restle and A. Gnudi. Internal probing of submicron FETs and photoemission using individual oxide traps. *IBM J. Res. Dev.*, **34**:227-42, 1990.

[72] P. Restle. Individual oxide traps as probes into submicron devices. *Appl. Phys. Lett.*, **53**(19):1862, 1988.

[73] P. Zhang, Y. Q. Zhuang, Z. F. Ma, L. Bao, L. Du, and J. L. Bao. The impact of gate–image charge on RTS amplitudes in ultra-thin gate oxide n-MOSFETs. *Semicond. Sci. and Technol.*, **23**(12):125037, 2008.

[74] M. E. Welland and R. H. Koch. Spatial location of electron trapping defects on silicon by scanning tunneling microscopy. *Appl. Phys. Lett.*, **48**(11):724, 1986.

[75] K. R. Farmer, C. T. Rogers, and R. A. Buhrman. Localized-State Interactions in Metal-Oxide-Semiconductor Tunnel Diodes. *Phys. Rev. Lett.*, **58**(21):2255, 1987.

[76] K. R. Farmer and R. A. Buhrman. Defect dynamics and wear-out in thin silicon oxides. *Semicond. Sci. and Technol.*, **4**(12):1084-108, 1989.

[77] A. Avellan, W. Krautschneider, and B. Sell. Softbreakdown and RTS in gate and drain currents of MOS transistors with ultrathin oxides. *Proc. Of the 16th Int. Conf. on Noise in Physical Systems and 1/f Fluctuations (ICNF2001)*, 375-8, 2001.

[78] A. Avellan, D. Schroeder, and W. Krautschneider. Modeling random telegraph signals in the gate current of metal–oxide–semiconductor field effect transistors after oxide breakdown. *J. Appl. Phys.*, **94**(1):703, 2003.

[79] F. Martinez, C. Leyris, G. Neau, M. Valenza, A. Hoffmann, J. C. Vildeuil, E. Vincent, F. Boeuf, T. Skotnicki, M. Bidaud, D. Barge, and B. Tavel. Oxide traps characterization of 45 nm MOS transistors by gate current R.T.S. noise measurements. *Microelectron. Eng.*, **80**:57-57, 2005.

[80] J. Armand, F. Martinez, M. Valenza, K. Rochereau, E. Vincent. Slow oxide trap density profile extraction using gate current low-frequency noise in ultrathin oxide MOSFETs. *Microelectron. Eng.*, **84**(9-10):2382-2385, 2007.

[81] B. Kaczer, M. Toledano-Luque, W. Goes, T. Grasser, G. Groeseneken. Gate current random telegraph noise and single defect conduction. *Microelectron. Eng.*, **109**:123-125, 2013.

[82] C. M. Chang, S. S. Chung, Y. S. Hsieh, L. W. Cheng, C. T. Tsai, G. H. Ma, S. C. Chien, and S. W. Sun. The observation of trapping and detrapping effects in high-k gate dielectric MOSFETs by a new gate current Random Telegraph Noise (IG-RTN) approach. *Tech. Dif. Int. Electron Devices Meeting (IEDM08).*, 2008.

[83] S. Lee, H. J. Cho, Y Son, D. S. Lee, and H. Shin. Characterization of oxide traps leading to RTN in high-k and metal gate MOSFETs. *Tech. Dif. Int. Electron Devices Meeting (IEDM09).*, 2009.

[84] H. J. Cho, S. Lee, B. G. Park, H. Shin. Extraction of trap energy and location from random telegraph noise in gate leakage current (Ig RTN) of metal–oxide semiconductor field effect transistor (MOSFET). *Solid-State Electron.*, **54**(4):362-367, 2010.

[85] H. J. Cho, Y. Son, B. C. Oh, S. Lee, J. H. Lee, B. G. Park. Study on Time Constants of Random Telegraph Noise in Gate Leakage Current Through Hot-Carrier Stress Test. *IEEE Electron Dev. Lett.*, **31**(9):1029-1031, 2010.

[86] P. Benioff. The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by Turing machines. *J. Stat. Phys.*, **22**(5):563–591, 1980.

[87] R. P. Feynman. Simulating physics with computers. *Int. J. Theor. Phys.*, **22**(5):563-591, 1980.

[88] D. Deutsch. Quantum Theory, the Church-Turing Principle and the Universal Quantum Computer. *Proc. of the Royal Society of London A*, **400**(1818):97–117, 1985.

[89] N. Gershenfeld and I. L. Chuang. Quantum Computing with Molecules. *Sci. Am.*, **278**(6):66-71, 1998.

[90] M. Isida and H. Ikeda. Random number generator. *Ann. Inst. Stat. Math.*, **8**:119-126, 1956.

[91] C. H. Vincent. The generation of truly random binary numbers. *J. Phys. E.*, 3:594, 1970.

[92] H. Friedman. Geiger Counter Tubes. *Proc. IREE Aust.*, **37**:791-808, 1949.

[93] M. P. Silverman, W. Strange, C. R. Silverman, and T. C. Lipscombe. Tests of alpha-, beta-, and electron capture decays for randomness. *Phys. Lett. A.*, **262**:265–73, 1999.

[94] G. Lutz. *Semiconductor Radiation Detectors*. Springer-Verlag, 2007.

[95] G. F. Knoll. Radiation Detection and Measurement. Wiley, 2010.

[96] A. Alkassar, T. Nicolay, and M. Rohe. Obtaining true-random binary numbers from a weak radioactive source. *Computational Science and Its Applications—ICCSA 2005*, **3481**:634–646, 2005.

[97] R. Duggirala, A. Lal, and S. Radhakrishnan, *Radioisotope Decay Rate Based Counting Clock.* Springer-Verlag, 2010.

[98] J. Walker, HotBits: Genuine random numbers, generated by radioactive decay. *http://www.fourmilab.ch/hotbits/*, 1996.

[99] H. Schmidt. Quantum Mechanical Random Number Generator. *J. Appl. Phys.*, **41**:462–468, 1970.

[100] W. Schottky. Über spontane Stromschwankungen in verschiedenen Elektrizitätsleitern. *Ann. Phys.*, **362**:541–567. 1918.

[101] S. A. Wilber. Entropy Analysis and System Design for Quantum Random Number Generators in CMOS Integrated Circuits. *https://comscire.com/files/whitepaper/Pure_Quantum_White_Paper.pdf*, 2014.

[102] P. I. Somlo. Zener-diode noise generators. *Electron. Lett.*, **11**:290, 1975.

[103] M. Stipčević. Fast nondeterministic random bit generator based on weakly correlated physical events. *Rev. Sci. Instrum.*, **75**:4442, 2004.

[104] R. Landauer. Solid-state shot noise. *Phys. Rev. B*, **47**:16427–16432, 1993.

[105] J. R. Klauder and E. C. G. Sudarshan. *Fundamentals of Quantum Optics.* Benjamin, 1968.

[106] R. Loudon. *The Quantum Theory of Light.* Oxford University Press, 2001.

[107] J. G. Rarity, P. C. M. Owens, and P. R. Tapster. Quantum Random-number Generation and Key Sharing. *J. Mod. Opt.*, **41**:2435–2444, 1994.

[108] T. Jennewein, U. Achleitner, G. Weihs, H. Weinfurter, and A. Zeilinger. A Fast and Compact Quantum Random Number Generator. *Rev. Sci. Instrum.*, **71**:1675–1680, 2000.

[109] M. Fürst, H. Weier, S. Nauerth, D. G. Marangon, C. Kurtsiefer, and H. Weinfurter. High speed optical quantum random number generation. *Opt. Express*, **18**:13029-13037, 2010.

[110] Y. Jian, M. Ren, E. Wu, G. Wu, and H. Zeng. Two-bit quantum random number generator based on photon-numberresolving detection. *Rev. Sci. Instrum.*, **82**:073109, 2011.

[111] M. Stipčević and B. M. Rogina. Quantum random number generator based on photonic emission in semiconductors. *Rev. Sci. Instrum.*, **78**:045104, 2007.

[112] A. Khanmohammadi, R. Enne, M. Hofbauer, and H. Zimmermanna. A Monolithic Silicon Quantum Random Number Generator Based on Measurement of Photon Detection Time. *IEEE Photonics J.*, **7**:1–13, 2015.

[113] M. J. Collett, R. Loudon, and C.W. Gardiner. Quantum Theory of Optical Homodyne and Heterodyne Detection. *J. Mod. Opt.*, **34**:881-902, 1987.

[114] T. Symul, S.M. Assad, and P. K. Lam. Real time demonstration of high bitrate quantum random number generation with coherent laser light. *Appl. Phys. Lett.*, 98:231103, 2011.

[115] K. Uchida, T. Tanamoto, R. Ohba, S. Yasuda, and S. Fujita. Single-electron random-number generator (RNG) for highly secure ubiquitous computing applications. *Digest. Int. Electron Devices Meeting*, 177-180, 2002.

[116] T. P. Ma and Paul V. Dressendorfer. *Ionizing Radiation Effects in MOS Devices and Circuits*. John Wiley & Sons, 1989.

[117] R. W. Keyes. Fundamental limits of silicon technology. *Proc. of the IEEE 89*, 227–239, 2001.

[118] S. Takagi, M. Iwase, and A. Toriumi. On the universality of inversion-layer mobility in n- and p-channel MOSFETs. *Digest. Int. Electron Devices Meeting*, 398-401, 1988.

[119] H. Grabert and M. H. Devoret. *Single Charge Tunneling: Coulomb Blockade Phenomena in Nanostructures.* Springer, 1992.

[120] Y. Taur and T. H. Ning. *Fundamentals of Modern VLSI Devices*. Cambridge University Press, 2009.

[121] Z. Li, M. K. Husain, H. Yoshimoto, K. Tani, Y. Sasago, D. Hisamoto, J. D. Fletcher, M. Kataoka, Y. Tsuchiya, and S Saito. Single carrier trapping and de-trapping in scaled silicon complementary metal-oxide-semiconductor field-effect transistors at low temperatures. *Semicond. Sci. and Technol.*, **32**(7):075001, 2017.

[122] T. Ando, A. B. Fowler, and F. Stern. Electronic properties of two-dimensional systems *Rev. Mod. Phys.*, **54**:437, 1982.

[123] A. Hartstein and N. F. Albert. Determination of the inversion-layer thickness from capacitance measurements of metal-oxide-semiconductor field-effect transistors with ultrathin oxide layers. *Phys. Rev. B*, **38**:1235, 1988.

[124] S. Saito, K. Torii, M. Hiratani, and T. Onai. Analytical quantum mechanical model for accumulation capacitance of mos structures. *IEEE Electron Device Lett.* **23**:348–50, 2002.

[125] N. Yamauchi, J. Hajjar, and R. Reif. Polysilicon thin-film transistors with channel length and width comparable to or smaller than the grain size of the thin film. *IEEE Trans. Electron Dev.*, **38**:55–60, 1991.

[126] W. B. Jackson, N. M. Johnson, and D. K. Biegelsen. Density of gap states of silicon grain boundaries determined by optical absorption. *Appl. Phys. Lett.*, **43**:195-7, 1983.

[127] A. Karwath and M. Schulz. Deep level transient spectroscopy on single, isolated interface traps in field-effect transistors. *Appl. Phys. Lett.*, **52**(8):634, 1988.

[128] L. K. J. Vandamme and M. Macucci. 1/f and RTS noise in submicron devices: Faster is noisier. *In 4th International Conference on Unsolved Problems of Noise and Fluctuations in Physics, Biology and High Technology*, 436–443, 2005.

[129] E. P. Vandamme and L. K. J. Vandamme. Critical discussion on unified 1/f noise models for mosfets. *IEEE Trans. Electron Dev.*, **47**:2146–2152, 2000.

[130] R. Nuryadi, H. Ikeda, Y. Ishikawa, and M. Tabe. Current fluctuation in single-hole transport through a two-dimensional Si multi dot. *Appl. Phys. Lett.*, **86**:133106, 2005.

[131] V. Golovach et al. Single-dopant resonance in a single-electron transistor. *Phys. Rev. B*, **83**:075401, 2011.

[132] B. J. Villis et al. Direct detection of a transport-blocking trap in a nanoscaled silicon single-electron transistor by radio-frequency reflectometry. *Appl. Phys. Lett.*, **104**:233503, 2014.

[133] D. J. Griffiths and D. F. Schroeter. *Introduction to quantum mechanics*. Cambridge University Press, 2018.

[134] *High Field Measurement System User Manual. Cryogenic Ltd*. Company, 2009.

[135] H. Mizuta and T. Tanoue. *The physics and applications of resonant tunnelling diodes*. Cambridge University Press, 2006.

[136] R. S. Wagner and W. C. Ellis. Vapor-liquid-solid mechanism of single crystal growth. *Appl. Phys. Lett.*, **4**:89, 1964.

[137] M. Kobayashi and T. Hiramoto. Experimental study on quantum confinement effects in silicon nanowire metal-oxidesemiconductor field-effect transistors and single-electron transistors. *J. Appl. Phys.*, **103**(5):053709, 2008.

[138] M. Lundstrom. *Fundamentals of carrier transport*. Cambridge University Press, 2009.

[139] D. D. D. Ma, C. S. Lee, F. C. K. Au, S. Y. Tong, and S. T. Lee. Small-Diameter Silicon Nanowire Surfaces. *Science*, 299(5614):1874-1877, 2003.

[140] X. Zhao, C. M. Wei, L. Yang, and M. Y. Chou. Quantum Confinement and Electronic Properties of Silicon Nanowires. *Phys. Rev. Lett.*, **92**(23):236805, 2004.

[141] G. P. Lansbergen, R. Rahman, C. J. Wellard, I. Woo, J. Caro, N. Collaert, S. Biesemans, G. Klimeck, L. C. L. Hollenberg, and S. Rogge. Gate-induced quantum-confinement transition of a single dopant atom in a silicon FinFET. *Nat. Phys.*, **4**:656–661, 2008.

[142] F. Liu, K. Ibukuro, M. K. Husain, Z. Li, J. Hillier, I. Tomita, Y. Tsuchiya, H. Rutt and S. Saito. Manipulation of random telegraph signals in a silicon nanowire transistor with a triple gate. *Nanotechnology*, **29**(47):475201, 2018.

[143] Z. Li, M. Sotto, F. Liu, M. K. Husain, H. Yoshimoto, Y. Sasago, D. Hisamoto, I. Tomita, Y. Tsuchiya, and S. Saito. Random telegraph noise from resonant tunnelling at low temperatures. *Sci. Rep.*, **8**:250, 2018.

[144] W. F. Koehl, B. B. Buckley, F. J. Heremans, G. Calusine, and D. D. Awschalom. Room temperature coherent control of defect spin qubits in silicon carbide. *Nature*, **479**:84-87, 2011.

[145] F. N. Hooge, T. G. M. Kleinpenning, and L. K. J. Vandamme. Experimental studies on 1/f noise. *Rep. Prog. Phys.*, **44**:479, 1981.

[146] P. Dutta, and P. M. Horn. Low-frequency fluctuations in solids: 1/f noise. *Rev. Mod. Phys.*, **53**:497, 1981.

[147] L. K. J. Vandamme, and M. Macucci. 1/f And RTS Noise In Submicron Devices: Faster Is Noisier. *AIP Conf. Proc.*, **800**:436, 2005.

[148] X. Chen, L. Wang, B. Li, Y. Wang, X. Li, Y. Liu, and H. Yang. Modeling Random Telegraph Noise as a Randomness Source and its Application in True Random Number Generation. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.*, **35**(9):1435-1448, 2016.

[149] A. Mohanty, K. B. Sutaria, H. Awano, T. Sato, and Y. Cao. RTN in Scaled Transistors for On-Chip Random Seed Generation. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, **25**(8):2248-2257, 2017.

[150] J. Brown, R. Gao, Z. Ji, J. Chen, J. Wu, J. Zhang, B. Zhou, Q. Shi, J. Crowford, and W. Zhang. A low-power and high-speed True Random Number Generator using generated RTN. *IEEE Symposium on VLSI Technology*, 95-96, 2018.

[151] T. M. Cover, J. A. Thomas. *Elements of Information Theory*. Wiley, 2006.

[152] A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. https://csrc.nist.gov/publications/detail/sp/800-22/rev-1a/final. 2019.

[153] The TestU01 web site. http://simul.iro.umontreal.ca/testu01/tu01.html. 2019.

[154] F. Liu, M. K. Husain, Z. Li, M. S. H. Sotto, D. Burt, J. D. Fletcher, M. Kataoka, Y. Tsuchiya, and S. Saito. Transport properties in silicon nanowire transistors with atomically flat

interfaces. *2017 IEEE Electron Devices Technology and Manufacturing Conference (EDTM)*, 193-195, 2017.