University of Southampton Research Repository
ePrints Soton

Copyright © and Moral Rights for this thesis 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 cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder/s. The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given e.g.

AUTHOR (year of submission) "Full thesis title", University of Southampton, name of the University School or Department, PhD Thesis, pagination

http://eprints.soton.ac.uk

UNIVERSITY OF SOUTHAMPTON

Suspended Gate Silicon Nanodot Memory

by

Mario Alberto García Ramírez

A thesis submitted in partial fulfillment for the degree of Doctor of Philosophy

in the
Faculty of Physical and Applied Sciences
Department of Electronics and Computer Science

Supervisors:
Professor Hiroshi Mizuta
Dr. Yoshishige Tsuchiya

October 2011

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL AND APPLIED SCIENCES
DEPARTMENT OF ELECTRONICS AND COMPUTER SCIENCE

Doctor of Philosophy

Suspended Gate Silicon Nanodot Memory

by Mario Alberto García Ramírez

The non-volatile memory market has been driven by Flash memory since its invention more than three decades ago. Today, this non-volatile memory is used in a wide variety of devices and systems from pen drives, mp3 players to cars, planes and satellites. However, the conventional floating gate memory technology in use for flash memory is facing a serious scalability issue, the tunnel oxide thickness cannot be reduced to less than 7 nm as pointed out in the latest international technology roadmap for semiconductors (ITRS2010) [1].

The limit imposed on the tunnel oxide layer reduces the programming and erasing times, the scalability and endurance among other parameters. To overcome those inherent issues, this research is focused on the co-integration of nano-electromechanical systems (NEMS) with metal-oxide-semiconductor (MOS) technology in order to generate a new non-volatile and high speed memory. The memory device that we are proposing is a high-speed non-volatile memory structure called the Suspended Gate Silicon Nanodot Memory (SGSNM) cell. This non-volatile memory device features a MOSFET as a readout element, a silicon nanodot (SiNDs) monolayer as the floating gate and a movable suspended control gate isolated from the floating gate by an oxide layer and by an air-gap. The fundamental component in this novel device is the introduction of a doubly-clamped beam as a movable control gate, in which through this element, the programming and erasing operations take place. To understand the behaviour of the doubly-clamped beam structure, it is analysed by using analytical models such as the doubly-plate capacitor model and also by using two- and three-dimensional (2D and 3D) finite element method (FEM) analysis.

The programming and erasing operations within the SGSNM occur when the suspended control gate is in contact with the tunnel oxide layer. This is the point at which the quantum-mechanical tunnelling mechanism (Fowler-Nordheim) takes place. Through this mechanism, the electrons are allowed to tunnel from the suspended control gate into the memory node and vice versa as a function of the applied voltage (bias). The tunnelling process is numerically analysed by implementing the Tsu-Esaki equation and the transfer matrix method within a homemade program which calculates the current density as a function of the tunnel oxide material and thickness.

Both the suspended control gate and tunnelling process are implemented as analog behavioural models within the SGSNM cell that is simulated by using a commercial circuit simulator. From a transient analysis of the suspended control gate, it was found that the suspended control gate takes 0.8 nsec in pull-in on the tunnel oxide layer for a 1 μm-long doubly-clamped structure. In contrast, the time that the memory node takes in charge and discharge is 1.7 nsec. Hence, the programming and erasing times are a combination between the mechanical pull-in and the charging time, which is 2.5 nsec due the fact that to both operations are symmetrical. Moreover, the suspended control gate was successfully fabricated and suspended. This process was performed by depositing a thin layer of aluminium (500 nm) over the sacrificial layer (poly-Si) by using an e-beam evaporator, which was patterned with doubly-clamped beam features through the photolithographic process. By using a combination of wet and dry etching processes, the aluminium and the sacrificial layer were successfully removed without affecting the substrate (Si-based) or the suspended control gate beam. In addition, Capacitance - Voltage measurements were performed on a set of doubly-clamped beams from which the pull-in effect was successfully obtained. Finally, the footprints for the memory device fabrication process were developed and sketched within the document as well as the design of three photomasks.

Contents

Nomenclature xv

Declaration of Authorship xix

List of Publications xxi

List of Conferences xxiii

Acknowledgements xxv

1 Memories Background 1
- 1.1 Introduction . . . 1
- 1.2 Semiconductor memory devices . . . 2
  - 1.2.1 Flash Memory . . . 3
  - 1.2.2 Channel Hot Electron tunnelling process . . . 4
  - 1.2.3 Fowler-Nordheim tunnelling process . . . 5

2 Suspended Gate Silicon Nanodots Memory 11
- 2.1 Operations performed by the suspended control gate . . . 12
  - 2.1.1 Programming sequence . . . 12
  - 2.1.2 Erasing Sequence . . . 13
  - 2.1.3 Reading Sequence . . . 14

3 Suspended Control Gate Analysis 17
- 3.1 Pull-in analysis . . . 17
- 3.2 Pull-out analysis . . . 19
- 3.3 Numerical Analysis of the Doubly-Clamped Structure . . . 21
  - 3.3.1 Implementation as behavioural models . . . 27
- 3.4 Transient Analysis . . . 31
- 3.5 Suspended Control Gate Scaling Down Process . . . 32

4 Quantum-Mechanical Tunnelling Process: Current Density Analysis 35
- 4.1 Tunnelling Current Density Analysis . . . 37
- 4.2 Tunnel Barrier Analysis . . . 40

5 Suspended Gate Silicon Nanodots Memory implementation 47
- Initial value . . . 49
- Discriminate process . . . 49
- Identification segment . . . 49

Encoding towards the correct pipeline 50
Output value 50
5.1 Electric Circuit Analysis of the SGSNM cell 50
5.2 Charging Time Analysis 55
5.3 Energy analysis 56
5.4 Short-Range Forces Analysis 57
5.5 Comparison with other emerging non-volatile memories 60

6 Fabrication and Characterisation of Suspended Gate Structures 61
6.1 Readout Element 62
6.2 Floating Gate (SiNDs) Layer Deposition 63
6.3 Suspended Control Gate Fabrication Process 66
6.4 Electro-Mechanical Beam Measurements 76

7 Conclusions and Future Work 83
7.1 Future Work 84

A Calibration for the C - V Meter 87

B Photomask Design 89

Bibliography 93

List of Figures

1.1 Schematic diagram of the flash memory structure . . . 3

1.2 Tunnelling mechanisms used to program the Flash memory structures (NAND and NOR). a) Channel hot electron process, when the control gate and the MOSFET are biased (beyond saturation), within the depletion region, the channel is formed and since the bias on the MOSFET is beyond saturation, it creates a pinch-off region. This region gives enough lateral energy to the electrons to tunnel into the floating gate. b) Fowler-Nordheim tunnelling process, in the band diagram (right side) shows that by applying a large positively biased voltage the energy band diagram of the tunnel oxide layer bends until a triangular shape is obtained. Once triangular, electrons are allowed to tunnel through this layer into the floating gate . . . 5

1.3 Schematic diagrams of the programming and erasing operations through both tunnelling mechanisms used in NOR and NAND-type flash memories . . . 6

1.4 Defects generated within the tunnel oxide layer produced by the Fowler-Nordheim or Channel Hot Electron tunnelling processes . . . 6

1.5 Schematic diagram of the silicon nanodots non-volatile memory structure 7

1.6 Defects created within the tunnel oxide layer by action of the programming and erasing operations due to the Fowler-Nordheim or Channel Hot Electron tunnelling processes . . . 8

1.7 Volatile and non-volatile memory characteristics (Courtesy of Prof. Mizuta) . . 9

1.8 The MEMS/NEMS downscaling trend compared with the CMOS extension, in which the hybrid integration of NEMS and CMOS may offer a new approach to the More-than-Moore and Beyond-CMOS domains (Courtesy of Prof. Mizuta) . . . 10

2.1 Suspended Gate Silicon Nanodots Memory (SGSNM) structure . . . 11

2.2 Schematic diagram of the SGSNM cell . . . 13

2.3 Programming sequence performed by the suspended gate when a) the suspended gate electrode is biased, b) the suspended gate bends downwards and eventually collapses allowing the tunnelling process (electrons) into the floating gate layer and c) the suspended gate detaches from the oxide layer and the electrons are trapped within the floating gate . . . 14

2.4 Erasing sequence of the SGSNM cell performed by the suspended control gate when a) the doubly-clamped beam has been biased, b) by action of the applied voltage and the electrostatic force the beam collapses on the oxide layer allowing the electrons to tunnel into the floating gate, and c) by reducing the applied voltage an imbalance between the restoring force with the electrostatic and short-range forces occurs allowing the suspended gate to return to its initial flat position 15

2.5 Schematic diagrams that represent the reading operations when a) the SGSNM cell has been erased, b) programmed and c) shows the I_{ds} - V_{SG} transfer characteristics . . . 16

3.1 Double-plate capacitor schematic diagram used to analytically obtain the pull-in and pull-out voltage equations . . . 18

3.2 Schematic diagram of the doubly-plate capacitor model modified to calculate the pull-out voltage . . . 19

3.3 Schematic diagram of the doubly-clamped beam structure implemented in CoventorWare . . . 21

3.4 Set of schematic diagrams that show the main points within the implementation of the doubly-clamped structure in CoventorWare . . . 22

3.5 Pull-in 3D-FEM analysis of a set of parameters shown in Table 3.1 . . . 23

3.6 3D-FEM analyses used to find the pull-in and pull-out trajectories for two structures Al / SiO_2 and Al / Si_3N_4, respectively . . . 24

3.7 Pull-in trajectories for the set of parameters specified in Table 3.2 that fulfils the requirements of pull-in under 10 V . . . 25

3.8 Pull-in and pull-out sequence of the doubly-clamped Al beam (30 nm) coupled to SiO_2 tunnel oxide layer (20 nm) . . . 26

3.9 Hysteresis cycles. (a) Displacement-Voltage, showing the pull-in/pull-out voltages and the stiction endurance; (b) Capacitance-Voltage characteristics . . . 27

3.10 Schematic diagram that describe the programming, reading and erasing operations with its associated hysteresis cycles . . . 28

3.11 Capacitance-Voltage Hysteresis Curves Implementation . . . 28

3.12 Region 1 approximated by using a third degree algebraic polynomial. The dot-line represents the discreet voltage dependent function and the continuous line represents the algebraic approximation . . . 29

3.13 Region 2 defined by the pulled-in beam and the increased contact area. In here, the dot-line represent the discreet C-V curve and the continuous line the approximation generated by the third degree polynomial . . . 29

3.14 Region 3 is defined as the stiction window starting from the maximum voltage applied until the suspended gate is detached from the tunnel oxide layer. The dot line represents the discreet erratic detachment behaviour and the continuous line the approximation curve generated by the algebraic model . . . 30

3.15 Programming and erasing sequences performed by the suspended control gate through the bias voltage . . . 32

3.16 Transient analysis of the doubly-clamped beam via numerical analysis . . . 32

3.17 Characteristic curve of a transient analysis performed to calculate the mechanical switching time (pull-in effect) of the doubly-clamped beam . . . 33

3.18 3D FEM scaling-down analysis by shrinking the suspended gate parameters according to a scaling factor described in Table 3.5 . . . 34

4.1 Programming and erasing sequence according to the SG and tunnelling processes. a) SG without bias, b) a negative voltage is applied on the SG and the electrons tunnelling is allowed because of the triangular shape of the tunnel barrier, c) by reducing the applied voltage the SG returns as well as the tunnel barrier to its initial position, d) considering the SGSNM cell programmed, a positive voltage is applied on the SG (d) and the electrons are removed from the FG (e). Finally, by reducing the applied voltage both the SG and the tunnel barrier return to their initial position (f) . . . 36
4.2 Small potential steps used for the transfer matrix calculations of a single barrier 38
4.3 Tunnel barrier equivalent circuit. Where C_{tb} is the tunnelling capacitance and R_{tb} is the tunnel variable resistance . . . 40
4.4 Current density-voltage set of curves for several SiO_2 thicknesses. In here the Fowler-Nordheim tunnelling equation and current are specified for a 5 nm SiO_2 layer . . . 41
4.5 Full current density-voltage characteristic curve for a SiO_2 layer at 7 nm thickness 42
4.6 Current density-voltage characteristic curve for a sweeping pull-in voltage (6.7 to 9 V) . . . 43
4.7 Current density-voltage pull-in voltage approximation. Circles represent the pull-in effect and the squares represent the approximation function . . . 43
4.8 Current density-voltage curves for the erasing (a) pull-in and (b) stiction - pull-out sequence . . . 44
4.9 Current density-voltage curves for the programming (a) pull-in (b) stiction - pull-out sequence . . . 45

5.1 Schematic diagrams that represent the suspended control gate and the tunnel oxide layer, a) shows the suspended gate pull-in, stiction and pull-out effects represented for three capacitors and b) shows the tunnel oxide layer as a variable resistor and a tunnel capacitor both in parallel . . . 48
5.2 Flow Diagram representation for the C - V and J - V hysteresis cycle . . . 49
5.3 Schematic diagram that represents the algorithm implemented as an analog model by using Verilog-AMS, in which T_{BP} is the tunnel barrier potential and M_N describes the memory node . . . 51
5.4 Schematic diagram for the suspended gate silicon nanodots memory . . . 51
5.5 SGSNM schematic models for a) programming, erasing and b) reading processes 52
5.6 Programming, reading and erasing voltage signals for the bi-stable piece-wise linear control source and the pulse signal used in the readout element activation 53
5.7 Set of signals obtained for each one of the SGSNM cell nodes a) suspended control gate, b) readout pulse signal, c) memory node signals in which the non-volatility of the SGSNM is displayed and d) once the readout element is activated, it displays a current pulse when the memory node is positively charged and no signal when the memory node is negatively charged . . . 54
5.8 Charging/discharging time as function of the tunnel oxide layer thickness . . . 55
5.9 Programming and erasing time obtained as a combination of the mechanical switching time and the charging/discharging time as function of the tunnel oxide thickness . . . 56
5.10 Energy consumption analysis in which the mechanical, electrostatic and kinetic as the principal energies . . . 57
5.11 Schematic representation used to calculate the Hamaker constant by using the Lifshitz model . . . 58

5.12 D - V characteristic curves for the pull-out analysis in which the short range forces are neglected in one analysis and included for the other . . . 59
5.13 Comparison between the emerging non-volatile memory devices with the SGSNM device (Courtesy of Prof. Mizuta) 60

6.1 Cross-section of the transistor foundations provided by our counterparts at Hitachi CRL, which shows schematically the set of layers on the transistor 62
6.2 Schematic set of diagrams that describe the fabrication process of the suspended control gate over the dummy samples. a) e-beam deposition of the aluminium layer, b) photolithography process by using a double-clamped features mask, c) etching process for the aluminium layer and finally, d) isotropic dry etching process that suspend the double-clamped structure 63
6.3 Schematic diagram of the apparatus used to deposit silicon nanodots by using SiH_4 plasma with pulsed H_2 64
6.4 Distribution of the silicon nanodots size fabricated by pulsed-H_2 supply into SiH_4 plasma [2] 64
6.5 Set of SEM images that shows a silicon nanodot monolayer a) coated by HMDS with a density of 7.33×10^{11} dots/cm^2 and b) a zoom of the top view of the film in which a lattice-like structure is observed [3] 65
6.6 Set of schematic diagrams that a) display the thicknesses within the silicon nanodot, b) roughness of the surface after oxidising the silicon nanodot layer, c) after polishing the rough surface produced by the silicon nanodots oxidising process and d) atomic layer deposition process to deposit SiO_2 as tunnel oxide layer . . 66
6.7 Set of optical images obtained after the sample, with MOSFET features, have been coated with aluminium. The images obtained are displayed with a zoom of a) 100×, b) 500×, c) 1000× and d) 10000× 68
6.8 Set of optical images obtained after the photolithography process is performed. In here the aluminium beam features looks rather sharp and well defined as shown at different zooms such as a) 5×, b) 10×, c) 20× and d) 100× 70
6.9 Set of optical images in which the double-clamped features are analysed after the aluminium wet-etching process took place. These images shows the aluminium beams sharpness at different zooms such as a) 5×, b) 10×, c) 20× and d) 100× . 71
6.10 Set of SEM images tilted 54° shows the double-clamped beam structures after the aluminium wet etching process took place as shown in image a), image b) shows that by zooming on the two beams inner area, the side walls and the beam surface are not over-etched, in image c), the beam surface looks uniform and smooth and in image d) it shows that the sidewalls are sharp and the substrate do not show apparent damage . 72
6.11 Set of SEM images obtained after the anisotropic dry etching process was performed, a) shows the set of images as an entire beam array, b) by zooming on the beams, it is possible to see the bi-layer beam in which the aluminium and the poly-Si layer are visible, c) and d) show a zoom over the central beam area, which shows a sidewall not as smooth as in previous image 73
6.12 Set of SEM images that shows the double-clamped beams successfully suspended when two dry-etching plasma processes are used. a) Shows the air-gap at the corner of the set of beams and b) shows the under-cut at the centre of the double-clamped beam structure 74

6.13 Set of SEM images that represents the one-step-plasma process that isotropically removes the sacrificial layer of the double-clamped beam features without affecting the other layers. Image a) shows two double-clamped beam structures before the cross-section process, b) SEM cross-section of the beams showing successfully the undercut in both structures, c) and d) images display individually a zoom over each beam 75
6.14 Double-clamped suspended beam structure under test in which the beam electrode is visible over the substrate 77
6.15 Pull-in effect obtained as a result of the electrical C - V measurements of a single double-clamped beam of 59.3 µm in length at 4.8 µm width, with a beam thickness of 500 nm and an air-gap of 320 nm. By using the information obtained from the C - V characterisation, the pull-in effect and the displacement - voltage characteristic curve as displayed within the main figure in the upper-left part 78
6.16 Pull-in effect obtained as a result of the electrical C - V measurements of a set of double-clamped beams 79
6.17 Pull-in analysis obtained by a numerical simulation once the structural parameters of the double-clamped beam structure have been updated . . 80
6.18 2D FEM analysis in which the readout element is included for more accurate pull-in voltage estimate 81

A.1 Interconnection between the C - V meter with the test probe system through the interconnection panel 87
A.2 Stage probe system and the array used to perform the C - V measurements . . . 88

B.1 Suspended Gate Silicon Nanodot Memory cell fabrication process. a) Memory device foundation, b)readout element, silicon nanodot floating gate, sacrificial layer and metal deposition layers; c) suspended gate patterned and wet etched (mask 1); d) by using anisotropic dry etching process, the sacrificial layer is anisotropically removed; e) Source-Drain contact holes are created using wet etching (mask 2); f) Contact pad deposition and lift-off (mask 3); g) finally, by using isotropic dry etching the sacrificial layer is removed and the beam is suspended, h) top, side and lateral views of the SGNM cell 89
B.2 Schematic diagram of the SGSNM cell that shows the three photomask such as pads, doubly-clamped beam features and the holes for the highly-doped regions . 90
B.3 Set of parameters that shows the length and width of the doubly-clamped beam features 91
B.4 Mask of the doubly-clamped beam features 91

List of Tables

2.1 Programming, reading and erasing characteristics of the SGSNM cell . . . 16

3.1 Set of materials and parameters proposed as suspended control gate . . . 22
3.2 Parameters used to optimise the structure 25
3.3 Coefficients obtained by approximating the C - V hysteresis curve with third order algorithms 31
3.4 Coefficients obtained by approximating the C - V hysteresis curve with third order algorithms 31
3.5 Scaling-down parameters implemented within de 3D FEM analysis 33

4.1 Comparison between voltage and current densities currents 42
4.2 Coefficients used to approximate the pull-in and stiction curves as function of an algebraic polynomial algorithm for the erasing process 44
4.3 Coefficients used to approximate the pull-in and stiction curves as a function of an algebraic polynomial algorithm for the programming process . . 45

5.1 Parameters used to control the SGSNM sequences, as function of the inner voltage 50
5.2 Control parameters used to define the voltage sources in the circuital analysis 52
5.3 Set of parameters implemented within the Lifshitz model to obtain the Hamaker constant 58

6.1 Spin parameters optimised to obtain 1 μm thick coating layer for a S1813 positive photoresist (Shipley) 68
6.2 Mix of gases and conditions used within the RIE 80+ plasma system to remove anisotropically the sacrificial layer (poly-Si) 71
6.3 Mix of gases and conditions implemented to isotropically remove the sacrificial layer by using 72
6.4 Mix of gases and parameters used within the RIE 80+ plasma system to remove isotropically the sacrificial layer by using a one-step plasma process 75
6.5 Comparison of the double-clamped beam structure while analysed by using a simple model, 3D FEM, *in situ* measured and an up-dated 3D FEM analysis 80

7.1 Set of parameters required to design the new photomask 85

B.1 Dimensions used in the photomasks for the length and width of the beams, respectively, as well as the pad dimensions 90

Nomenclature

C - V	Capacitance - Voltage Hysteresis Cycle
D - V	Displacement - Voltage Hysteresis Cycle
e.g.	exempli gratia
i.e.	id est
J - V	Current Density - Voltage Characteristic Curve
$^{\circ}$C	Temperature in Celsius
ϵ_0	Free Space permittivity
ϵ_{ox}	Tunnel oxide permittivity
η	Poisson's ratio
κ	Dielectric Constant
Ψ	Wave Function
µm	Micrometres
E	Electric Field
2D	Two-Dimensional
3D	Three-Dimensional
A	Area
A	Ampere
A_h	Hamaker Constant
AHDL	Analog Hardware Description Language
Al	Aluminium
BL	Bit Line
C_{tb}	Tunnelling Capacitance
CD	Current Direct
CHE	Channel Hot Electron
cm	Centimetre
CMOS	Complementary Metal Oxide Semiconductor
CMOS RAM	Complementary Metal Oxide Semiconductor RAM
Cu	Copper
d_0	Initial air-gap
DI water	Deionised Water
DRAM	Dynamic Random Access Memory
E	Young's modulus

E	Erase
E_F	Fermi Level
EEPROM	Electric Erasing/Programming Read Only Memory
EPROM	Erasing/Programming Read Only Memory
F	Farads
$F_{Casimir}$	Casimir Force
F_{vdW}	van der Waals Force
Fe-RAM	Ferromagnetic Random Access Memory
FEM	Finite Element Method
FG	Floating Gate
FN	Fowler-Nordhein Tunnelling Process
Fortran77	Fortran Language version 77
FPM RAM	Fast Page Mode Random Access Memory
h	Plank Constant
k	Spring constant
K	Temperature in Kelvin
k^i	Wave number
L_{SG}	Suspended Gate Length
MEMS	Micro-Electromechanical Systems
min	Minutes
MOS	Metal Oxide Semiconductor
MOSFET	MOS Field Effect Transistor
MRAM	Magnetic Random Access Memory
NEMS	Nano-Electromechanical Systems
nm	Nanometres
NV-RAM	Non-Volatile Random Access Memory
P	Programming
P/E	Programming and Erasing operations
PC-RAM	Phase Change Random Access Memory
Poly-Si	Poly Silicon
R	Read
R_{tb}	Tunnel Resistor
RAM	Random Access Memory
RDRAM	Rambus Dynamic Random Access Memory
ROM	Read Only Memory
rpm	Revolutions Per Minute
SDRAM	Synchronous Dynamic Random Access Memory
sec	Seconds
SG	Suspended Control Gate
SGSNM	Suspended Gate Silicon Nanodot Memory
Si_3N_4	Silicon Nitride

SiNDs	Silicon Nanodots
SiO_2	Silicon Dioxide
SRAM	Static Random Access Memory
T^i	Matrix
t_{ox}	Thickness of the Tunnel oxide layer
t_{SG}	Thickness of the Suspended Control Gate
TB	Tunnel Barrier
UHF	Ultra High Frequency
V	Voltage
V(z)	Potential Energy
V_d	Drain Voltage
V_g	Gate Voltage
$V_{Pull-in}$	Pull-in Voltage
$V_{Pull-out}$	Pull-out Voltage
V_{SD}	Source - Drain Voltage
VHF	Very High Frequency
VRAM	Video Random Access Memory
WL	Word Line

Declaration of Authorship

I, Mario Alberto García Ramírez, declare that the thesis entitled:

“Suspended Gate Silicon Nanodot Memory”

and the work presented in it are my own. I confirm that:

- this work has done wholly or mainly while in candidature for a research degree at this University;
- 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;
- where I have consulted the published work of others, this is always clearly attributed;
- 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;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly that I have contributed by myself and for the integrality of projects, I have made clear what was done by others;
- parts of this work have been published in research journals. A list of publications is provided with this manuscript.

Signed:

Date:

List of Publications

Journals

1. **Hybrid Analysis, Fabrication and Characterisation of the Suspended Gate Silicon Nanodot Memory (SGSNM)**
   García-Ramírez, Mario A., Tsuchiya, Y. and Mizuta, H., *Journal of Applied Physics*, *Manuscript in preparation*
2. **Hybrid Numerical Analysis of a High-Speed Non-Volatile Suspended Gate Silicon Nanodot Memory (SGSNM)**,
   García-Ramírez, Mario A., Tsuchiya, Y. and Mizuta, H., *Journal of Computational Electronics*, 10, 1, pp. 248 - 257 (2011), DOI:10.1007/s 10825-011-0361-z
3. **Hybrid Circuit Analysis of a Suspended Gate Silicon Nanodot Memory (SGSNM) cell**,
   García-Ramírez, Mario A., Tsuchiya, Y. and Mizuta, H., *Microelectronic Engineering*, 87 . pp. 1284-1286 (2009), DOI:10.1016/j.mee.2009.10.019
4. **Multi-scale Simulation of Hybrid Silicon Nanoelectromechanical (NEM) Information Systems**,
   Mizuta, H., **García-Ramírez, Mario A.**, Tsuchiya, Y., Nagami, T., Sawai, S., Oda, S. and Okamoto, M., *Journal of Automation, Mobile Robotics & Intelligent Systems, 3, pp. 58-61 (2009)*

List of Conferences

1. **Scaled Silicon Nanoelectromechanical (NEM) Hybrid Systems**,

   Mizuta, H., **García Ramírez, Mario A.**, Arab Hassani, F., Ghiass, M. A., Kalhor, N., Moktadir, Z., Tsuchiya, Y., Sawai, S., Ogi, J. and Oda, S., In: International Conference on Solid-State and Integrated Circuit Technology (Invited Talk), Nov. 1 - 4, Shanghai, China (2010)

2. **Suspended Gate Silicon Nanodot Memory (SGSNM) cell towards Non-Volatile RAM Memories**,

   García Ramírez, Mario A., Tsuchiya, Y. and Mizuta, H., In: International Symposium on Atom-scale Silicon Hybrid Nanotechnologies for "More than Moore" & "Beyond CMOS" era, held in the University of Southampton, March 1 - 2, Southampton, UK (2010)

3. **Hybrid Circuit Analysis of a Suspended Gate Silicon Nanodot Memory (SGSNM) cell**,

   García Ramírez, Mario A., Tsuchiya, Y. and Mizuta, H., In: 35th International Conference on Micro and Nano Engineering (MNE2009), September 28th, Ghent, Belgium (2009)

4. **Multi-scale simulation of hybrid silicon nano-electromechanical (NEM) information devices**,

   Mizuta, H., **García Ramírez, M.**, Arab Hassani, F., Ghiass, Mario A., Tsuchiya, Y., Nagami, T., Pruvost, B., Ogi, J., Sawai, S., Oda, S. and Okamoto, M. In: 8th International Conference on Global Research and Education Inter-Academia (ICERI2009), November 16 - 18, Madrid, Spain (2009)

5. **NEM - MOS Co-integration for Fast & Nonvolatile Memory Applications (Invited Talk)**,

   Mizuta, H., **García Ramírez, Mario A.**, Tsuchiya, Y., Nagami, T., Uchida, K. and Oda, S., In: IEEE NANO Satellite Workshop on Emerging Nonvolatile Memories, July 26 - 30, Genoa, Italy (2009)

6. **Co integration of Silicon Nanodevices and NEMS for Advanced Information Processing**,

   Mizuta, H., Nagami, T., Ogi, J., Pruvost, B., **García Ramírez, Mario A.**, Yoshimura, H., Tsuchiya, Y. and Oda, S., In: 9th International Conference on Solid-State and Integrated-Circuit Technology (ICSICT2008), October 20 - 23, Beijing, China (2008)

7. **Suspended Gate Silicon Nanodots memory**,

   García-Ramírez, Mario A., Yoshimura, H., Tsuchiya, Y. and Mizuta, H., In: ESSDERC/CIRC Fringe (ESS-Fringe), September 15 - 19, Edinburgh, Scotland, UK (2008)

Acknowledgements

These few lines can not give all my gratitude to many, many, $\cdots$ many people who made possible this PhD. I want to express my gratitude to my supervisor Professor Hiroshi Mizuta for his guidance, support, and endless knowledge. A special acknowledgement to the Mexican Council of Science and Technology (CONACyT) for funding my Master & PhD and to Hitachi CRL for provided me samples to work with. Special thanks to Dr. Arturo Sarmiento for his "peculiar" way to show us the tree that was hiding the forrest, Dr. Manuel Corona for his advices, help and support and Dr. Ernesto Aguilar for his support when needed, cheers Chief.

I want to thank all the people for the countless hours of hard work in the cleanroom, pub, corridor, office, desk, courts, etc. Special thanks to Dr. Harold Chong, Dr. Yoshishige Tsuchiya, Dr. Stuart Pearce, Dra. Barbara Cortese, Dr. Owain, Dr. Sun Kai, Dr. Kian, Dr. Ashwin, Katie, Denny, Peter Ayliffe, Richard Bayly, thanks for all your time, support, training, comments, suggestions and chats, you guys really improve my research as well as made my life and work easier and better.

También quiero expresar mi agradecimiento a todas las personas que directa e indirectamente coadyuvaron para la realización de este proyecto. En forma muy especial quiero agradecer a Yuyi Gwe (Grace) por su paciencia durante mis tiempos más obscuros, los viajes, por incrementar mi visión acerca de la vida, y por $n + 1$ cosas, Sank you very much!!!!!. Al grupo de estudiantes, trabajadores, amigos en Southampton, Europa y Asia, en forma especial a Omar, Juan José, Nadia, D.D.S (a.k.a Lourdes), David (Potter), Cesar, Ivo, Allan, Juan de Dios, Mario Enrique, Lucila (muchas gracias boluuua), Fabiola, Renato, Luis (Hipster), Victor Castillo, David, Livia, Faezeh, Marek, Adel, Rares (Rari), Hakim, Asa, Mr & Mrs Kalita, Mr & Mrs Sámson, Khaled, Zakaria, Akhtar, Yudong, Hwanjit, Suhana, Asa, Tess de Leon, $\cdots$ for $\cdots$ everything.

A rather big part of my life has been shared with loads of people since the dark times at INAOE and earlier till now and I really hope it can last for a long time. I want to thank for the support to all the people who was around, specially to Luis Abraham, Carlos Muñiz, Carmen Paleta, Rosa (Ajo), Enrique (Zopilas), Jacobo, Miguel, Isabel, Xicotencatl, Jonathan Tavares, Victor Rodolfo, $\cdots$, Thanks.

Dedication

To my Family

Elsa, Mario, Ana Iris, Juan Luis, Doña Anita, Don Luis ✝, Doña Lupe ✝

Alicia, Clara, Lourdes, Gema, Amalia, Ana, Salvador, Gustavo, $\cdots$

Alondra Itzel, Ana Fernanda

Y

a un Hippo Amarillo

Chapter 1

Memories Background

1.1 Introduction

Electronic devices have had a major impact on almost every aspect of human life since they were first invented a few decades ago. Nowadays, global position systems (GPS), digital cameras, computers, MP3 players, portable video games or the capability to make transatlantic video calls are widespread and commonly used by most people. Invariably all sorts of apparatuses need to store different kind of information from documents, images or music to whole internet. There are three ways to store large amounts of information: magnetic, optical or in semiconductor devices [4]. Regardless of the storage capabilities of optical and the magnetic devices, the semiconductor based memories are more suitable for implementation in a wide variety of devices due to its characteristics such as low power consumption, low cost and reliability. Semiconductor memories are suitable for numerous devices such as watches, pen drives, mobile phones to quite complex systems such as satellites or planes. This is the main reason why the memory market has grown exponentially in the last decades and it is this that has driven my research. The aim of this research is to design, analyse and fabricate a new non-volatile and high speed memory as a promising candidate for the non-volatile RAM device. The memory device that we are proposing is the Suspended Gate Silicon Nanodot Memory (SGSNM) which co-integrates nano-electromechanical systems (NEMS) with MOS technology. The proposed structure features a MOSFET as readout element, silicon nanodots (SiNDs) as floating gate (FG) and a movable suspended gate (SG) which is isolated from the FG by an air gap and by a thin oxide layer. The advantages that the SGSNM has to offer over the other non-volatile memories include high programming and erasing speed operations, endurance, low power consumption, low size and virtually a negligible leakage current from the floating gate towards the readout element due to a thick oxide layer and to the suspended control gate because the thin tunnel oxide layer and air-gap. A serious non-volatility is expected due to the presence of the air gap except for programming and erasing operations. In order to achieve the goals above described a

review of non-volatile memories as well as the programming and erasing operations are provided in Chapter 1. The new non-volatile high speed memory structure is introduced in Chapter 2. The electro-mechanical behaviour of the suspended control gate is modelled and analysed numerically in Chapter 3. Tunnelling process is analysed in Chapter 4. The implementation as behavioural models of the above analyses, within a SGSNM circuital simulation, are described in Chapter 5. The fabrication process of the suspended control gate, the capacitance - voltage characterisation and the footprints of the entire fabrication of the SGSNM cell are discussed in Chapter 6. Finally, in Chapter 7 the summary and future work of the SGSNM are presented. In order to complement the understanding of the internal processes, two appendices are attached to the document.

1.2 Semiconductor memory devices

Semiconductor memories are divided into two main groups: volatiles and non-volatiles. Volatile memory devices require, by definition, to be always connected to an energy source in order to keep the information (data) stored. Otherwise, the information stored will be lost when the device is unplugged from the energy supply. Most random access memories (RAM) are ranged within these kinds of devices. There are several types of RAM devices such as dynamic RAM (DRAM), static RAM (SRAM), fast page mode RAM (FPRAM), synchronous dynamic RAM (SDRAM), double-data rate RAM (DDR RAM), video RAM (VRAM) and CMOS RAM, among others [4, 5, 6]. Each one of these memories has its particular application, *e.g.* DDR RAM is widely used within the PC and lap-top market. This memory is a variation of the SDRAM which uses the internal clock (rising and falling edges) to send twice the amount of data by using the same amount of time rather than the rising edge as for SDRAM memory. In contrast, read only memories (ROM) or non-volatile memories have the capacity to maintain the information stored all the time even when the power supply has been disconnected or switched off. This kind of memory devices are encompassed within the following groups:

- Mask ROM
- Programmable ROM (PROM)
- Erasable/programmable ROM (EPROM)
- Electric erase/programmable ROM (EEPROM or E^2PROM)
- Flash memory

Within this group of non-volatile memories, there are those that can be reprogrammed and erased. These devices are electrically programmed such as EPROM and erased by using UV light. There are also those that can be programmed and erased electrically

such as E^2PROM and flash memory [7, 8, 9]. Within those non-volatile memory devices, EPROM takes longer to be erased in the order of minutes (24 minutes) and special hardware is required [10]. In contrast, for the E^2PROM device, the erasing process is relatively faster than that of its predecessor. This process takes as long as $\approx$ 10 msec to erase and program a single byte. However, both operations are performed per single byte blocks and the erasing time is long [4, 10]. Despite the fact that it uses a similar programming and erasing process to E^2PROM, flash memory is rather faster in both programming and erasing operations. This is why both EPROM and E^2PROM are considered the ancestors of flash memory [11]. In this Chapter an introduction to non-volatile semiconductor flash memory is presented that includes the programming and erasing operations and the tunnelling mechanism such as Fowler-Nordheim and Channel Hot Electron. Finally, an overview of non-volatile memory devices is presented.

1.2.1 Flash Memory

Flash memory has been playing a key role in the semiconductor industry and in our daily lives since its invention three decades ago [12]. Nowadays, it is quite common to see this kind of non-volatile semiconductor memory co-integrated with a wider range of devices and apparatuses. Flash memory features a stacked MOSFET as the readout element, a floating gate layer isolated from the substrate by a thermally grown oxide layer and buried within an inter-poly-Si dielectric which isolates the floating gate from the control gate. A schematic diagram of non-volatile memory structure is depicted in Fig. 1.1.

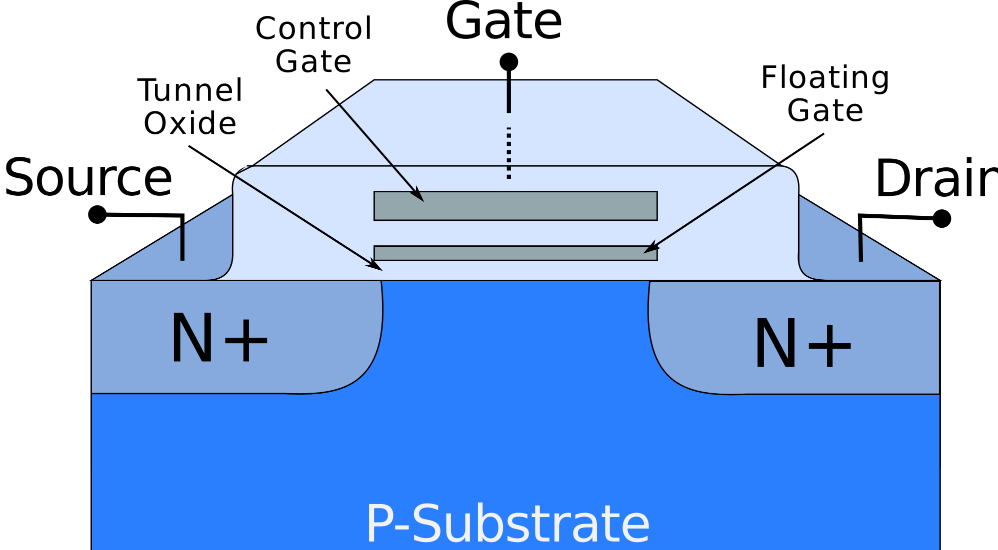


FIGURE 1.1: Schematic diagram of the flash memory structure

Flash memory operations are controlled via the applied voltage on the control gate electrode, substrate (back-contact) and by activating (bias) the readout element [11, 13].

To program the memory device, a positively biased voltage is applied on the control gate electrode, the back contact is grounded and the MOSFET is biased. Once the channel is created, the electrons are injected through the tunnel oxide layer into the floating gate due to the channel hot electron (CHE) tunnelling mechanism [14]. The memory node remains negatively charged (programmed) until the electrons are removed from the floating gate. To erase the memory device, the control gate electrode and the back contact are biased. In this process, the tunnelling mechanism is implemented in a different way for NAND and NOR cells. When the control gate is negatively biased and the source of the MOSFET is positively biased, the electrons tunnel from the floating gate into the source through the tunnel oxide layer by using the Fowler-Nordheim tunnelling process (NOR cell). In contrast, when the back contact is positively biased and the control gate is grounded, the electrons tunnel from the floating gate into the substrate through the oxide layer (NAND cell) [14]. The schematic diagrams for the Fowler-Nordheim and Channel Hot Electron tunnelling mechanisms are depicted in Fig. 1.2.

1.2.2 Channel Hot Electron tunnelling process

The Channel Hot Electron tunnelling process is typically used as the programming process in NOR-type flash memory [15, 16]. In here, the control gate electrode and the drain electrode are positively biased whilst both the source and back contact electrodes are grounded as depicted in Fig. 1.3a. Once the readout element is biased and the channel is formed, the electrons gain energy under the lateral electric field ($\vec{E}_{lat}$). By increasing $\vec{E}_{lat}$ to more than 100 kV/cm, the electrons are no longer in equilibrium but heated by the high lateral $\vec{E}$. By using the extra energy provided by the lateral field, a small fraction of the electrons defined by Eq. (1.1) can gain energy large enough to surmount the barrier between the oxide and silicon conduction band edges as depicted in Fig. 1.2 [11, 15].

$$P = \frac{1}{2}\left[1 - \sqrt{\frac{\Phi_b}{\Phi_b + \Delta\Phi}}\right] = \frac{\Delta\Phi}{4\Phi_b} \tag{1.1}$$

Where P is the probability of an electron possessing energy $\Phi = \Phi_b + \Delta\Phi$ to surmount the barrier, Φ_b is defined as the Si/SiO_2 potential barrier (V) and $\Delta\Phi$ as the increment.

To overcome the potential barrier, electrons must meet the following conditions to surpass the barrier:

- The kinetic energy of the electrons need to be higher than the potential barrier
- The electrons must be directed towards the barrier

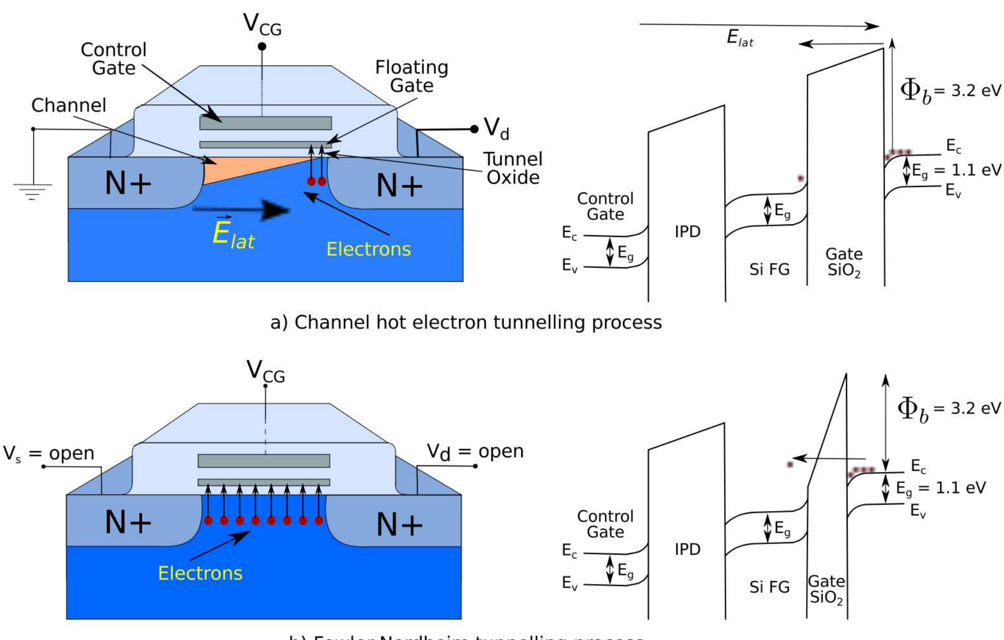


FIGURE 1.2: Tunnelling mechanisms used to program the Flash memory structures (NAND and NOR). a) Channel hot electron process, when the control gate and the MOSFET are biased (beyond saturation), within the depletion region, the channel is formed and since the bias on the MOSFET is beyond saturation, it creates a pinch-off region. This region gives enough lateral energy to the electrons to tunnel into the floating gate. b) Fowler-Nordheim tunnelling process, in the band diagram (right side) shows that by applying a large positively biased voltage the energy band diagram of the tunnel oxide layer bends until a triangular shape is obtained. Once triangular, electrons are allowed to tunnel through this layer into the floating gate

1.2.3 Fowler-Nordheim tunnelling process

The Fowler-Nordheim tunnelling mechanism allow carriers (electrons) to tunnel through a barrier in the presence of a high electric field [14, 17, 18]. This mechanism is used within the programming and erasing operations within the NAND-type flash memory. In order to program this device, a positively biased voltage is applied on the control gate electrode. When applied, the energy band diagram of the tunnel oxide layer bends from a square shape to a triangular one [17, 19]. Once it becomes triangular, the electrons tunnel through the thinner section and are trapped within the floating gate layer as depicted in Fig. 1.2b. This tunnelling mechanism presents a number of advantages such as:

- It does not require a source-to-drain bias, therefore low power consumption and short channel length are obtained
- The current involved within the process is rather low

Programming and erasing operations of the NOR and NAND-type flash memories and quantum-mechanical tunnelling processes associated to each type are depicted in Fig.

1.3.

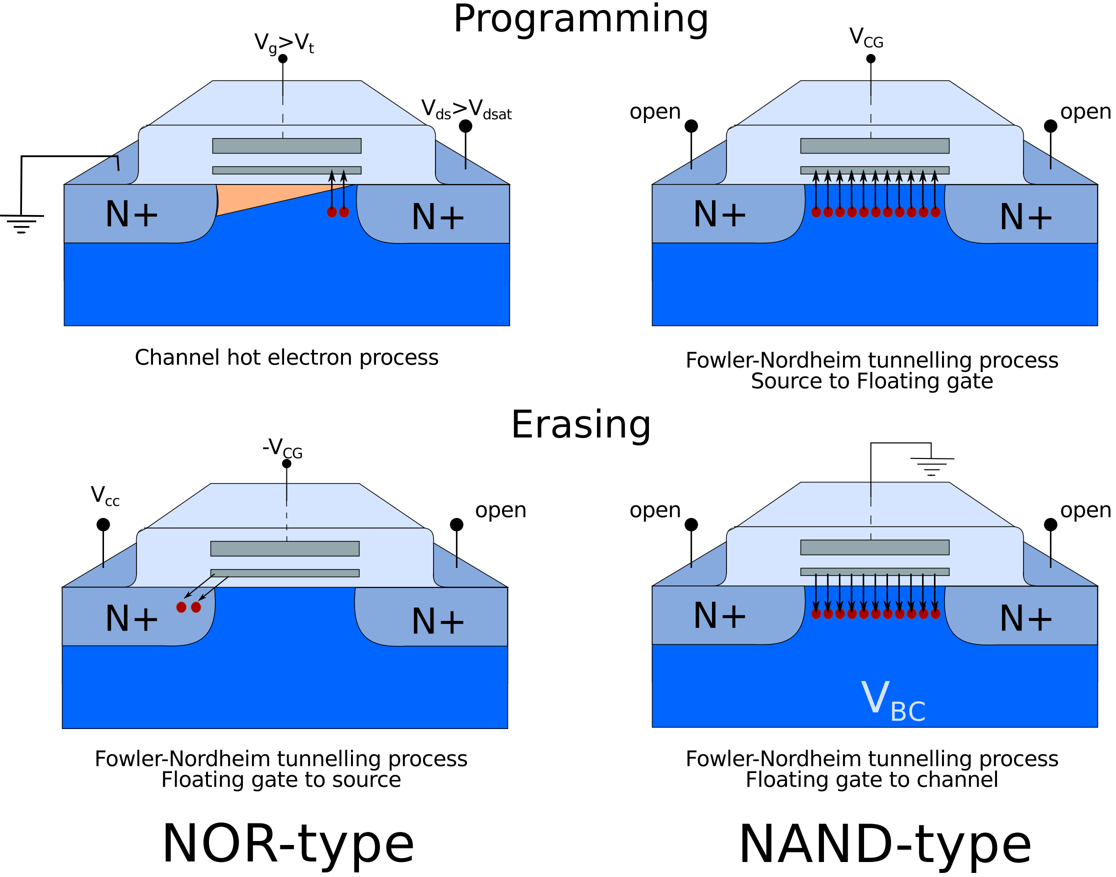


FIGURE 1.3: Schematic diagrams of the programming and erasing operations through both tunnelling mechanisms used in NOR and NAND-type flash memories

The continuous use of flash memory via the programming and erasing operations degrades the tunnel oxide layer. This degradation induces the defect creation within the tunnel oxide layer that generates leakage paths towards the substrate which results in the reduction of the memory lifetime (Fig. 1.4) [20].

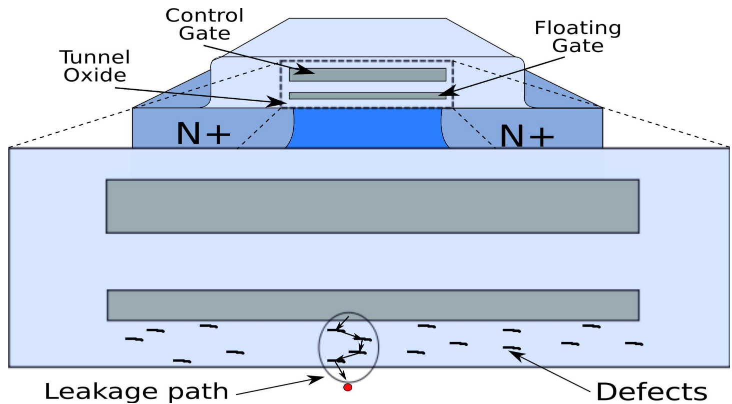


FIGURE 1.4: Defects generated within the tunnel oxide layer produced by the Fowler-Nordheim or Channel Hot Electron tunnelling processes

Considering that flash memory is following the scaling-down trend exposed in Moore's law, this non-volatile memory is facing a serious integration problem due to scalability issues mainly because the tunnel oxide layer is reaching the thickness limit described in the latest International Technology Roadmap for Semiconductors (ITRS2010) [1, 21]. When this limit is reached, several key features of flash memory such as speed, endurance and lifetime will be severely affected by the degradation of the tunnel oxide layer due to its inherent operations. There have been several proposals to overcome such drawbacks. These include the introduction of high-κ materials as the tunnel oxide layer [22, 23, 24], and the implementation of silicon nanodots as the floating gate layer instead of the poly-Si layer [3, 25, 26, 27]. In this case, the silicon nanodot layer is proposed as an alternative to reduce the leakage current issue [23, 28]. The schematic diagram of such memory device featuring the silicon nanodot as floating gate layer is depicted in Fig. 1.5.

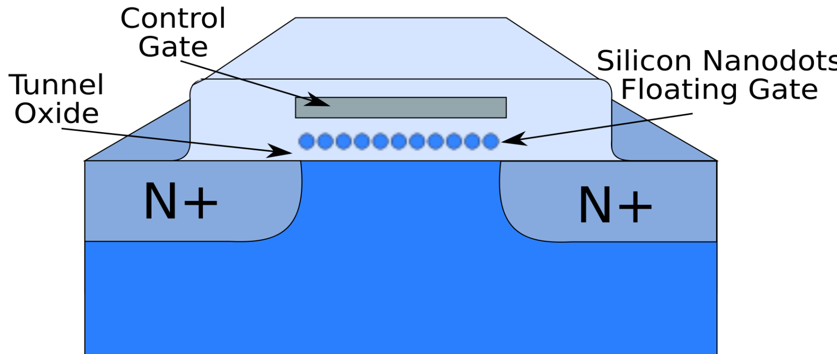


FIGURE 1.5: Schematic diagram of the silicon nanodots non-volatile memory structure

Given that this memory device uses the same tunnelling mechanisms as its predecessors, consequently some defects are created within the oxide layer due to the continuous use of the memory device. Nevertheless, only the electrons stored within the silicon nanodot adjacent to the leakage paths might leak which retains most of the charge stored at the floating gate. By including the silicon nanodots as the floating gate, it relies in an increment of the memory life time but does not solve the leakage path generation. The leakage process in this structure is depicted in Fig. 1.6

Programming and erasing operations within flash memory are linked to the time that these processes take to perform a single cycle. In this context, another drawback of flash memory emerges. The average time that the memory device takes to perform these operations is in the range of microseconds [4]. In today's world, the time it takes to perform these processes is jeopardising the prevalence of flash memory as the non-volatile memory market leader. If flash memory wants to retain market hegemony, it has to improve its timings making them closer to those of DRAMs and SRAMs. However, in the non-volatile memory market various different approaches have been examined in order to gather the qualities that are needed for the up-coming technologies such as high-speed, non-volatility, reliability, low-power consumption and cost production, amongst others.

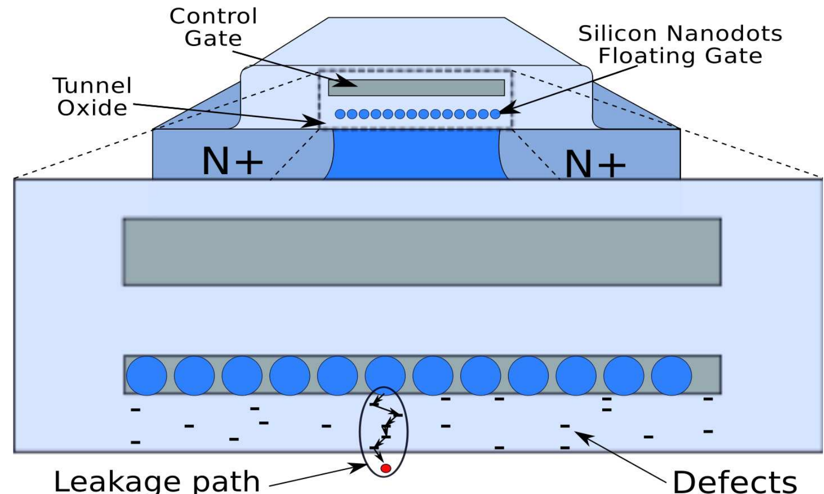


FIGURE 1.6: Defects created within the tunnel oxide layer by action of the programming and erasing operations due to the Fowler-Nordheim or Channel Hot Electron tunnelling processes

Few structures have been proposed that fulfil those requirements known as emerging memories such as Ferro-electric RAM (Fe-RAM), Magneto-resistive RAM (MRAM) and Phase-change RAM (PC-RAM) [4, 13, 29, 30]. The introduction of the emerging memory devices in the competition for the title of non-volatile RAM devices inherently carries some drawbacks. One of these requires the introduction of exotic materials within the mass production process, where these materials are not used within conventional Si-based technology. Issues such as this leave the door open for new competitors to enter the non-volatile memory market.

Figure 1.7 describes the ideal position of the non-volatile memory device at fast programming & erasing operations, large retention time, lower stand-by power and a major functionality/speed. In here, Flash memory shows that has a strong non-volatility, large retention time but also shows a rather low writing and erasing characteristics as well as low functionality. It also shows the other non-volatile memories and how are they trying to close the path between the non-volatile capabilities towards the non-volatile RAM. On the other hand, random memories such as SRAM or DRAM shows a large random access, high writing and erasing time operations and poor retention time. From both fronts are trying to get every time closer to the ideal high speed, long retention time, random access and fast non-volatile RAM memory device.

Given that a new window of opportunities have been opened in the non-volatile memory market in which hybrid devices are showing rather interesting results in several fields such as sensors and non-volatile memories. The co-integration of micro-/nano-electromechanical systems (MEMS/NEMS) with the conventional metal-oxide-semiconductor (MOS) technology enables the increase of the functionalities of the devices/systems and meets the challenging specifications in advanced applications. To date, it is possible to see in the literature some efforts to introduce this hybrid technology in high-precision

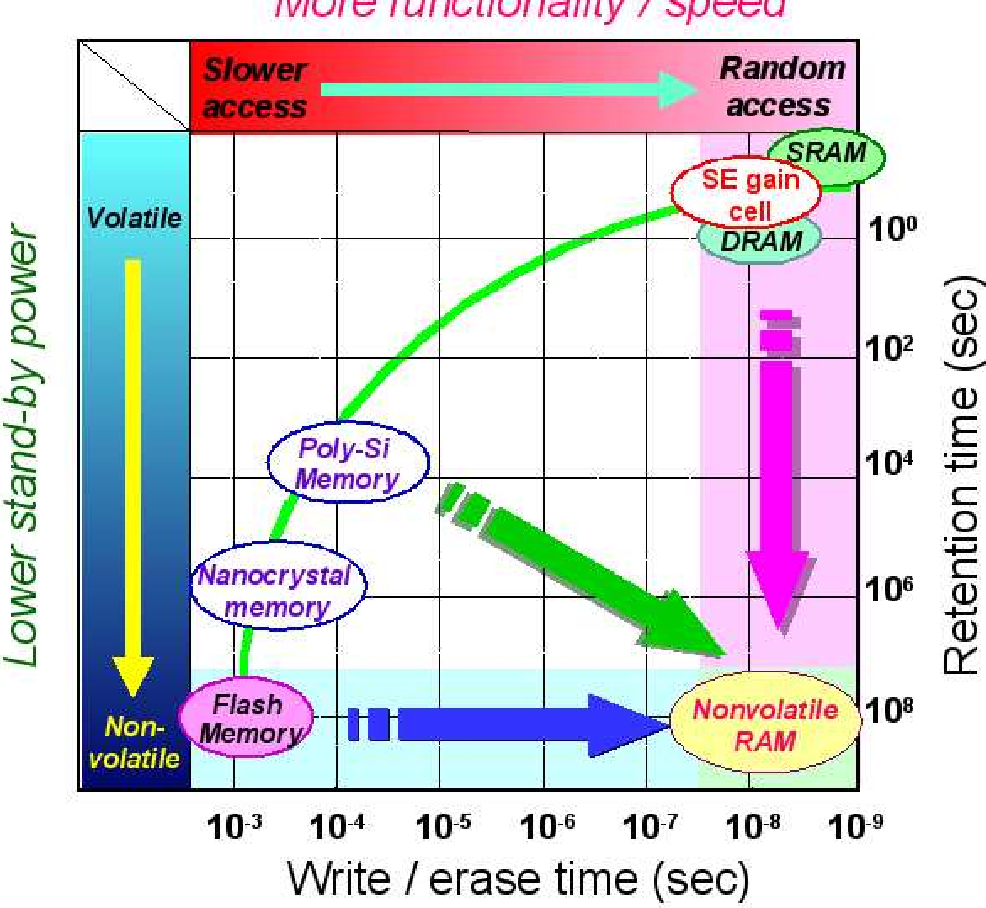


FIGURE 1.7: Volatile and non-volatile memory characteristics (Courtesy of Prof. Mizuta)

sensitive sensors [31], non-volatile memories [32], among others. Figure 1.8 describes the miniaturisation trend of several technologies in which some of them are superimposed CMOS-NEMS producing hybrid devices [33].

By co-integrating the two technologies as depicted in Fig. 1.8, hybrid devices based on MOS and MEMS/NEMS technologies may help us to develop fast, sensitive, rather small, low power, broadband, non-volatile among other characteristics that were difficult to obtain by using individual technologies such as MOS or MEMS/NEMS technology. In here it is possible to identify where the flash memory and the hybrid devices (non-volatile memory) are going to meet.

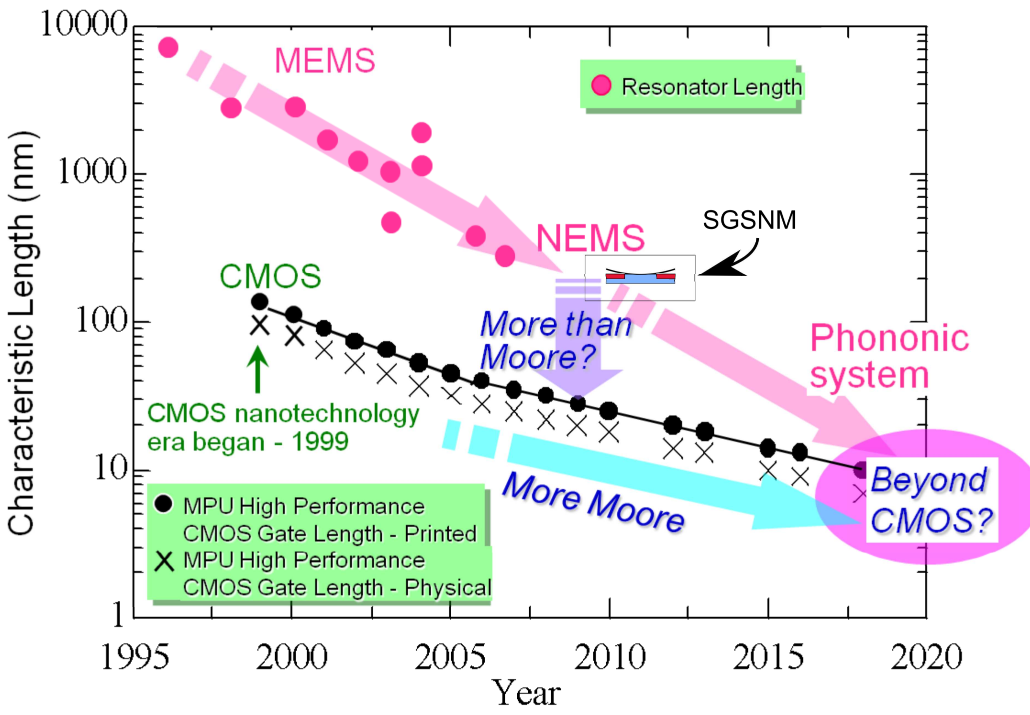


Figure 1.8: The MEMS/NEMS downscaling trend compared with the CMOS extension, in which the hybrid integration of NEMS and CMOS may offer a new approach to the More-than-Moore and Beyond-CMOS domains (Courtesy of Prof. Mizuta)

Chapter 2

Suspended Gate Silicon Nanodots Memory

In this Chapter the new non-volatile memory device is introduced. The memory structure is a hybrid device that co-integrates two different technologies nano-electromechanical systems (NEMS) with metal-oxide-semiconductor (MOS) technology, in pursuit of high-speed, negligible leakage current and large endurance cycle (>10^6) as well as low power consumption and fast programming, reading and erasing operations. This hybrid non-volatile memory device is called the Suspended Gate Silicon Nanodot Memory (SGSNM). A schematic diagram of the memory structure is depicted in Fig. 2.1.

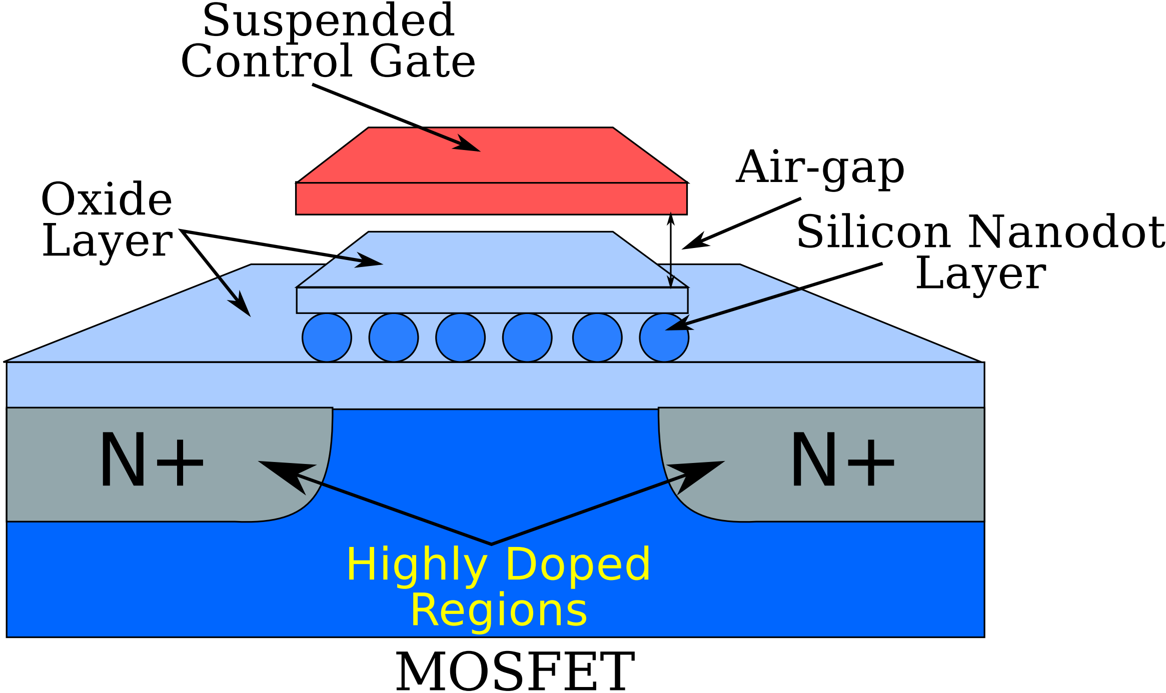


FIGURE 2.1: Suspended Gate Silicon Nanodots Memory (SGSNM) structure

The SGSNM device features a MOSFET as a readout element, a silicon nanodot (SiNDs) monolayer as the floating gate and a movable suspended control gate isolated from the floating gate by an oxide layer and by an air-gap. To perform the programming, reading and erasing operations within the non-volatile memory device, several processes are involved which are encompassed in three main sections such as:

- Suspended control gate movement (pull-in/pull-out)
- Tunnelling mechanism
- Readout memory node process

The programming and erasing operations are performed through the movable suspended control gate by the pull-in and pull-out effects produced as a consequence of the bias voltage applied on the suspended gate electrode. In order to analyse those effects in a reliable way, an exhaustive set of analyses that includes analytical and numerical analyses as well as the electromechanical behaviour of the suspended control gate are performed. In the following sections, the programming, erasing and reading techniques employed within the SGSNM device are described.

2.1 Operations performed by the suspended control gate

The programming and erasing operations are performed as a combination of the mechanical movement of the suspended control gate and the tunnelling mechanism that take place when both layers are in contact. It is necessary to pointed out that the contact area produced by both layers also covers the entire silicon nanodot monolayer (floating gate) whilst in contact. The tunnelling process in this device occurs through top carrier injection rather than from the substrate (channel formation) as with other memory devices such as flash memory [5, 13]. To read the SGSNM cell, the suspended control gate remains isolated from the floating gate (flat) and a readout element (a normally 'On' MOSFET) is activated. This reads the state of the memory node and a current pulse is displayed if the device has been erased. When the memory device has been programmed the current pulse is absent. Figure 2.2 shows a schematic diagram of the memory device.

2.1.1 Programming sequence

To program the Suspended Gate Silicon Nanodot Memory a negative bias voltage is applied across the suspended control gate electrode. As the applied voltage increases, the suspended gate bends downwards linearly until it reaches an inflection point known as pull-in voltage. After the pull-in point is reached and by increasing the applied

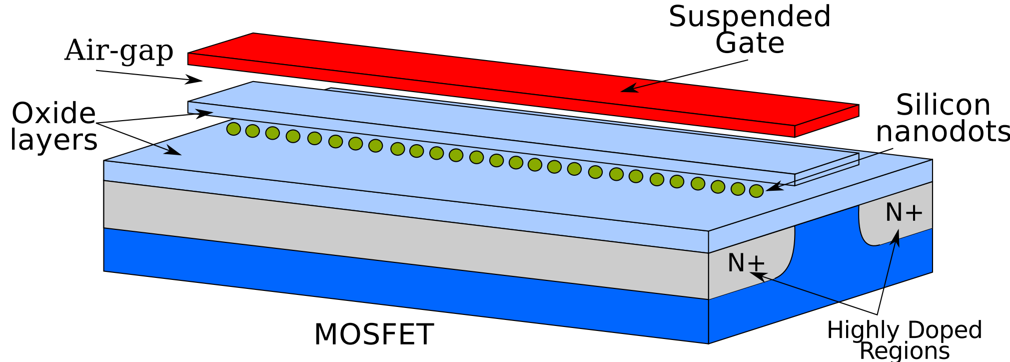


FIGURE 2.2: Schematic diagram of the SGSNM cell

voltage, the suspended gate collapses on the tunnel oxide layer allowing the carriers to tunnel from the suspended gate into the SiNDs through the tunnel oxide layer by using the Fowler-Nordheim tunnelling mechanism as shown in Fig. 2.3b [17, 34, 30]. Once in contact, if the applied voltage is increased, the contact area between those layers will also increase as well as the current density. In contrast, by reducing the applied voltage, the contact area between the suspended gate and the tunnel oxide layers is reduced but remains attached even at voltages lower than the pull-in voltage. Those layers remain in contact due to the stiction generated. This occurs due to the electrostatic and short range forces such as the van der Waals and Casimir are in equilibrium with the mechanical restoring force of the doubly-clamped structure [35, 36, 37, 38, 39]. The equilibrium between the forces is disturbed when the applied voltage is reduced beyond the pull-in voltage. By reducing the applied voltage, the restoring force overcomes the electrostatic and short-range forces allowing the suspended gate to return sharply to its initial flat position. This point of imbalance is known as the pull-out voltage. When pull-out occurs, the tunnelling mechanism (Fowler-Nordheim) is quickly interrupted. As a result, the electrons are kept stored within the floating gate on the SiNDs layer as depicted in Fig. 2.3c. Taking into account that the SiNDs are isolated from the substrate and from the control gate by the tunnel oxide layer and the air-gap, the presence of leakage current, under these circumstances, towards the substrate, between them or to the control gate, is negligible.

2.1.2 Erasing Sequence

In order to erase the memory device, a positive bias voltage is applied into the suspended gate electrode. As a consequence a similar sequence as for the programming process is performed. When the control electrode is biased, the suspended gate bends downwards by action of the electrostatic force due to the floating gate being negatively charged (programmed). By increasing the biased voltage, the doubly-clamped beam structure reaches the pull-in point and collapses on the tunnelling oxide layer allowing the electrons to tunnel from the floating gate into the suspended gate through the tunnel oxide layer

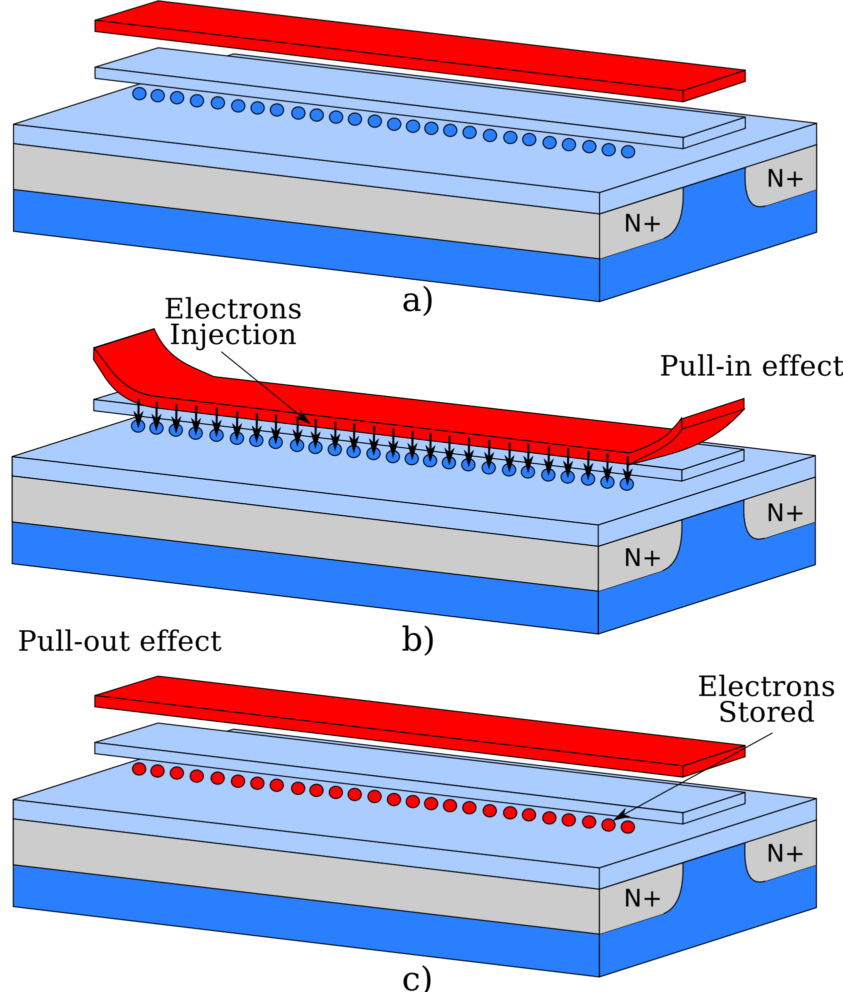


FIGURE 2.3: Programming sequence performed by the suspended gate when a) the suspended gate electrode is biased, b) the suspended gate bends downwards and eventually collapses allowing the tunnelling process (electrons) into the floating gate layer and c) the suspended gate detaches from the oxide layer and the electrons are trapped within the floating gate

as shown in Fig. 2.4b. By reducing the applied voltage, the suspended gate remains attached to the tunnel oxide layer due to the short-range and electrostatic forces being in equilibrium with the mechanical restoring force. Further reduction in the applied voltage creates an unbalance between those forces allowing the suspended gate to return to its initial flat position and remain isolated from the floating gate. As a result, the floating gate is positively charged as depicted in Fig. 2.4c.

2.1.3 Reading Sequence

The reading operation in the Suspended Gate Silicon Nanodot Memory is performed through activation (bias) of the readout element V_{DS} (MOSFET) whilst the suspend gate is unbiased (V_{SG}= 0 V). Without prior knowledge about the status of the memory node, when the readout element is activated, a current pulse is displayed. This shows

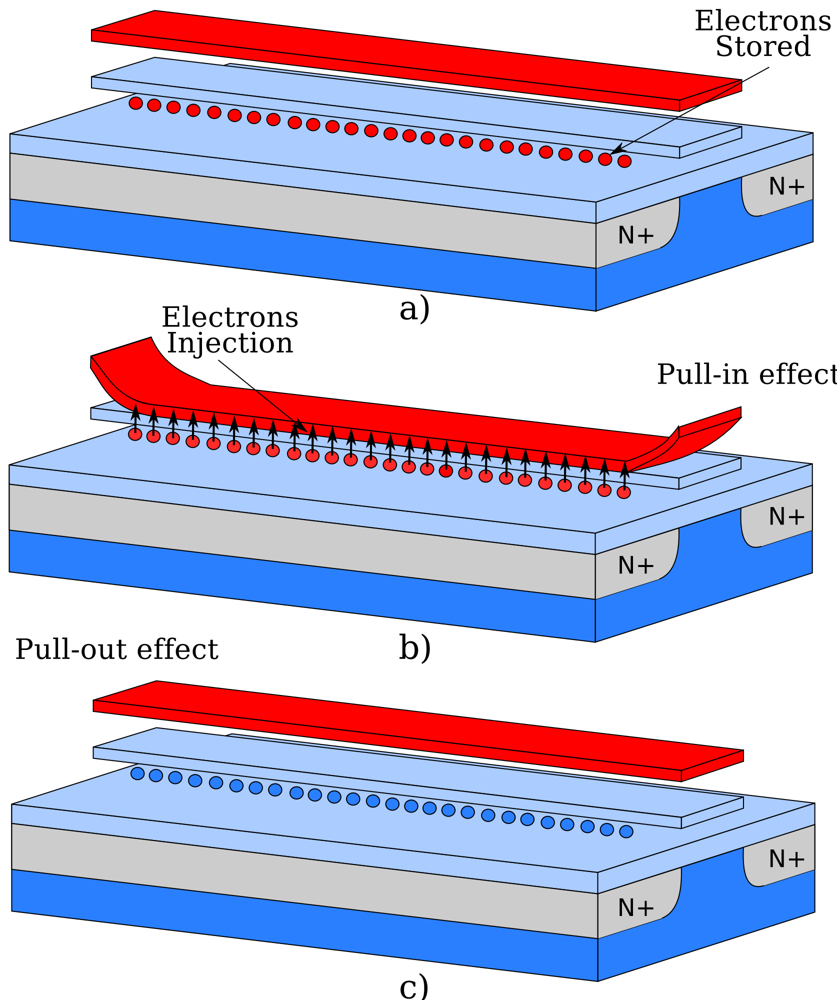


FIGURE 2.4: Erasing sequence of the SGSNM cell performed by the suspended control gate when a) the doubly-clamped beam has been biased, b) by action of the applied voltage and the electrostatic force the beam collapses on the oxide layer allowing the electrons to tunnel into the floating gate, and c) by reducing the applied voltage an imbalance between the restoring force with the electrostatic and short-range forces occurs allowing the suspended gate to return to its initial flat position

that the memory node is either positively charged or it has been erased. The process that occurs within the SGSNM cell at its read operation is that, upon activating the MOSFET (biased), it "sees" the charge stored within the floating gate, which for this operation becomes the control gate. As the memory node has been positively charged, within the depletion region, an inversion layer is formed by action of the charge contained in the floating gate, which as a consequence forms the channel as depicted in Fig. 2.5a. Conversely, when the readout element is activated and no-current pulse is displayed, it is indicative that the SGSNM cell is programmed. Within the memory cell, as the floating gate has been negatively charged, there is no inversion layer formation within the depletion region and as a consequence no current pulse is displayed (Fig. 2.5b). Figure 2.5c shows the status of the readout element when it is biased for the programming and erasing memory node status at $V_{SG} = 0$ V.

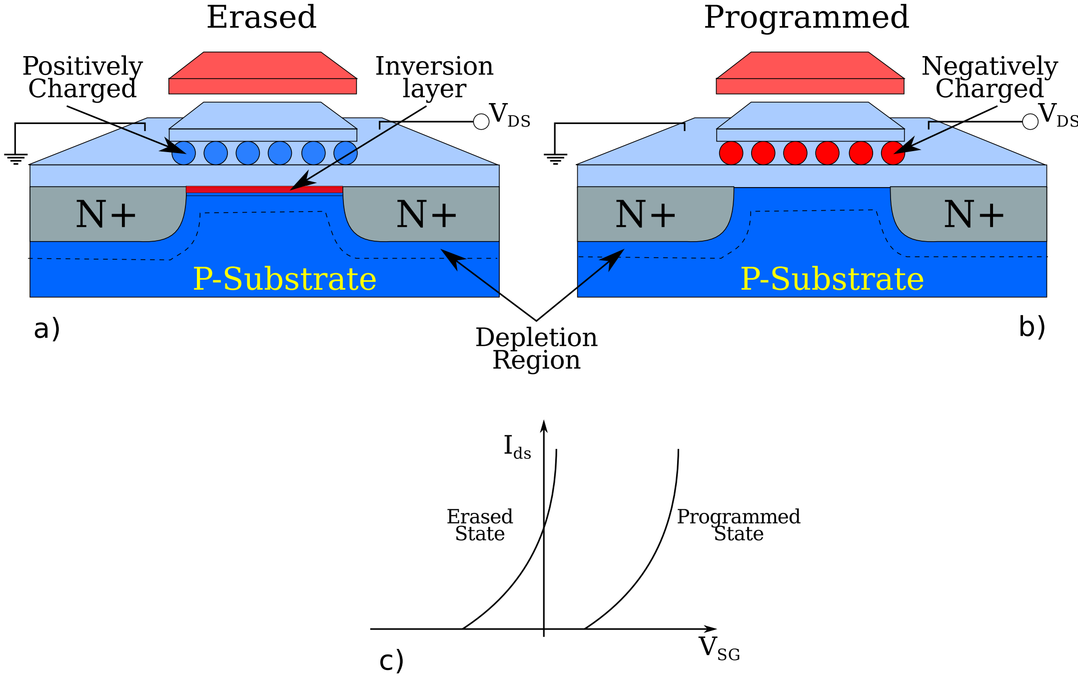


FIGURE 2.5: Schematic diagrams that represent the reading operations when a) the SGSNM cell has been erased, b) programmed and c) shows the I_{ds} - V_{SG} transfer characteristics

Table 2.1 summarises the programming, reading and erasing operations performed by the different control signals within the SGSNM cell such as the control voltage (V_{SG}), readout bias voltage (V_{DS}) and the current signal (I_{ds}).

TABLE 2.1: Programming, reading and erasing characteristics of the SGSNM cell

Process	Bias	Pull-in	Pull-out	Reading process
Prog.	$-V_{SG} \leq 0$ V	$-V_{SG} \leq -V_{Pull-in}$	$-V_{SG} \geq -V_{Pull-out}$	$V_{SG} = 0$ V & $V_{DS} > 0$ V
	$V_{DS} = 0$ V	$V_{DS} = 0$ V	$V_{DS} = 0$ V	$I_{ds} \approx 0$ A
Erase	$V_{SG} \geq 0$ V	$V_{SG} \geq V_{Pull-in}$	$V_{SG} \leq V_{Pull-out}$	$V_{SG} = 0$ V & $V_{DS} > 0$ V
	$V_{DS} = 0$ V	$V_{DS} = 0$ V	$V_{DS} = 0$ V	$I_{ds} > 0$ A

Then, the programming, reading and erasing operations are described as functions of the electro-mechanical movement of the suspended control gate as well as the readout element. It is essential to investigate the full electro-mechanical behaviour of the doubly-clamped structure by using reliable and fast models in which specific materials such as aluminium, copper or poly-Si can be implemented as the suspended gate and oxide materials such as the SiO_2 and Si_3N_4 can be implemented as the tunnel oxide layers [40, 41, 42, 43, 44].

Chapter 3

Suspended Control Gate Analysis

The Suspended Gate Silicon Nanodot Memory has its foundations in the dynamic movements of the suspended control gate via the pull-in and pull-out effects. The understanding of those effects through electromechanical analyses and the interaction between the suspended control gate with the short range forces (when in contact), will lead to the acquisition of reliable models for the programming and erasing operations of the SGSNM device. This is why the suspended control gate structure is scrutinised in detail by performing mathematical derivations as well as trustworthy numerical simulations. In this Chapter a detailed analysis of the electromechanical properties of the suspended control gate is performed. The analysis starts by considering a double-plate capacitor structure which is used as the most basic and simplest representation of the doubly-clamped suspended control gate. Here, the equations that govern the electromechanical behaviour are analytically obtained. In addition, this analysis is complemented with numerical three-dimensional (3D) finite element method (FEM) analysis. The Chapter concludes with the implementation of the electromechanical behaviour as compact voltage dependent models in an analog hardware description language (AHDL) [45, 46].

3.1 Pull-in analysis

The pull-in effect is defined by the action in which the suspended gate collapses over the tunnel oxide layer as a consequence of the voltage applied on the control gate electrode. This effect controls two key processes that involve voltage and speed parameters within the SGSNM, the programming and erasing operations. The sharper the mechanical pull-in, the higher the current density and the faster the transfer of carriers is performed. Those characteristics are obtained by designing the suspended control gate taking into account materials, shapes, thicknesses and dimensions implemented in accurate, robust and trustworthy models among others. Electric and electromechanical characteristics within the model are analysed in order to obtain the programming and erasing voltages

through a combination of the pull-in and pull-out effects. A schematic diagram of the doubly-clamped capacitor model is depicted in Fig. 3.1. From this model, the analytical equations that describe the pull-in and pull-out effects are obtained.

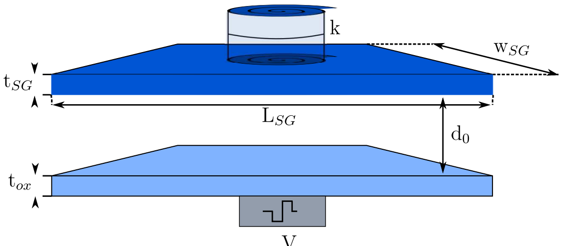


FIGURE 3.1: Double-plate capacitor schematic diagram used to analytically obtain the pull-in and pull-out voltage equations

Where L_{SG} and w_{SG} define the length and width of the plates, respectively, t_{SG} represents the thickness of the top plate, d_0 is defined as the initial air-gap, ϵ_0 is the free space permittivity, k represents the spring constant and V is the applied bias voltage on the electrode. Given that it depends on the structural shape, the spring constant is not a fixed entity that can be used indistinctly within any analysis. The spring constant can be modelled as a membrane, cantilever or as a beam [42]. In our case, the model required for the suspended control gate is a beam fixed at both ends [42, 44]. The function that defines the doubly-clamped beam structure as spring constant is defined as

$$\mathrm{k} = \frac{16\mathrm{E}\mathrm{w}_{SG}\mathrm{t}_{SG}^3}{\mathrm{L}_{SG}^3} \tag{3.1}$$

where E represents the Young's modulus [47].

The equations that govern the behaviour of the double-plate capacitor are divided into electrical and mechanical force. The diagram shown in Fig. 3.1 is first transformed to its equivalent circuit. Hence, the equations that govern the structure are defined as

$$F_E = \frac{d^2M}{dz^2} + kz \tag{3.2}$$

$$F_M = -kz \tag{3.3}$$

By equalising Eqs. (3.2) & (3.3) we obtain

$$\frac{1}{2}\frac{\epsilon_0\epsilon_r AV^2}{d^2} = kz \tag{3.4}$$

$$V_{Pull-in} = \sqrt{\frac{8kd_0^3}{27\epsilon_0 A}}$$

Once the equations that govern the movements of the capacitor model have been obtained, those are re-arranged as a function of the applied voltage by including the spring constant function in it. As a consequence, the equation to calculate the pull-in voltage ($V_{Pull-in}$) is defined as [42, 47]

$$V_{Pull-in} = 8\sqrt{\frac{2Et_{SG}^3 d_0^3}{27\epsilon_0 L_{SG}^4}} \tag{3.5}$$

where ϵ_0 represents the free space permittivity.

The above equation calculates the inflection point as a function of the applied voltage at which the upper plate capacitor is attracted by action of the electrostatic force. This force overcomes the stiffness of the material (mechanical restoring force), allowing the plate to collide over the bottom plate. In here, the equation (3.5) shows that the pull-in voltage has a strong relationship with the length of the double-plate structure. It also shows that it is possible to scale-down the dimensions of the structure in a linear way.

3.2 Pull-out analysis

The equation governing the pull-out voltage assumes that both plates are initially in contact. The electrostatic and electromechanical forces that act at this stage are controlled by the applied voltage. When this voltage is reduced, the spring constant of the structure (stiffness of the material) has a stronger presence each time the voltage is reduced. By reducing the applied voltage, the forces are overcome by the spring constant and it is here that the top plate returns to its initial separated position. The schematic diagram that is analysed to obtain the pull-out voltage is depicted in Fig. 3.2.

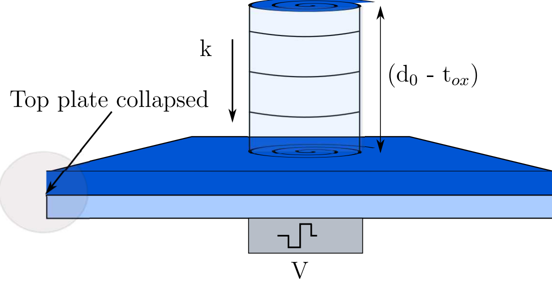


FIGURE 3.2: Schematic diagram of the doubly-plate capacitor model modified to calculate the pull-out voltage

The difference between the substrate and the initial air-gap between plates is assumed

to be the initial condition generated by the spring constant. Taking into account these changes and by analysing the above structure, the pull-out equation ($V_{Pull-out}$) is defined as

$$V_{Pull-out} = \sqrt{\frac{2kt_{ox}^2}{\epsilon_0 \kappa_{ox} A}(d_0 - t_{ox})} \tag{3.6}$$

where κ_{ox} and t_{ox} represent the dielectric constant and the thickness of the dielectric material, respectively.

The above model is useful to estimate the pull-in and pull-out voltages taking into account a minimum set of parameters such as electrostatic and electromechanical forces. Nevertheless, the set of forces that are implemented to calculate these voltages refers to electrostatic nature and does not consider other kinds of forces such as short-range forces that have a major impact at the nano-scale [35, 36, 37, 48]. Hence, to complement the analysis the pull-out analysis requires the inclusion of short-range forces such as van der Waals and Casimir in the calculations in order to obtain an accurate pull-out voltage equation. Taking those short-range forces into consideration and rearranging Eq. (3.6), the pull-out equation is re-defined as

$$V_{Pull-out} = \sqrt{\frac{2kt_{ox}^2}{\epsilon_0 \kappa_{ox} A}(d_0 - t_{ox}) - \frac{A_h}{3\pi \kappa_{ox} t_{ox}}} \tag{3.7}$$

where A_h is defined as the Hamaker constant [35, 49, 50]. By including the short range force (van der Waals force) into the pull-out voltage equation, it relies in a accurate and reliable behaviour of the pull-out voltage.

The above set of equations obtained to calculate the pull-in and pull-out voltages considers ideal and homogeneous parameters such as materials and shapes for its analysis. Those parameters, despite their well known physical properties such as Young's modulus or permittivity, do not physically represent the work performed by the structure at each point. Despite these improvements, these analytical models based on electromechanical structures are far too complex and require a long time which does not guarantee the accuracy or robustness required. By increasing the complexity of the model to include "extra" parameters which are required to obtain high accuracy, the equation will give reliable information to some extent and every time that a new parameter arises, it needs to be included within the function which in the end becomes inefficient. This is why, instead of using a model, we are looking to use a tool capable of analysing the pull-in and pull-out effects of the doubly-clamped structure in a dynamic way. Within its core, this tool should include electrical and mechanical properties such as elasticity, plasticity, stress, tension and compression required by the analysis [41, 51]. The finite element method enables us to analyse structures from different perspectives such

as electric, electromechanical or magnetic in a reliable, consistent and robust manner in a dynamic way not only in one-dimension (1D) but in two- (2D) and three- (3D) dimensions without being affected by the complexity in the implementation. Nowadays, there are several commercial software companies that provide FEM packages such as Matlab, Comsol, CoventorWare and Ansys, among others that allow us to analyse a broad variety of structures in 2D and 3D in a considerably low computational time, over a short learning period and at a low cost [43, 44, 52, 53]. Although those FEM packages are capable to analyse the doubly-clamped beam structure and calculate the pull-in and pull-out trajectories, each one has its own advantages as well as drawbacks. According to the characteristics present in each package; CoventorWare fulfils the requirements that enable the analysis of the behaviour of the suspended control gate. The qualities that differentiate this package in comparison to the others are the built-in libraries that contain *in situ* measurements for micro/nano-electromechanical structures as well as for CMOS technology, the fabrication process simulator, the circuit level simulator, and the engine that creates and solves the meshing algorithms that fit the structural shapes at inter-layers, among other built-in algorithms present in it [44].

3.3 Numerical Analysis of the Doubly-Clamped Structure

Analysing the behaviour of the suspended control gate through CoventorWare requires following a sequence that involves several steps in which the material thickness of each layer, the dimensions and shapes of the device under test are defined as initial conditions. It is at this early stage that the sacrificial layer is defined within the CoventorWare model. A schematic diagram of the structure implemented in CoventorWare is shown in Fig. 3.3.

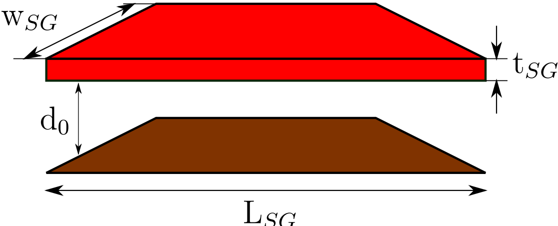


FIGURE 3.3: Schematic diagram of the doubly-clamped beam structure implemented in CoventorWare

The structure described above is solved by taking advantage of each of the features that CoventorWare has to offer, by using one of the built-in algorithms' trajectory analysis present in the package. Several materials are proposed as suspended control gate candidates. These materials and their physical properties are listed in Table 3.1

The implementation process of the doubly-clamped structure within CoventorWare in which the initial set of layers is defined, the etching-out process of the sacrificial layer

TABLE 3.1: Set of materials and parameters proposed as suspended control gate

Materials	Thickness (nm)	Dimensions (µm × nm)	Poisson's ratio	Young's modulus (GPa)
Aluminium (Al)	30	1 × 300	0.3	77
Copper (Cu)	30	1 × 300	0.36	128
Poly-Si	30	1 × 300	0.22	160
SiO_2	10	1 × 300	0.17	73
Si_3N_4	10	1 × 300	0.27	222

and the meshing analysis are displayed in Fig. 3.4. Once the structures have been implemented on the simulator, a processing operation that involves boundary conditions, size and density of the grid, inter-layer analysis, mesh algorithm, resources and the computing time required for the process are defined within CoventorWare.

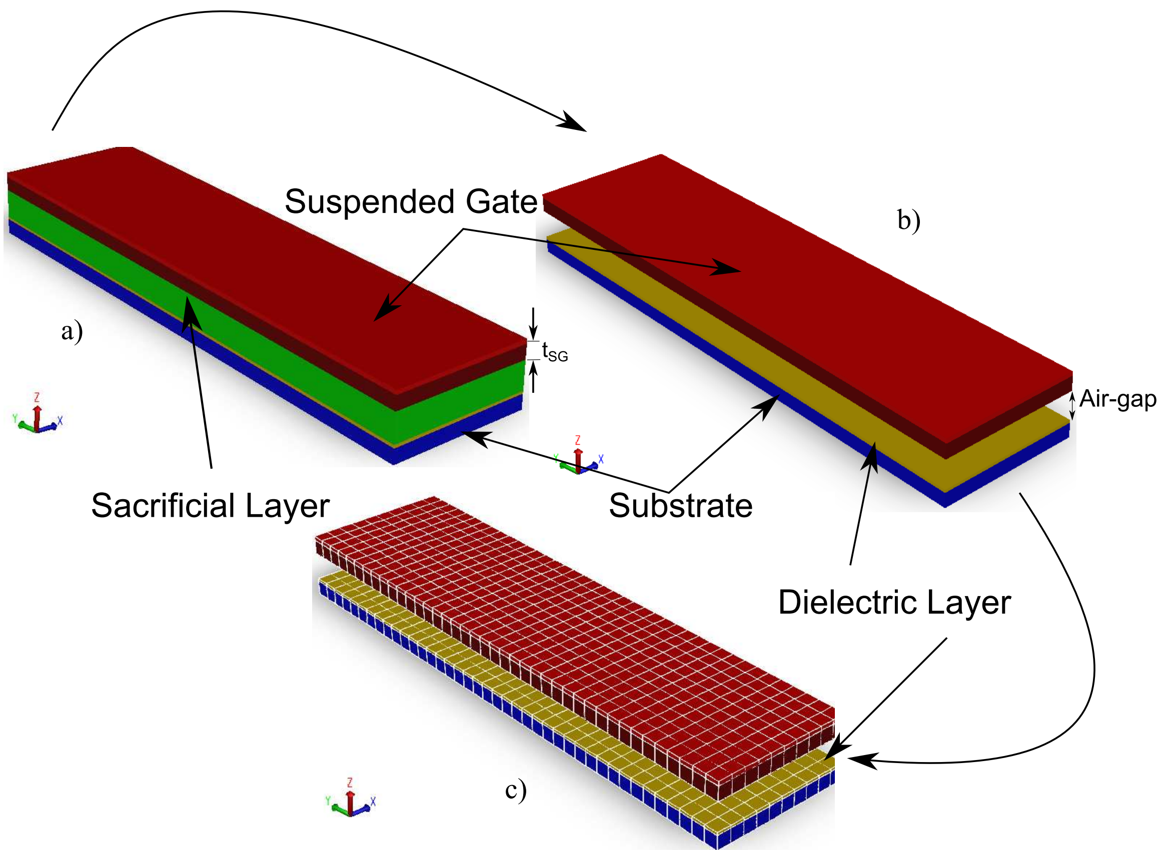


FIGURE 3.4: Set of schematic diagrams that show the main points within the implementation of the doubly-clamped structure in CoventorWare

The initial stage (Fig. 3.4a) it is possible to identify a whole set of layers that form the suspended control gate: substrate (blue), thermal oxide layer (yellow), sacrificial layer (green) and control gate (red). By using the fabrication process tool, the sacrificial layer is etched-out to form a suspended control gate, this sacrificial layer is required by the software in order to have a rather realistic fabrication process (Fig. 3.4b). As a part of processing, the entire structure is meshed by using a built-in mesh algorithm

known as Manhattan bricks. Within this part both sides, top and bottom of the beam and the top part of the tunnel oxide layer, are named and fixed with the exception of the beam top and bottom faces. After specifying the voltage and type of analysis, the pull-in detection algorithm is implemented. This algorithm analyses the tension, compression and deformation at each node within the mesh as mechanical properties. The electromechanical analysis leads to finding the inflection point at which the beam collapses over the substrate. The results obtained from this numerical analysis are depicted in Fig. 3.5.

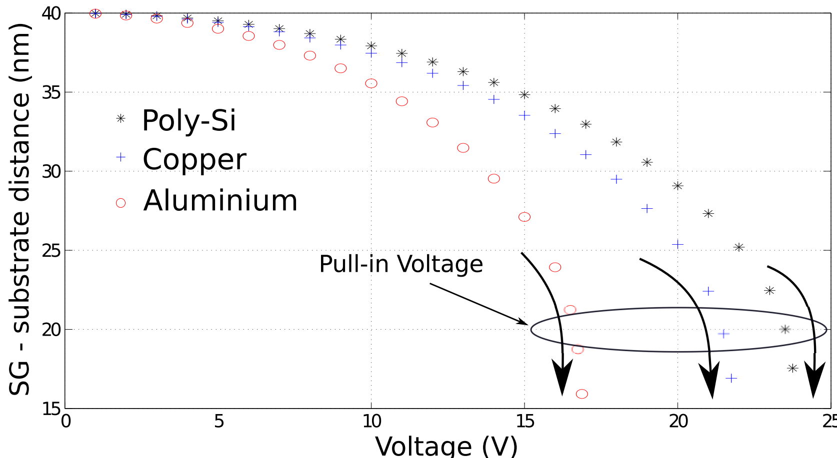


FIGURE 3.5: Pull-in 3D-FEM analysis of a set of parameters shown in Table 3.1

The set of pull-in curves obtained from the FEM analysis shows that the three suspended gate materials pulled-in under 25 V. The 3D FEM analysis assumes the interaction between the electrostatic and mechanical forces in the analysis without taking into account other forces such as short range forces. From those structures it is possible to see the non-linearity which is highly desired for the control gate. From the point of view of the operation voltage, it is desirable to drive the programming and erasing operations at as low voltage as possible. According to the results obtained from this analysis, aluminium seems to be a suitable candidate for the suspended control gate due to its low pull-in voltage (17 V) compared with the other structures (copper @ 21 V and poly-Si @ 24.2 V). The pull-in voltages obtained by using the 3D FEM analysis are quite dissimilar to those voltages obtained by using the double-plate capacitor model *e.g.* for the aluminium beam, it pulls-in at $\approx$ 4.5 V rather that 17 V as the obtained by using the FEM analysis. Once the control gate material has been chosen, the pull-out voltage is analysed. To calculate the pull-out voltage, the control gate is coupled with SiO_2 and Si_3N_4 tunnel oxide layers, respectively. Once coupled, both structures are analysed through the trajectory analysis in which the applied voltage is swept from 0 to 25 V and vice versa. This analysis is performed in order to obtain the pull-in and pull-out curves merged in a single graph as a displacement - voltage (D - V) hysteresis cycle. The hysteresis curves for both analyses are overlapped and depicted in Fig. 3.6

By overlapping both hysteresis curves, several characteristics inherent to the structures

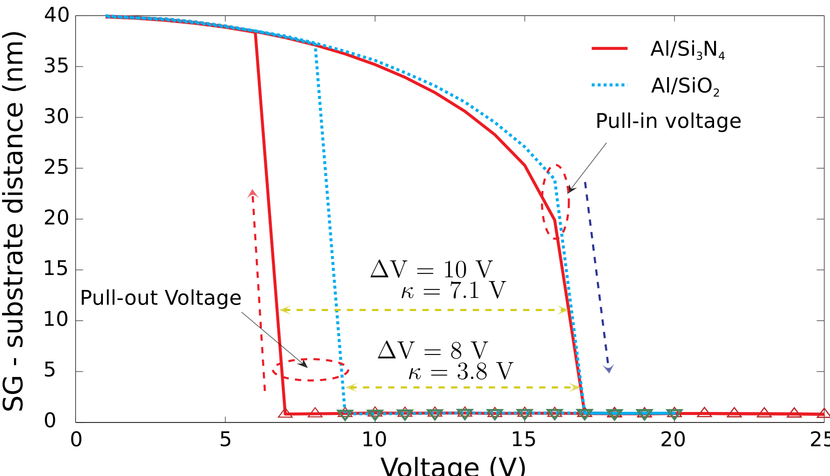


FIGURE 3.6: 3D-FEM analyses used to find the pull-in and pull-out trajectories for two structures Al / SiO_2 and Al / Si_3N_4, respectively

arise. These characteristics show the effects that the tunnel oxide permittivity has on the behaviour of the suspended control gate. In the first instance, it shows that the tunnel oxide permittivity has a negligible effect on the pull-in trajectory as shown in both D - V hysteresis curves due to both pull-in trajectories converging at the same point. For the pull-out effect, the small variation between SiO_2 ($\kappa = 3.8$) and Si_3N_4 ($\kappa = 7.1$) creates a fairly large stiction window (2 V) between both hysteresis cycles [40, 54, 55]. The difference in the stiction window between both materials is produced, mainly, by the permeability of the material in which the thickness of the layer plays an important role. Thus, the pull-out voltages for both structures occur at 7 V and 9 V, respectively. It is also desirable to obtain a short stiction window due to endurance and degradation caused by the tunnelling mechanisms on the tunnel oxide layer. By considering the shortest stiction window, SiO_2 is defined as the tunnel oxide layer for the SGSNM. Taking into account those parameters, the programming and erasing operation voltages are not competitive for today's non-volatile memory market. In order to be competitive, those voltages are required to have a threshold under 10 V to match the voltages proposed by emerging memories such as MRAM or PCRAM [1]. In order to reduce the high operation voltage, there are a few modifications in the suspended gate structure that need to be addressed in pursuit of lower programming and erasing voltages.

This is why the area as well as the beam material are kept constant while the beam thickness and the air-gap are defined as the parameters to be modified. The goal of modifying those parameters is to find a programming voltage under 10 V which can be competitive within the non-volatile memory market. The set of parameters that are tested as thickness and as air-gap are listed in Table 3.2.

As a result of the parameters proposed, the numerical analysis shows a satisfactory result in which the set of parameters under test show that whole pull-in under 10 V. Figure 3.7 shows the set of pull-in trajectories and its electro-mechanical properties that can

TABLE 3.2: Parameters used to optimise the structure

Material SG	Suspended Gate (nm)	Air-Gap (nm)	Dimensions (μm × nm)
Al	30	30	1 × 300
Al	30	20	1 × 300
Al	20	20	1 × 300
Al	20	15	1 × 300

be used in combination with the area as the suspended control gate parameters.

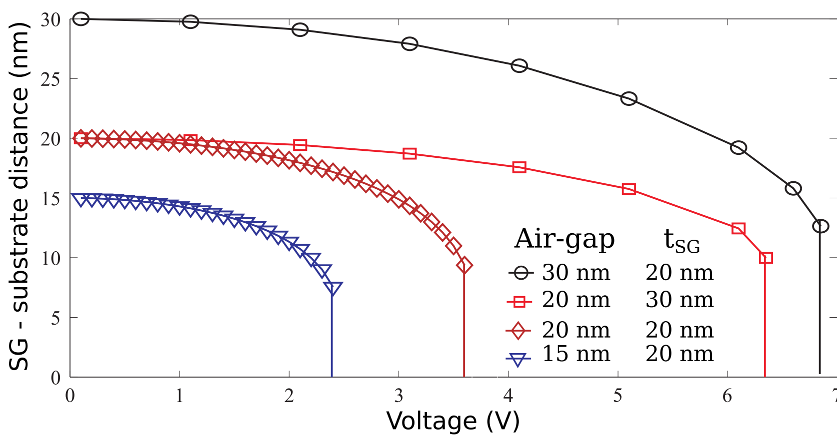


FIGURE 3.7: Pull-in trajectories for the set of parameters specified in Table 3.2 that fulfils the requirements of pull-in under 10 V

From the set of pull-in trajectories, the suspended control gate is carefully chosen based on its mechanical properties such as thickness and robustness, as well as the inherent non-linear behaviour as result of the combination between the air-gap and material thickness. The requirements for the suspended control gate are gathered by the structure with 30 nm thickness at 20 nm air-gap. Chosen parameters are coupled to the SiO_2 tunnel oxide layer and analysed numerically in order to obtain the D - V hysteresis cycle. For this analysis, a bias voltage is swept on the suspended gate electrode from 0 to 10 V and vice versa. As a result the pull-in and pull-out trajectories are obtained. These trajectories are merged to form a hysteresis cycle which is displayed as a 3D FEM sequence, as it is displayed in Fig. 3.8.

When a low positive bias voltage is applied on the control gate electrode, it does not seem to have any influence on it due to the fact that the stiffness of the material is greater than the electrostatic force generated from 0 to 3.5 V. The influence of the applied voltage on the electrode is invisible until it has reached 4 V as depicted in Fig. 3.8b. By increasing the applied voltage, the beam deformation becomes visible until it collapses over the tunnel oxide layer due to the fact that the electrostatic force has overcome the restoring force of the beam (Fig. 3.8e). Further increments on the applied voltage rely on a major contact area as can be seen in Fig. 3.8f. When the applied

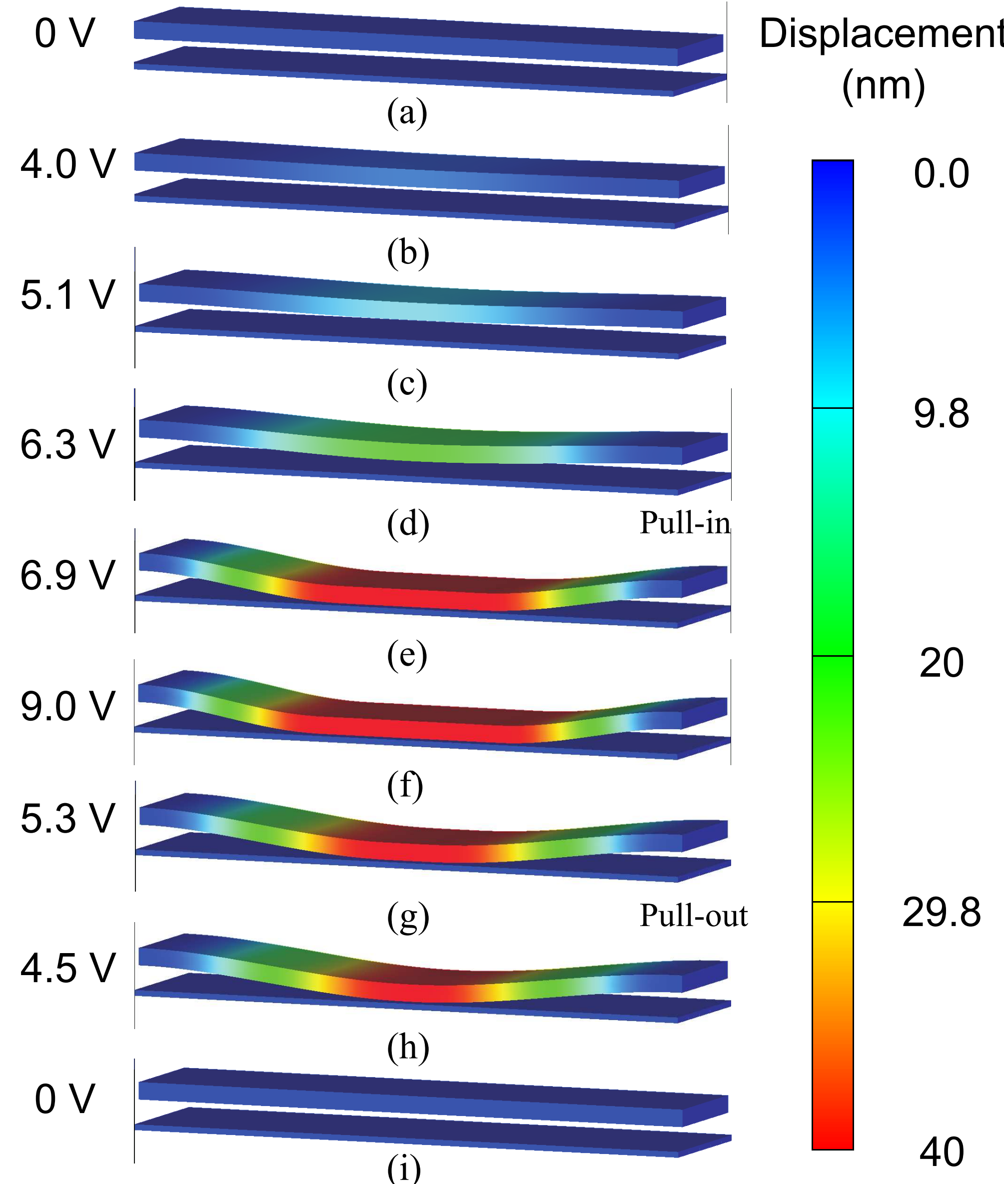


FIGURE 3.8: Pull-in and pull-out sequence of the doubly-clamped Al beam (30 nm) coupled to SiO_2 tunnel oxide layer (20 nm)

voltage is reduced, the control gate remains attached to the tunnel oxide layer even at voltages similar to the pull-in (Fig. 3.8g). By reducing the applied voltage beyond pull-in voltage, those forces are not in equilibrium any more and the control gate detaches from the tunnel oxide layer returning to its initial flat position as depicted in Fig. 3.8i. A complementary analysis to the mechanical analysis, an electrical-based analysis is performed on the doubly-clamped beam structure. In here, the capacitance - voltage (C - V) hysteresis cycle is obtained. These hysteresis cycles describe the electrical and electromechanical comportment of the suspended control gate. Both hysteresis cycles are displayed in Fig. 3.9.

The suspended control gate is directly responsible for the programming and erasing

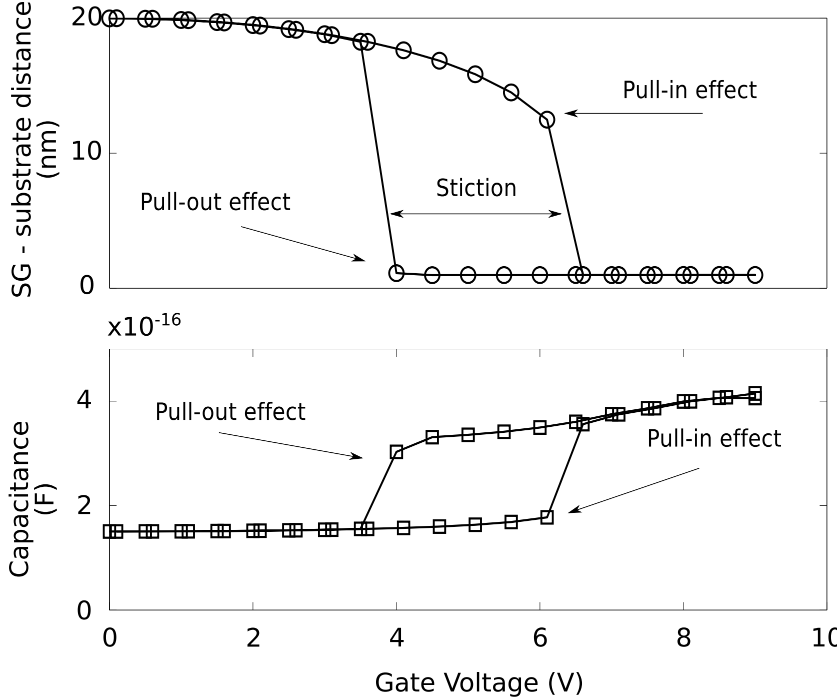


FIGURE 3.9: Hysteresis cycles. (a) Displacement-Voltage, showing the pull-in/pull-out voltages and the stiction endurance; (b) Capacitance-Voltage characteristics

operations of the SGSNM due to the pull-in and pull-out effects and indirectly for the reading operation due to it remains isolated from the memory node by a combination of the air-gap and the tunnel oxide layer. According to the numerical analysis, these operations are performed at a relatively low voltage. To sum up, Fig. 3.10 shows schematic diagrams for the programming, reading and erasing operations as well as the hysteresis curves that intervene when the SGSNM perform those processes. It also need to be considered that the programming and erasing processes are no-symmetrical due to the charging/discharging processes.

3.3.1 Implementation as behavioural models

The implementation of the C - V hysteresis curves within a circuit simulator is performed through the conversion of the C - V hysteresis cycle into analog voltage dependent functions. Such implementation is made through an analog hardware description language (AHDL) which is interpreted within the CAD simulators as embedded analog libraries [56, 57, 58, 59]. The C - V hysteresis curve is adapted to the mnemonic. In here the analog implementation can be considered as voltage or current dependent functions and a graphical representation of the mnemonic is depicted in Fig. 3.11. The mnemonic used for this implementation is adaptive at a three word system. Once divided, the analog implementation is made by using algebraic algorithms. The behavioural functions have several advantages such as easy to implement, fast analysis, reduced number of coefficients rather than complex and heavy, computationally speaking, mathematical

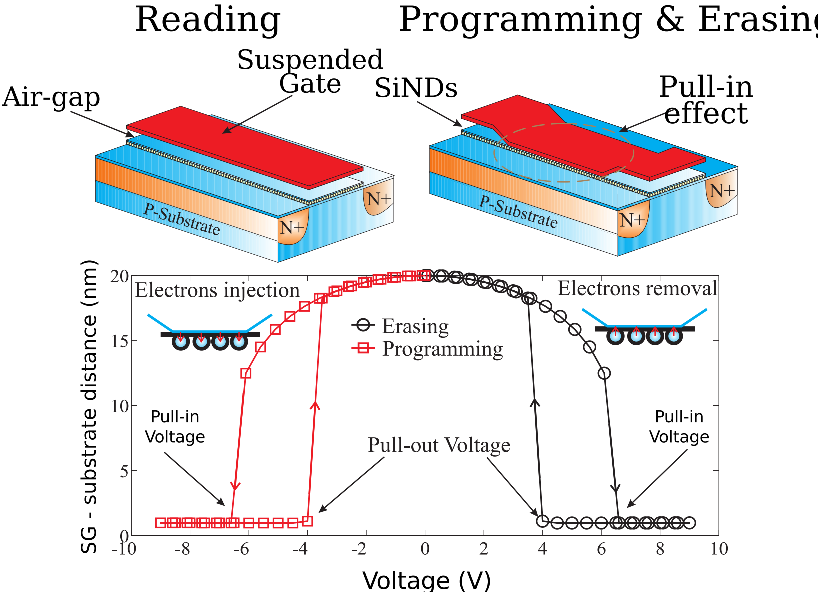


FIGURE 3.10: Schematic diagram that describe the programming, reading and erasing operations with its associated hysteresis cycles

functions [51, 60, 61, 62]. The AHDL language that is used for the implementation of those voltage dependent functions is Verilog-AMS [45, 63, 64].

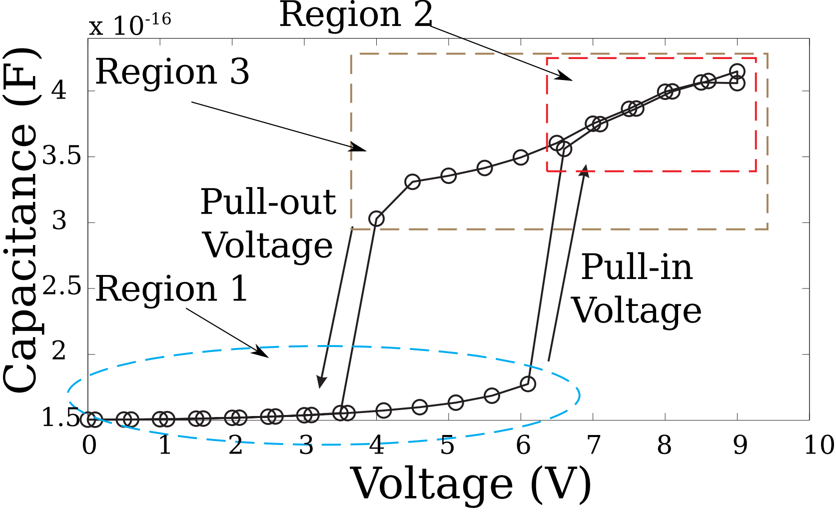


FIGURE 3.11: Capacitance-Voltage Hysteresis Curves Implementation

According to Fig. 3.11 the implementation for each mnemonic has its own particularities. Region 1 considers the section from the initial state of the suspended gate at zero volts until the beam collapses at pull-in voltage ($\approx$ 6.7 V). The capacitance variation associated to the beam displacement from the isolated flat position until the suspended control gate has collapsed on the tunnel oxide layer is 3.0×10^{-17} F. The algorithm used to fit region 1 is a third order polynomial. This algebraic polynomial used offers several advantages and a good trade-off between the accuracy and fast coding. The first mnemonic displayed as a dot-line and the approximation curve made by the algebraic

algorithm as a continuous line (Fig. 3.12).

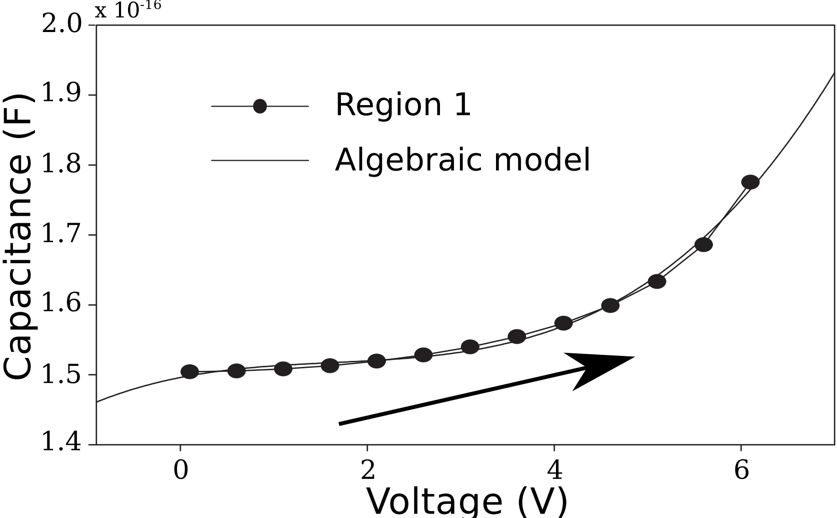


FIGURE 3.12: Region 1 approximated by using a third degree algebraic polynomial. The dot-line represents the discreet voltage dependent function and the continuous line represents the algebraic approximation

The second region is bounded from the point at which the suspended gate collapses over the tunnel oxide layer (pull-in voltage) until the maximum voltage applied on the suspended gate electrode is reached (≈ 10 V). In this region, the capacitance variation is produced, in the first instance, by the sharp contact between the suspended gate and tunnel oxide layer and later increased by the contact area (5.7×10^{-17} F). This region is also implemented by using a third order polynomial, similar as for region 1. Figure 3.13 depicts the curve for region 2 represented as a dot-line and the approximation curve produced by the algorithm as a continuous line.

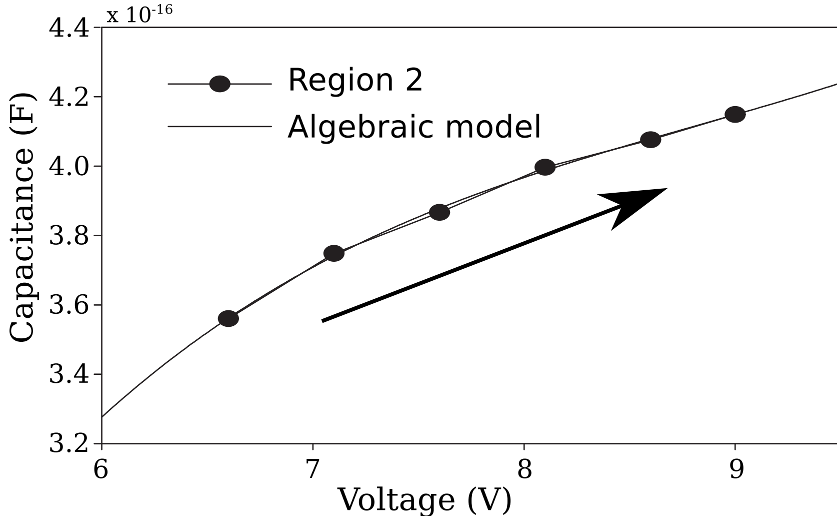


FIGURE 3.13: Region 2 defined by the pulled-in beam and the increased contact area. In here, the dot-line represent the discreet C-V curve and the continuous line the approximation generated by the third degree polynomial

Finally, region three defines the stiction - pull-out process that starts from the maximum

applied voltage (10 V) until the pull-out effect takes place (4 V). The erratic behaviour depicted by the curve is caused by the unbalance between the restore and the electrostatic forces [65]. The interaction at short-range between those forces allows the suspended gate to remain attached to the tunnel oxide layer. Whilst the applied voltage is reduced, the unbalance between those forces is visible and concludes with the suspended gate isolated from the tunnel oxide layer. In here, the capacitance variation produced within the stiction window is $\approx 1.0\times10^{-16}$ F. To approximate region 3, a similar algorithm as before is used in the analysis. Figure 3.14 shows the discreet C - V stiction curve (dot-line) and the approximation curve represented as the continuous line.

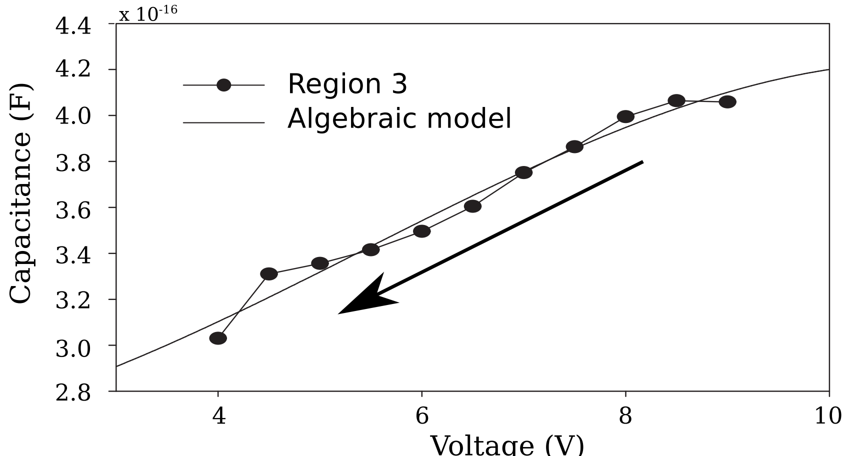


FIGURE 3.14: Region 3 is defined as the stiction window starting from the maximum voltage applied until the suspended gate is detached from the tunnel oxide layer. The dot line represents the discreet erratic detachment behaviour and the continuous line the approximation curve generated by the algebraic model

The approximation curve generated by the polynomial algorithm is not as smooth as in the previous trajectories. Although other algorithms were tested, the improvements provided by those functions did not reflect the trade-off between accuracy, implementation and processing time. Hence, the algebraic algorithm used to approximate this region was used. The set of curves obtained has the form of a third degree algebraic function defined as

$$f(x) = \mathrm{M}_3 x^3 + \mathrm{M}_2 x^2 + \mathrm{M}_1 x^1 + \mathrm{M}_0 x^0 \tag{3.8}$$

where $\mathrm{M}_{1,2,3}$ stands for the coefficients that represent each region within the algebraic function and x is defined as the applied voltage.

The set of coefficients obtained from the algebraic algorithms is implemented within AHDL as behavioural models by following a sequence to differentiate between the programming and erasing operations. When a positive bias voltage is applied on the suspended gate electrode, the set of C - V curves that are used within this operation are positive as depicted in Fig. 3.10. From these C - V curves, the coefficients of the voltage

dependent functions are shown in Table 3.3.

TABLE 3.3: Coefficients obtained by approximating the C - V hysteresis curve with third order algorithms

Region	M_3	M_2	M_1	M_0
1	2.4668×10^{-19}	-1.2126×10^{-18}	2.6251×10^{-18}	1.4961×10^{-16}
2	1.5988×10^{-18}	-4.2254×10^{-17}	3.8947×10^{-16}	-8.3338×10^{-16}
3	-4.4919×10^{-19}	4.0622×10^{-18}	3.0217×10^{-19}	2.5997×10^{-16}

In contrast, to program the memory node a negative bias voltage is applied on the suspended gate electrode. In this process, as depicted in Fig. 3.10, the C - V hysteresis cycle is negative. Therefore, the coefficients used within the voltage dependent functions are modified. Table 3.4 shows the set of coefficients associated with the negative C - V hysteresis curve.

TABLE 3.4: Coefficients obtained by approximating the C - V hysteresis curve with third order algorithms

Region	M_3	M_2	M_1	M_0
1	-2.4668×10^{-19}	-1.2126×10^{-18}	-2.6251×10^{-18}	1.4961×10^{-16}
2	-1.5988×10^{-18}	-4.2254×10^{-17}	-3.8947×10^{-16}	-8.3338×10^{-16}
3	-4.4919×10^{-19}	-4.0622×10^{-18}	-3.0217×10^{-19}	-2.5997×10^{-16}

The voltage dependent functions shown in Tables 3.3 and 3.4 are merged in order to show the full behaviour of the suspended control gate as it is depicted in Fig. 3.15. In here, it is possible to identify the programming, erasing and the reading sequences performed through the suspended control gate and the C - V hysteresis curve.

3.4 Transient Analysis

The pull-in and pull-out effects are linked to the speed that the SGSNM cell can achieve during the programming and erasing operations. In order to find the time at which the pull-in effect occurs, a 3D FEM transient analysis is performed. Within this numerical analysis the electrostatic as well as the Casimir forces are interacting between the suspended gate and the tunnel oxide layer. As a result of the transient analysis Fig. 3.16 shows the displacement - voltage analysis.

From this numerical analysis, it was found that for 1 μm-long beam the time at which the pull-in effect occurs for the programming and erasing operations is 0.8 nsec as depicted in Fig. 3.17 [43].

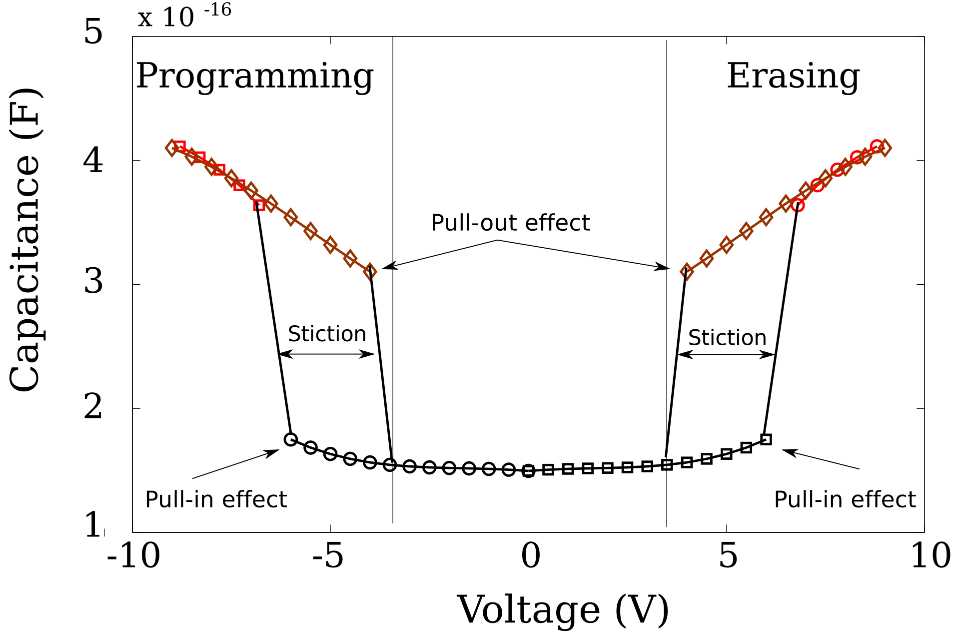


FIGURE 3.15: Programming and erasing sequences performed by the suspended control gate through the bias voltage

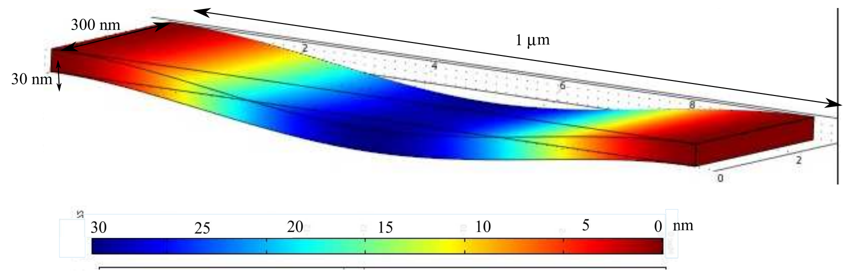


FIGURE 3.16: Transient analysis of the doubly-clamped beam via numerical analysis

3.5 Suspended Control Gate Scaling Down Process

The main aim of this process is to reduce, besides the dimensions of the suspended control gate, the voltage at which the pull-in effect takes place. This analysis is performed as a 3D FEM analysis by using CoventorWare. The scaling down process is defined as the suspended gate parameters divided by the scaling down factor (n) that is expressed as

$$\text{Scaling down} = \frac{\text{SG parameters}}{\text{Scaling Factor (n)}} \tag{3.9}$$

Table 3.5 shows the parameters used within the numerical analysis and the scaling down factor associated for each analysis

The set of D - V curves displayed in Fig. 3.18 obtained from the 3D FEM analysis shows the scaling-down trend obtained by scaling-down the dimensions according to the fixed factor. In here, each of the analyses performed shows that the pull-in voltage is lower

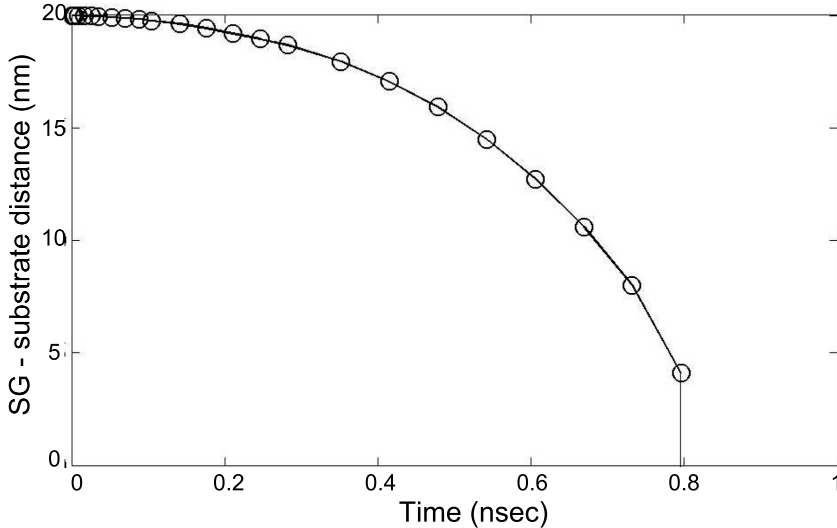


FIGURE 3.17: Characteristic curve of a transient analysis performed to calculate the mechanical switching time (pull-in effect) of the doubly-clamped beam

TABLE 3.5: Scaling-down parameters implemented within de 3D FEM analysis

Analysis	Length nm	Width nm	Thickness nm	Air-Gap nm	Scaling Factor (n)
1	1000	300	30	20	1.0
2	909.1	272.7	27.3	18	1.1
3	769.2	230.8	23.07	15.4	1.3
4	588.2	176.4	17.7	11.8	1.7
5	526.3	157.9	15.7	10.6	1.9
6	434.8	130.5	13.0	8.7	2.3

than that of its predecessor, whose operation voltage was substantially reduced from the typical 6 V for the largest feature to about 3 V for the shortest one. It is foreseen that by increasing the scaling factor, the operation voltage of the memory device can be improved.

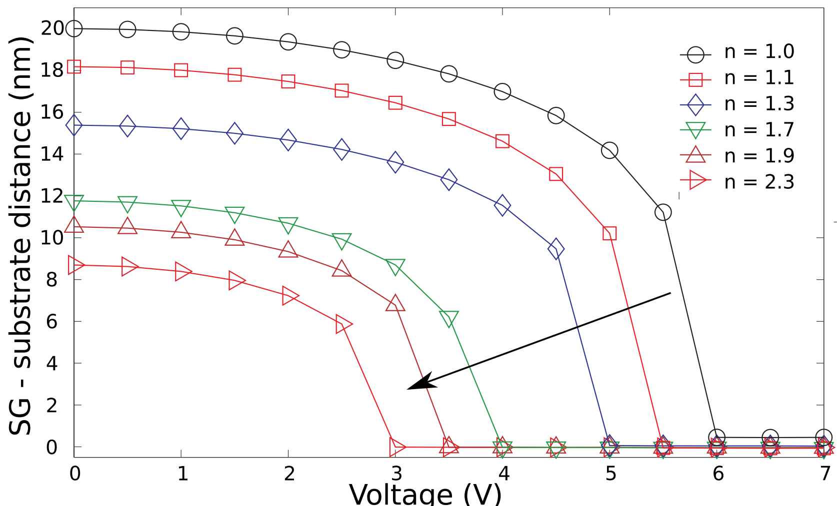


FIGURE 3.18: 3D FEM scaling-down analysis by shrinking the suspended gate parameters according to a scaling factor described in Table 3.5

Chapter 4

Quantum-Mechanical Tunnelling Process: Current Density Analysis

The programming and erasing operations consider the injection of electrons through the tunnel oxide layer. The tunnelling process occurs as a combination generated by the high voltage applied on the control gate electrode and the pull-in effect. From the point of view of the energy band diagram, the tunnel oxide layer bends its original square shape until it becomes triangular allowing the electrons to tunnel through the thinner triangular shape. This tunnelling process is known as the Fowler-Nordheim tunnelling process. In this Chapter, the quantum-mechanical tunnelling mechanism is analysed and implemented by considering the Tsu-Esaki equation within the transfer matrix method as voltage dependent functions. These are implemented within an algorithm in order to find the current density as function of the tunnel oxide material, thickness and applied voltage. Finally, the current density characteristic curve is segmented and transformed from a discreet function to analog voltage dependent functions. These functions are implemented as behavioural models within a commercial circuit simulator coded in an analog hardware description language such as Verilog-AMS.

The injection of electrons to program or to erase the memory node is performed by the dynamic movement of the suspended control gate. Here, the electrons are injected into the memory node via the suspended control gate (top injection), rather than from the channel (bottom injection) as other memory devices *e.g.* flash memory. The tunnelling mechanism starts once the suspended control gate has collapsed on the tunnel oxide layer. To exemplify the process, a set of schematic diagrams that describe the programming and erasing operations of the SGSNM and how the band diagram of the tunnel oxide layer, silicon nanodots and the thick oxide from the substrate change according to each bias voltage, is depicted in Fig. 4.1.

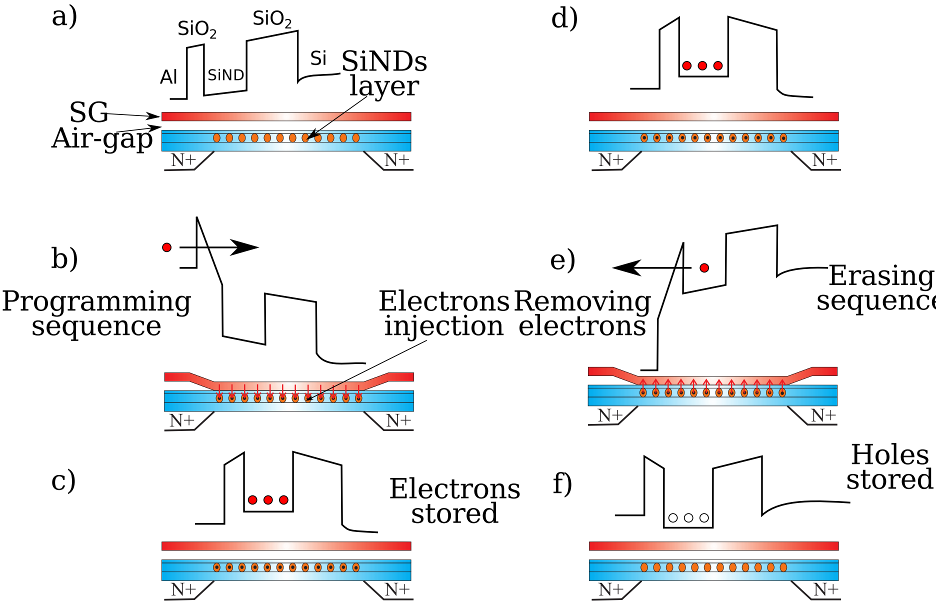


FIGURE 4.1: Programming and erasing sequence according to the SG and tunnelling processes. a) SG without bias, b) a negative voltage is applied on the SG and the electrons tunnelling is allowed because of the triangular shape of the tunnel barrier, c) by reducing the applied voltage the SG returns as well as the tunnel barrier to its initial position, d) considering the SGSNM cell programmed, a positive voltage is applied on the SG (d) and the electrons are removed from the FG (e). Finally, by reducing the applied voltage both the SG and the tunnel barrier return to their initial position (f)

At zero bias, the suspended control gate is isolated from the entire structure and the equivalent energy band diagram of the structure has the shape as depicted in Fig. 4.1a. By applying a negative bias voltage for the programming operation, the suspended gate bends downwards until it collapses on the tunnel oxide layer. The energy band diagram shows how the tunnel oxide potential barrier bends until it becomes triangular allowing the electrons to tunnel through it (Fig. 4.1b). Whilst the applied voltage is reduced, the energy barrier recovers its original squared shape trapping the electrons that firstly tunnel through it into the floating gate. Those trapped electrons within the SiNDs let the memory node become negatively charged as depicted in Fig. 4.1c. In contrast, when a positive bias voltage is applied on the control gate and the suspended control gate snaps on the oxide layer, the energy band diagram shows that the tunnel barrier becomes triangular allowing electrons to tunnel from the floating gate into the suspended gate

as depicted in Fig. 4.1e. By reducing the applied voltage, the suspended gate detaches from the tunnel oxide layer and the energy barrier recovers its initial shape letting the memory node become positively charged as depicted in Fig. 4.1f.

The tunnelling process described in Fig. 4.1 shows the behaviour of the tunnelling mechanism which is expected by co-integrating the transfer matrix method and the Tsu-Esaki equation within a finite element-based homemade algorithm [66, 67, 68, 69, 70]. Within this algorithm the Schrödinger and the Poisson equations are co-solved simultaneously in order to obtain the desired current density curve.

4.1 Tunnelling Current Density Analysis

The model that we are going to implement analyses the tunnel barrier which assumes that the energy and momentum are maintained due to no energy dissipation processes being included. Thus, the total energy of electrons can be divided into lateral and vertical components (x,y & z)

$$E(\vec{k}) = \frac{\hbar^2\left(k_x^2 + k_y^2\right)}{2m^*} + E_z \tag{4.1}$$

where m^* is defined as the effective mass of the electron, $\hbar$ is defined as the Planck constant and $\vec{k}$ represents the lateral wave vector.

The tunnelling current density known as the Tsu-Esaki equation at finite temperature is defined as

$$J = J_{\rightarrow} - J_{\leftarrow} \tag{4.2}$$

where

$$J_{\rightarrow} = 2 \sum_{k_x,k_x,k_x>0} ev_z T(E_z)\left[f_L(\vec{k})[1 - f_R(\vec{k})]\right]$$

$$J_{\rightarrow} = \frac{1}{2\pi^2}\int_0^\infty k_{x,y}dk_{x,y}\int_0^\infty dk_z ev_z T(E_z) f_L(k_{x,y,z})[1 - f_R(k_{x,y,z})] \tag{4.3}$$

$$J_{\leftarrow} = 2 \sum_{k_x,k_x,k_x<0} ev_z T(E_z) f_R(\vec{k})[1 - f_L(\vec{k})]$$

$$J_{\leftarrow} = \frac{1}{2\pi^2}\int_0^\infty k_{x,y}dk_{x,y}\int_{-\infty}^0 dk_z ev_z T(E_z) f_R(k_{x,y,z})[1 - f_L(k_{x,y,z})] \tag{4.4}$$

where $T(E_z)$ defines the transmission probability function, f_L and f_R are the Fermi distribution functions at barrier sides called emitter and collector regions.

$$f_{L,R}(\vec{k}) = \frac{1}{1 + \exp\left(\frac{E(\vec{k}) - E_F^{L,R}}{K_B T}\right)} \tag{4.5}$$

where

$$E_F^L = E_F^R + V$$

where V is defined as an external voltage applied to the barrier, by using

$$v_z = \frac{1}{\hbar}\frac{dE_z}{dk_z} \tag{4.6}$$

Integrating over the regions perpendicular to z

$$J = \int_0^\infty dE_z T(E_z) S(E_z) \tag{4.7}$$

From the above equation, $S(E_z)$ is defined as the electron supply function. This function is defined as

$$S(E_z) = \frac{m^* e k_B T}{2\pi^2 \hbar^3} ln \left[\frac{1 + \exp\left(\frac{1}{k_B T}(E_F^L - E_z)\right)}{1 + \exp\left(\frac{1}{k_B T}(E_F^R - E_z)\right)}\right] \tag{4.8}$$

The Tsu-Esaki equation and the transfer matrix method consider the Schrödinger equation as time independent and in one dimension (1D) as shown in Fig. 4.2.

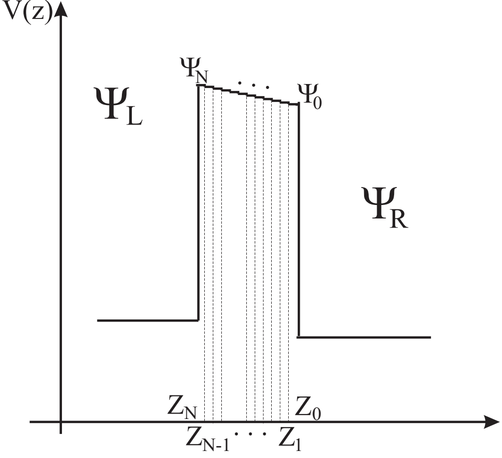


FIGURE 4.2: Small potential steps used for the transfer matrix calculations of a single barrier

$$-\frac{\hbar^2}{2}\nabla\left(\frac{1}{m^*(z)}\nabla\right)\Psi(z) + V(Z)\Psi(z) = E_z\Psi(z) \tag{4.9}$$

where $m^*(z)$ is defined as the z-dependent conduction-band effective mass, $V(z)$ as the potential energy used and $\Psi(z)$ is defined as the wave function. The solution for the wave function has the form of

$$\Psi^{(i)}_{k_z^{(i)}}(z) = A^{(i)}_{k_z^{(i)}} \exp(ik_z^{(i)} z) + B^{(i)}_{k_z^{(i)}} \exp(-ik_z^{(i)} z) \tag{4.10}$$

In Eq. (4.10), $k_z^{(i)}$ is defined as the complex wave number that is described as

$$k_z^{(i)} = \frac{\sqrt{2m^{*(i)}(E_z - V^{(i)})}}{\hbar} \tag{4.11}$$

The solution of the above coefficients are related as

$$\begin{pmatrix} A^{(i+1)}_{K_z^{(i+1)}} \\ B^{(i+1)}_{K_z^{(i+1)}} \end{pmatrix} = T^{(i)} \begin{pmatrix} A^{(i)}_{K_z^{(i)}} \\ B^{(i)}_{K_z^{(i)}} \end{pmatrix} \tag{4.12}$$

The matrix $T^{(i)}$ is defined as

$$T^{(i)} = \begin{pmatrix} \alpha_+^{(i)} P & \alpha_-^{(i)}/Q \\ \alpha_-^{(i)} Q & \alpha_+^{(i)}/P \end{pmatrix}$$

where

$$\alpha_\pm^{(i)} = \frac{1}{2}\left[1 \pm (\frac{m^{*(i+1)}}{m^{*(i)}})(\frac{k_z^{(i)}}{k_z^{(i+1)}})\right] \tag{4.13}$$

$$P = \exp[i(k_z^{(i)} - k_z^{(i+1)})z_{i+1}] \tag{4.14}$$

$$Q = \exp[i(k_z^{(i)} + k_z^{(i+1)})z_{i+1}] \tag{4.15}$$

$$T_{SB}(E_z) = \left[1 + \frac{V^2}{4E_z(V - E_z)} \sinh^2(\kappa_z d)\right]^{-1} \tag{4.16}$$

where

$$k_z = \sqrt{\frac{2m^* E_z}{\hbar^2}} \quad \& \quad \kappa_z = \sqrt{\frac{2m^*(V - E_z)}{\hbar^2}}$$

The above set of equations was implemented to calculate the current density characteristic curve. Taking into account the complexity of the algorithm due to the set of equations that it encompasses, the code is written in a language created for this purpose, Fortran [71].

4.2 Tunnel Barrier Analysis

The tunnelling process within the SGSNM is modelled by using a variable resistor in parallel with a tunnel capacitor. The tunnel capacitor within the model represents the memory node and the tunnel resistor describes how the current increases or decreases as a function of the applied voltage. Several parameters such as materials, number of nodes, thicknesses and other physical properties are required to represent those elements within the tunnel barrier analysis. The model is represented schematically as depicted in Fig. 4.3.

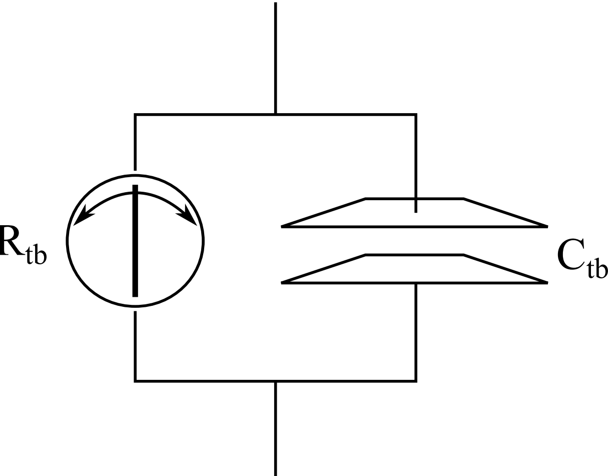


FIGURE 4.3: Tunnel barrier equivalent circuit. Where C_{tb} is the tunnelling capacitance and R_{tb} is the tunnel variable resistance

The current density analysis takes into account several parameters for both elements. Most of those parameters are physical properties such as material, electron effective mass of the material, band gap, dielectric constant, doping concentration of the emitter and collector and thickness. Within the FEM kernel other numerical parameters such as number of lines or nodes are required [66]. Those sets of parameters are coupled to the electromechanical characteristics of the suspended control gate (D - V and C - V hysteresis cycles) analysed in the previous Chapter. Once the whole set of parameters has been implemented within the algorithm and by considering that the tunnel oxide layer is SiO_2, the numerical analysis takes place in which five different thicknesses from 4 to 8 nm are analysed. The characteristic curves obtained from those analyses are depicted in Fig. 4.4.

The characteristic curves obtained from the 5 SiO_2 layers show that the curves at low

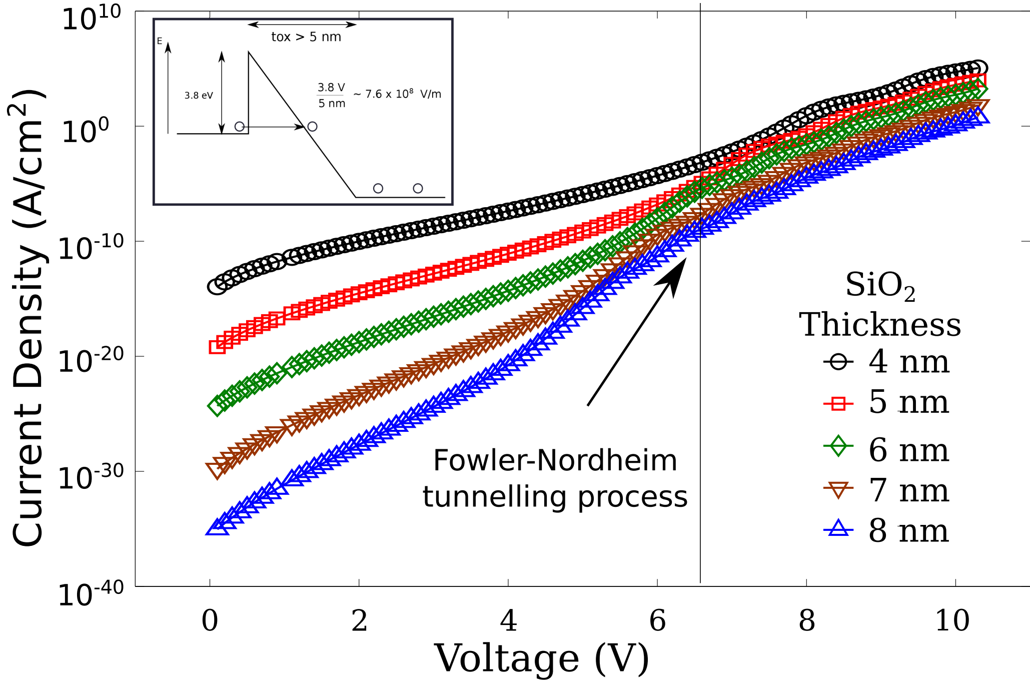


FIGURE 4.4: Current density-voltage set of curves for several SiO_2 thicknesses. In here the Fowler-Nordheim tunnelling equation and current are specified for a 5 nm SiO_2 layer

bias voltage from 0 to 2 V, the current density is rather low in which the maximum difference between each thickness is about 5-magnitude order. Whilst the applied voltage increases from 2 to 7 V, the current density also increases and the difference between each thickness reduces to three magnitude orders as it is displayed at pull-in and onward voltages. Further increment in the applied voltage results in fairly large currents that eventually reach the saturation point over the 18 V. Analysing the above set of curves, it is possible to see that at low voltages, the tunnelling mechanism is governed by direct tunnelling when the dielectric thickness analysed is rather thin <4 nm and by the Fowler-Nordheim tunnelling at high voltages and rather thick dielectric layers. Therefore, the thinner the dielectric layer, the larger the current density present is and as a consequence, the faster the programming and erasing operations occur. Table 4.1 lists the current density at low and high voltages for each one of the SiO_2 thicknesses. Here, Table 4.1 shows that whole layers at low bias ($\approx$ 0.1 V) have a rather negligible current density *e.g.* 8 nm at 9.24×10^{-36} A/cm^2 while at high voltage ($\approx$ 10 V) the current density reaches a rather high current *e.g.* 4 nm at 506 A/cm^2.

According to the set of curves and parameters displayed in Fig. 4.4 and Table 4.1 and by considering the minimum tunnel oxide layer for the scaled flash memory is 7 nm defined elsewhere, the value for the tunnel oxide layer hereafter adopted for the SGSNM analysis is 7 nm SiO_2 [1]. Once J - V characteristic curve for 7 nm has been obtained, it is transformed into an analog model in AHDL in the same manner as for the C - V hysteresis cycle. The implementation of the current density curve requires

Table 4.1: Comparison between voltage and current densities currents

Thickness (nm)	Voltage min (V)	Current Density (A/cm^2)	Voltage max (V)	Current Density (A/cm^2)
4	0.1	1.12×10^{-14}	10	506
5	0.1	6.26×10^{-20}	10	42.3
6	0.1	4.86×10^{-25}	10	5.86
7	0.1	1.66×10^{-30}	10	0.163
8	0.1	9.24×10^{-36}	10	9.23×10^{-3}

the specification of the boundary conditions in which those functions are going to take place. Considering that the tunnelling process takes place after the pull-in occurs until the pull-out effect takes place, outside of these boundaries, the tunnelling process is negligible due to the isolation in which the floating gate is immersed. As a consequence, this isolation avoids any leakage current into the substrate due to the thick oxide layer or towards the control gate due to the doubly isolation existent produced by the oxide layer and air-gap to the control gate. Figure 4.5 shows the current density curve segmented into two regions.

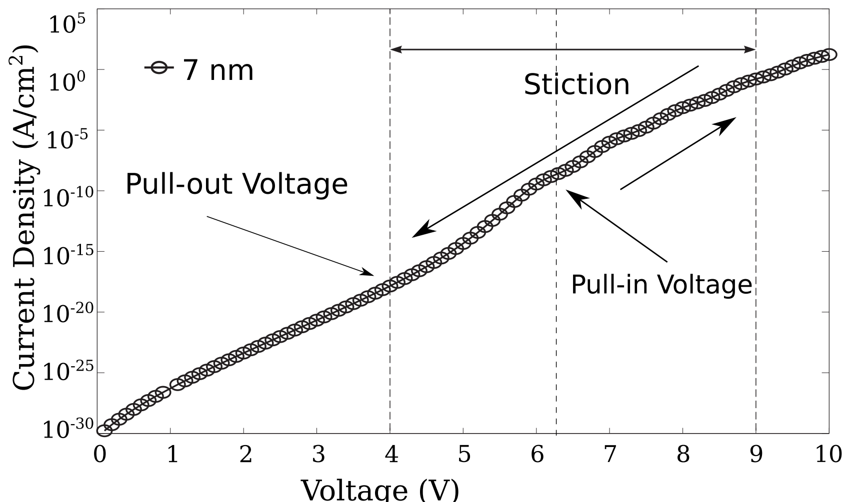


Figure 4.5: Full current density-voltage characteristic curve for a SiO_2 layer at 7 nm thickness

The implementation of the analog voltage functions requires the manipulation of the current density curve. As the curve is defined in a logarithmic scale, the first step is to transform the logarithmic scale into a linear function through mathematical manipulation. Once the function has been transformed, an algebraic algorithm such as polynomial degree 4, is used to approximate the linear function. The approximation curve and the linearised function for the pull-in section are depicted in Fig. 4.6. Here, the discreet curve is represented as the black line while the approximation by the polynomial algorithm is depicted as the red line.

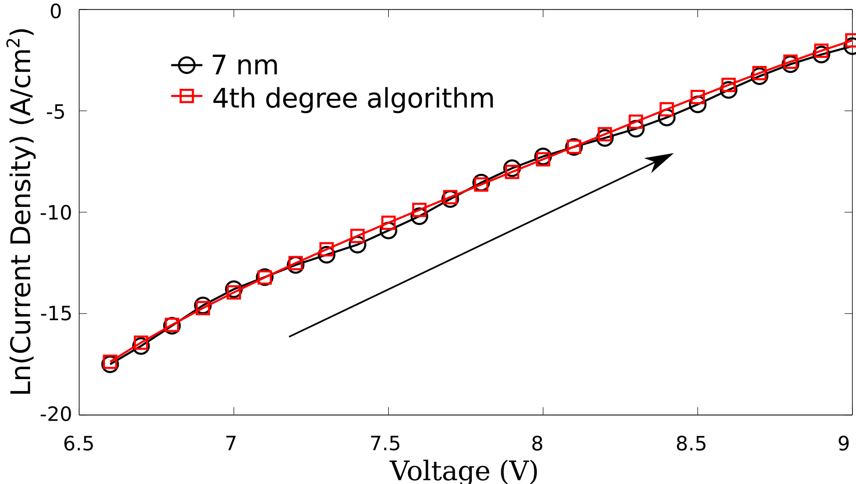


FIGURE 4.6: Current density-voltage characteristic curve for a sweeping pull-in voltage (6.7 to 9 V)

The stiction window (max. voltage to pull-out), is transformed into an analog voltage dependent function as depicted in Fig. 4.7.

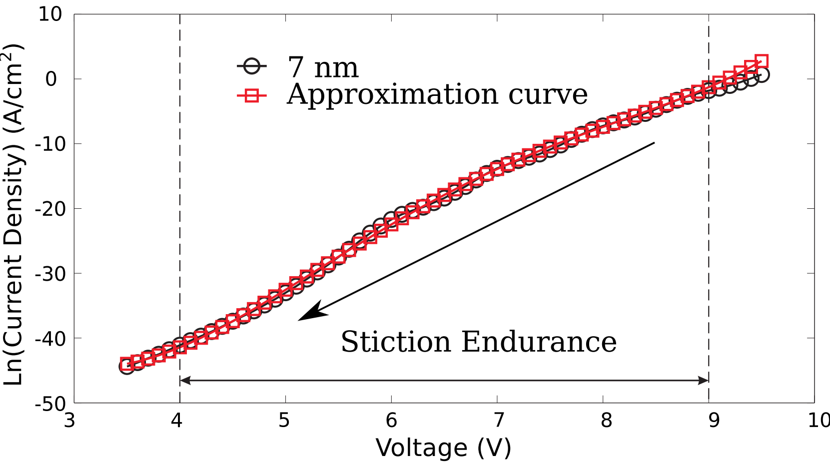


FIGURE 4.7: Current density-voltage pull-in voltage approximation. Circles represent the pull-in effect and the squares represent the approximation function

The mathematical implementation and approximation of those curves are represented for the polynomial equation of the form

$$f(x) = \mathrm{M}_0x^0 + \mathrm{M}_1x^1 + \mathrm{M}_2x^2 + \mathrm{M}_3x^3 + \mathrm{M}_4x^4$$

where the coefficients of the equation are defined as shown in Table 4.2.

Once the pull-in voltage and the stiction characteristic curves have been transformed as voltage dependent functions, these curves are implemented as mathematical functions that describe the programming and erasing operations. These functions are implemented within the analog hardware description language by following the form

TABLE 4.2: Coefficients used to approximate the pull-in and stiction curves as function of an algebraic polynomial algorithm for the erasing process

Function	E_0	E_1	E_2	E_3	E_4
$V_{Pull-in}$	-1316.2	633.95	-117.4	9.7618	-0.3045
$V_{Stiction}$	66.547	-88.631	23.829	-2.4718	0.0917

$$V_{Pull-in,Stiction}(x) = \pm \exp\left(\mathrm{E}_0 x^0 + \mathrm{E}_1 x^1 + \mathrm{E}_2 x^2 + \mathrm{E}_3 x^3 + \mathrm{E}_4 x^4\right) \tag{4.17}$$

Equation (4.17) defines the erasing process as a function of the coefficients specified in Table 4.2. Here, the functions are defined as an exponential function because of the nature of the current density obtained from numerical analysis. Figure 4.8 displays the curves for the pull-in and stiction window transformed as analog voltage dependent functions for the erasing sequence.

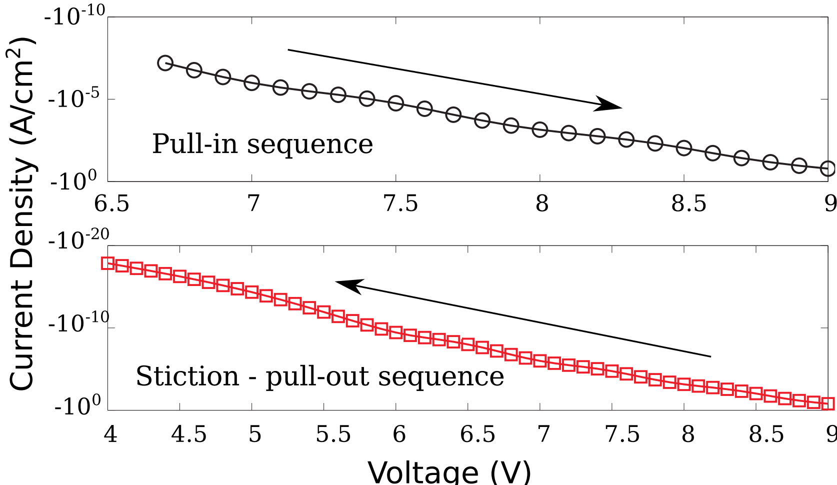


FIGURE 4.8: Current density-voltage curves for the erasing (a) pull-in and (b) stiction - pull-out sequence

The programming sequence considers different approaches as for the erasing one. Here, the function that is used to implement the voltage dependent functions is described as

$$V_{Pull-in,Stiction}(x) = \exp\left(\mathrm{P}_0 x^0 + \mathrm{P}_1 x^1 + \mathrm{P}_2 x^2 + \mathrm{P}_3 x^3 + \mathrm{P}_4 x^4\right) \tag{4.18}$$

where P are the coefficients that define those expressions. The coefficients that drive the above equation are defined in Table 4.3

Figure 4.9 shows the voltage dependent functions that describe the programming sequence.

The implementation of the analog voltage dependent functions as behavioural models within the SGSNM circuit are performed by using a commercial circuit simulation

TABLE 4.3: Coefficients used to approximate the pull-in and stiction curves as a function of an algebraic polynomial algorithm for the programming process

Function	P_0	P_1	P_2	P_3	P_4
$V_{Pull-in}$	-1316.2	-633.95	-117.4	-9.7618	-0.3045
$V_{Stiction}$	66.547	88.631	23.829	2.4718	0.0917

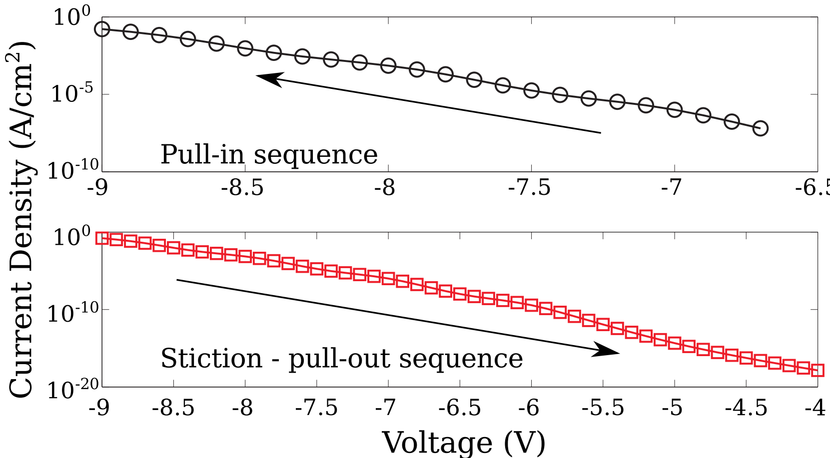


FIGURE 4.9: Current density-voltage curves for the programming (a) pull-in (b) stiction - pull-out sequence

software such as SmartSpice or Cadence that uses PSpice as the main kernel. The implementation of those voltage dependent functions coded in Verilog-AMS, allows the simulator to interpret the models as libraries and use them as an instance of the main library whenever those models might be required.

Chapter 5

Suspended Gate Silicon Nanodots Memory implementation

The electromechanical analysis of the suspended control gate describes the voltages at which the suspended gate collapses (pull-in) and detaches (pull-out) from the programming and erasing operations. The numerical analysis was conducted considering a doubly-clamped aluminium beam of 30 nm thick at a 20 nm air-gap coupled to a thin tunnel oxide layer. The hysteresis curves that describe the full mechanical and electromechanical behaviour of the suspended control gate were obtained from the 3D numerical analysis and were implemented as voltage dependent functions by using an analog hardware description language such as Verilog-AMS. Another key point within the programming and erasing operations was the analysis and implementation of the tunnelling process as a behavioural model. In order to calculate the current density through the tunnel oxide layer, an analysis of this layer was conducted by using a home-made code which encompasses the Tsu-Esaki equation embedded in the transfer matrix method. As a result of this implementation, the current density through a set of SiO_2 layers (4 to 8 nm) and in particular through a 7 nm SiO_2 layer was successfully obtained and implemented as voltage dependent functions. In this Chapter we carry out the circuital analysis of the SGSNM cell in which the behavioural models above mentioned are implemented as libraries within the circuit simulator described by a general algorithm. As a result of the implementation and simulation of the memory cell, the curves for each one of the nodes such as the control gate, memory node and readout element are obtained. By analysing the memory node, the non-volatility of the SGSNM is successfully demonstrated. Further analyses are performed in which the programming and erasing speeds, power consumption and influence of the short range forces are analysed.

The implementation of the behavioural models within the analog hardware description language requires the specification of the ranges on which those models have influence and in which those elements are negligible. The electromechanical behaviour of the suspended control gate is analysed through the C - V hysteresis cycle that represents the three regions in which the pull-in, stiction and pull-out processes take place. Those regions are driven and limited by the control voltage. Similarly, the tunnelling process is defined by the voltage at which the tunnelling mechanism takes place. Here, the tunnelling process is modelled as a variable resistor that depends on the applied voltage which is driven by the control gate. Each of those models is represented schematically as a function of the control voltage as depicted in Fig. 5.1.

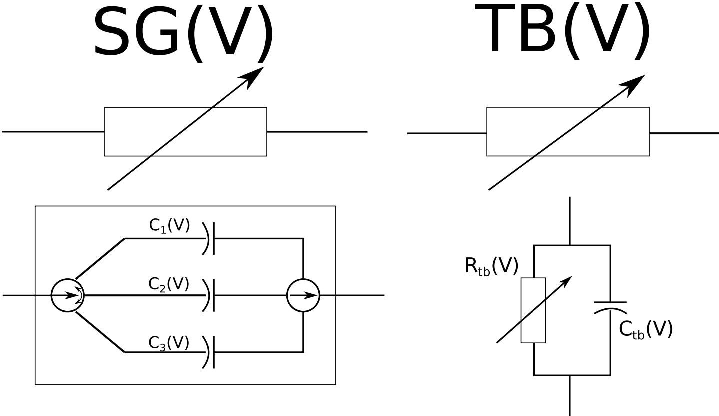


FIGURE 5.1: Schematic diagrams that represent the suspended control gate and the tunnel oxide layer, a) shows the suspended gate pull-in, stiction and pull-out effects represented for three capacitors and b) shows the tunnel oxide layer as a variable resistor and a tunnel capacitor both in parallel

The implementation of the behavioural models co-integrated within a hybrid entity requires a fine control over each one of the processes that take place at a particular voltage. This is why we are proposing an algorithm that controls the behavioural models as well as the boundaries for each of the processes. Initially, the set of libraries that drives the flags at which the functions are going to take place is settled as Table 5.1 states. Once settled, the body of the algorithm is divided into five sequential processes as follows:

1. Initial value
2. Selection of the process for programming or erasing
3. Localise if the process is going upwards or downwards
4. Encodes towards the correct pipeline
5. Send the value to the output

Figure 5.2 shows the flow diagram that encompasses these requirements.

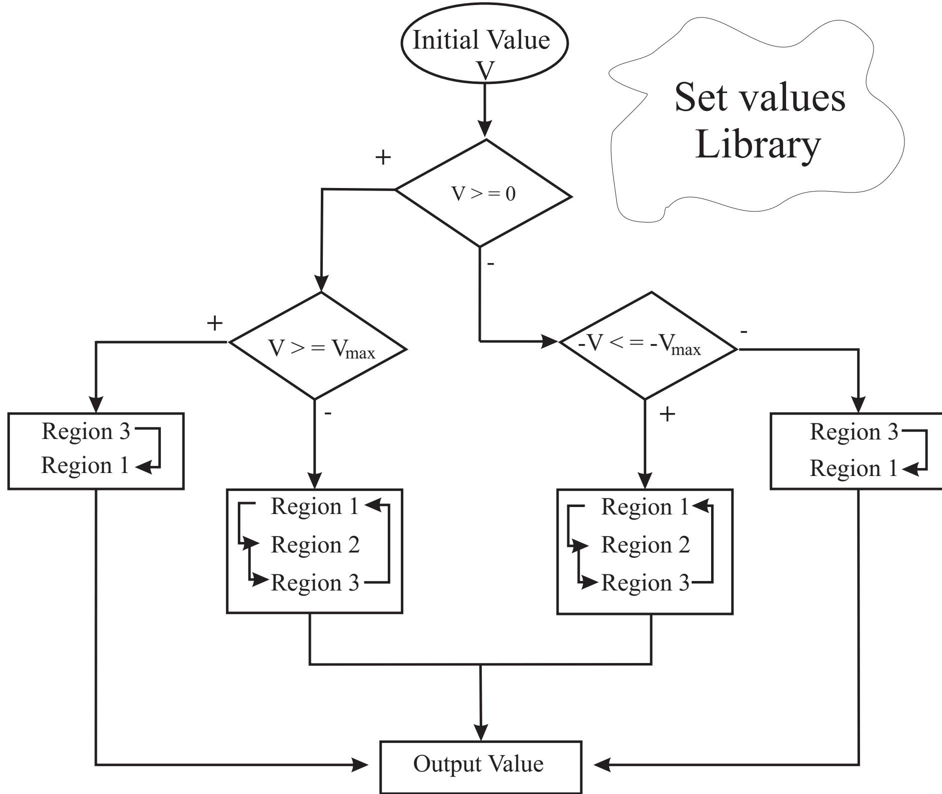


FIGURE 5.2: Flow Diagram representation for the C - V and J - V hysteresis cycle

Initial value At this stage, the algorithm clears whole variables and settles them to a known value. The set of flags that intervene is initialised clearing the post- and pre-processes in the pipeline. The first set of data is allowed, read and encoded towards the next stage.

Discriminate process Here the set of data is compared and the flags are modified as functions of the value (positive or negative). If the initial value is negative, it indicates that the process is referring to programming and as a consequence, a set of flags is activated to allow the programming operation. In contrast, if the initial value is positive, the action that is going to take place is the erasing process. Once the data are analysed and discriminated, they are moved on to the next level.

Identification segment According to the set of flags stated by the inner data and in the previous stage, if the inner value analysed is found to be out of the boundaries pre-settled on the master library, then the information is neglected and the flag that indicates that useless information has been fed is activated, which clears all flags and resets the process to allow a new set of data. Alternatively, if the value is within the pre-settled limits, then the value is mapped and directed to the next place in the algorithm ladder.

Encoding towards the correct pipeline Here, the numerical value is analysed and sent towards the mathematical function. This function is going to be used within the specific segment of the algorithm in which this value is mapped and selected. Once the value has been analysed, it is sent towards the last step. However, if the value analysed is the maximum value, then it has the maximum priority and overrides whole flags, with the exception of the minimum voltage or pull-out.

Output value After being analysed, the value is directed to the output section in which it is conditioned to interact within the commercial software. Also, at this point, the flags activated on the identification segment are cleared and the variables used within this analysis are set to a known value. The flags are mirrored to the library settled and in the case of a misplaced value, these are re-initiated.

Table 5.1: Parameters used to control the SGSNM sequences, as function of the inner voltage

Process	Set Values (V)	Flag. Erase	Flag Max. V	Flag P-out
Pull-in	$\mp$ 6.7	$\times$	$\surd$	$\times$
Pull-out	$\mp$ 4	$\surd$	$\times$	$\surd$
Stiction	$\mp$ 9 - 4	$\surd$	$\times$	$\surd$

Once the values are analysed, the results are stored in an output register as a stack in which the values are used as an internal sequence counter and as a buffer that feeds the next part of the circuit within the commercial software. The model that encompasses the algorithm within the commercial circuit simulator is displayed as a schematic diagram (Fig. 5.3).

5.1 Electric Circuit Analysis of the SGSNM cell

The electric circuit analysis of the Suspended Gate Silicon Nanodot Memory includes the co-integration of the behavioural models with the readout element which will display the state of the memory node at reading operations, considering the fact that modern circuit simulators have a broad range of libraries that include different technologies and models for MOS transistors. The transistor model that is considered as a readout element within the electric circuit analysis for the SGSNM is based on the BSIM3v3 model. The BSIM3 model is a PSpice model for CMOS technology which is physics-based, accurate, scalable, robust and predictive [55, 56]. Once the above elements have been implemented as depicted in the schematic diagram shown in Fig. 5.4, the SGSNM cell is analysed.

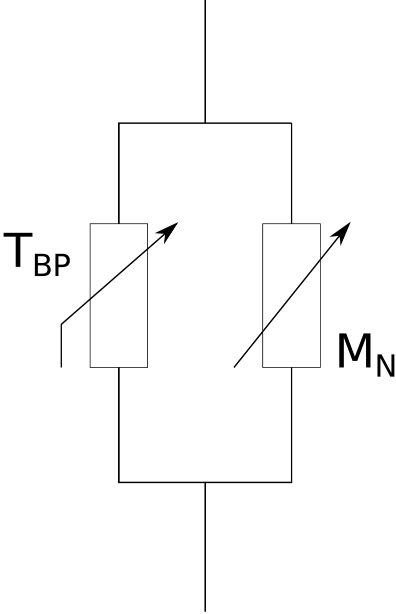


FIGURE 5.3: Schematic diagram that represents the algorithm implemented as an analog model by using Verilog-AMS, in which T_{BP} is the tunnel barrier potential and M_N describes the memory node

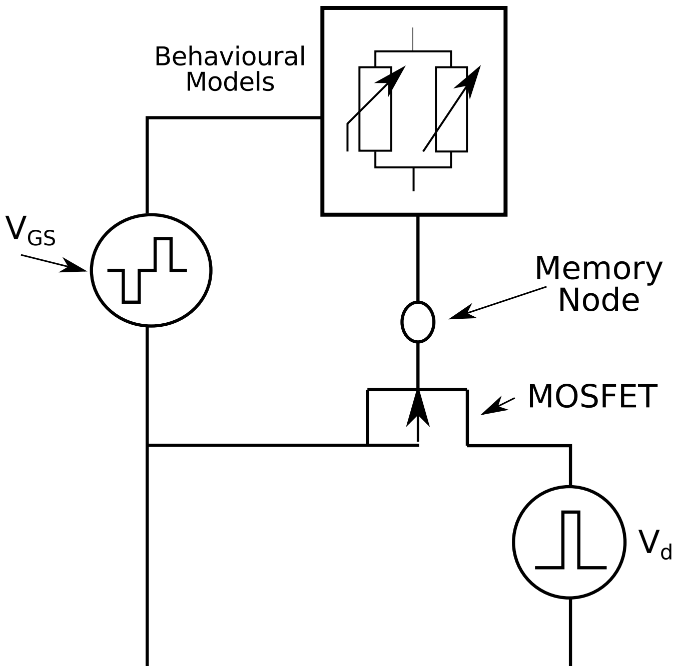


FIGURE 5.4: Schematic diagram for the suspended gate silicon nanodots memory

In the above diagram, the memory node is a virtual point where the electrons will be either stored or removed according to the control voltage. The control signal is defined as a bi-stable voltage source in which each section is fully controlled by a voltage signal as a function of time. This signal drives the slope that it uses to raise and lower/fall (slew-rate) the control signal. It also considers the high/low voltage levels and the repetition cycle of the entire process [56, 57]. The drain signal is generated

through a pulse voltage source built-in within the circuit simulator. The functions that process the programming and erasing operations are isolated from the reading process. This is why, when the programming or erasing operations take place, the readout element is deactivated. In contrast, when the memory node is read, the control voltage is deactivated and the readout element is biased which "reads" the memory node. Figure 5.5 shows the configuration of the SGSNM cell for programming and erasing as well as for reading operations.

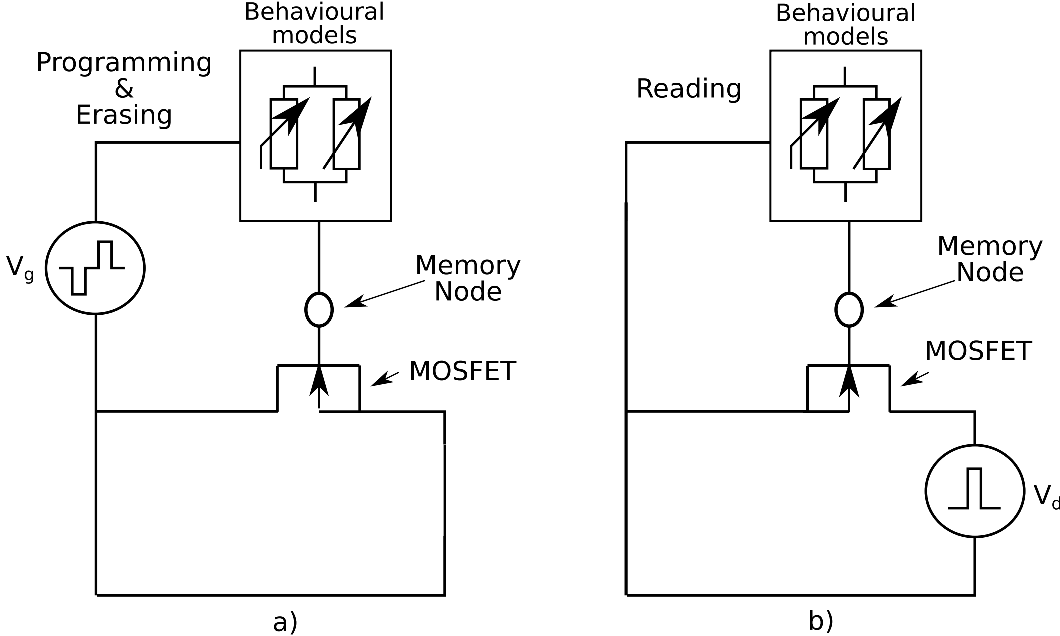


FIGURE 5.5: SGSNM schematic models for a) programming, erasing and b) reading processes

The times and voltages required to control the Suspended Gate Silicon Nanodot Memory cell, listed in Table 5.2, are implemented within the voltage sources in the commercial circuit simulator. The control voltage source uses a piece-wise linear source. The associated control voltage and the time that it takes to program, read or erase the memory cell, are co-dependently integrated within the source. In the same manner, the voltage source used to control the readout element is only active when the control source is not biased (0 V), otherwise the source is switched-off. In the pulse source the maximum voltage, dead-times (0 V) as well as rise, high and fall times are also specified.

TABLE 5.2: Control parameters used to define the voltage sources in the circuital analysis

Process	Time rise (sec)	Time fall (sec)	Time high (sec)	Period (sec)
Drain Voltage	1×10^{-12}	1×10^{-12}	1×10^{-9}	10.2×10^{-9}
Control Voltage	0.1×10^{-9}	0.1×10^{-9}	5×10^{-9}	-

By using the set of parameters depicted in Table 5.2, these are implemented within the

commercial circuit simulator to test the control as well as the drain voltage sources. The signals obtained for each one of those voltage sources are displayed in Fig. 5.6.

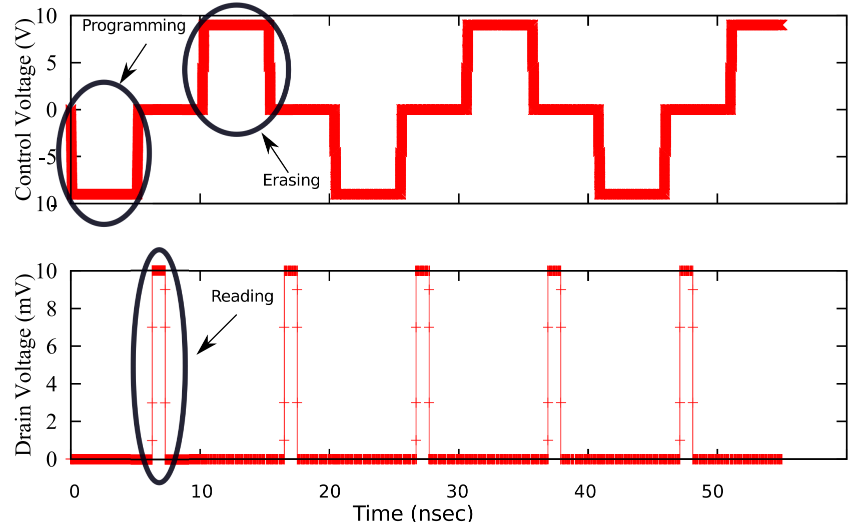


FIGURE 5.6: Programming, reading and erasing voltage signals for the bi-stable piece-wise linear control source and the pulse signal used in the readout element activation

A key parameter to consider within the circuit simulation is the power consumption by the readout element device. Within the commercial circuit simulator that is used for this analysis (SmartSpice) within the BSIM3V3 family, there is a model that corresponds to the level-81 family which fulfils the high-speed and low-power consumption requirements [64, 72, 73]. Once the readout element has been carefully chosen, the memory cell is implemented within the circuit simulator as a netlist in which it calls the behavioural models as libraries within the circuit simulator. As a result of this analog circuit simulation of the SGSNM memory cell, the signals for each node depicted in Fig. 5.4 are successfully obtained.

The electric circuit simulation depicted in Fig. 5.7 displays the signals for each node of the Suspended Gate Silicon Nanodot Memory cell. The chart is divided into four operations: programming (P), reading the memory state (R_{Prog}), erasing (E) and reading the memory (R_{Erase}) state. The programming sequence starts by applying a negative bias voltage on the suspended gate electrode as displayed in Fig. 5.7a. As previously described, the readout voltage (V_{drain}) is settled to 0 V (5.7b). Meanwhile in the memory node it shows how the floating gate is being charged by the injection of electrons that are tunnelling through the tunnel oxide layer into the memory node until the pulse finishes as shown in Fig. 5.7c. It is required to pointed out that the noise present in the memory node when the programming and erasing cycles of the control voltage are activated has a strong relationship with the defects created within the tunnel oxide layer [74, 75]. Given that the readout element has not been activated, there is no reason to look at the device. For the reading process, the control voltage is settled at 0 V and the readout element is activated. To activate the readout element, the drain source displays a rather short voltage pulse as depicted in Fig. 5.7b. Once the programming operation

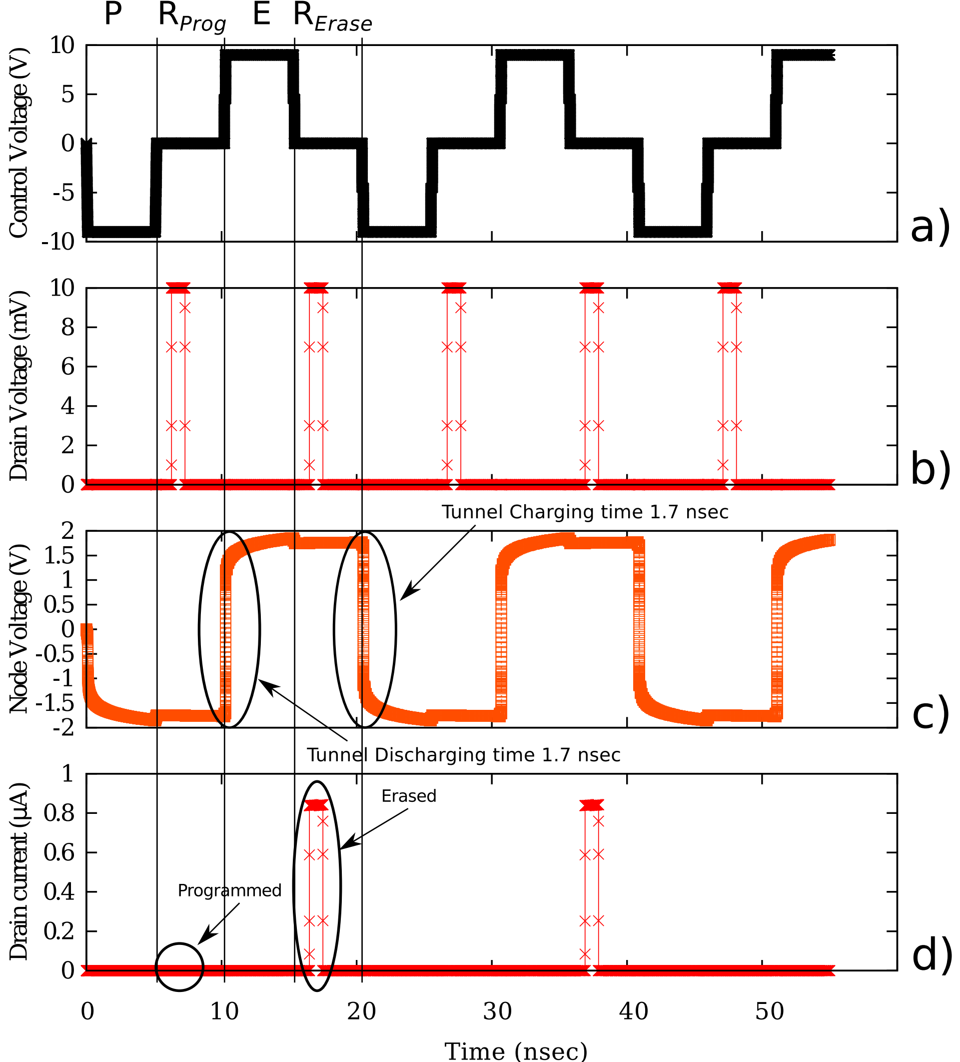


FIGURE 5.7: Set of signals obtained for each one of the SGSNM cell nodes a) suspended control gate, b) readout pulse signal, c) memory node signals in which the non-volatility of the SGSNM is displayed and d) once the readout element is activated, it displays a current pulse when the memory node is positively charged and no signal when the memory node is negatively charged

has finished, the memory node shows a negative DC plateau, which indicates that the memory node has been negatively charged. When the readout element is activated and it reads the memory node, it does not display any current pulse, which indicates that the memory node is negatively charged and, as a consequence, the SGSNM has been programmed as depicted in Fig. 5.7d. On the other hand, to erase the memory cell a positive bias voltage is applied on the suspended gate electrode (Fig. 5.7a). When the memory node is scrutinised, it shows how the electrons are tunnelling from the floating gate into the control gate leaving the memory node positively charged. After the erasing operation has finished and the reading process starts, the memory node displays a positive DC plateau. The readout signal displays a rather short current pulse. This current pulse indicates that there are no electrons stored within the floating gate, which means that the memory node has been erased (Fig. 5.7d). Considering these operations as a

sequence, the programming, erasing and reading operations continue indefinitely.

The co-integration of the behavioural models with CMOS technology in a hybrid circuit simulation proves successful the non-volatility characteristic of the SGSNM cell as well as its fast charging/discharging time ($\approx$ 1.7 nsec). Figures 5.7a & b show the signals that control the circuital analysis such as control (V_{SG}) and drain (V_d) voltage signals and, as a result, the signals obtained depicted in Fig. 5.7c & d such as memory node voltage (V_{Node}) and readout pulse current (I_{drain}). The charging/discharging times in combination with the suspended gate transient analysis ($\approx$ 0.8 nsec) described elsewhere show that the time that the SGSNM cell takes to be either programmed or erased is 2.5 nsec. The programming and erasing times that the SGSNM takes is, comparatively, at least 3-magnitude order smaller than nowadays flash memory. Considering those characteristics, the SGSNM cell shows the qualities and characteristics that are required for it to be the non-volatile RAM memory.

5.2 Charging Time Analysis

As shown in Chapter 3, the operation voltage of the memory device as well as its dimensions can be improved by applying a proper scaling-down factor. Nevertheless, not only can the dimensions of the memory device be improved, but also the charging/discharging time. To improve these characteristics, a few parameters within the original suspended gate structure such as area, thickness and the suspended gate - air-gap ratio are kept fixed whilst the tunnel oxide layer is varied. Taking into account that the current density increases as the tunnel oxide decreases, a set of thicknesses that represents the tunnel oxide layer is analysed within the circuit simulation. The analysis considers the thickness to be of 4 to 10 nm. The programming and erasing times obtained from this analysis as function of the tunnel oxide layer for the charging/discharging times are depicted in Fig. 5.8.

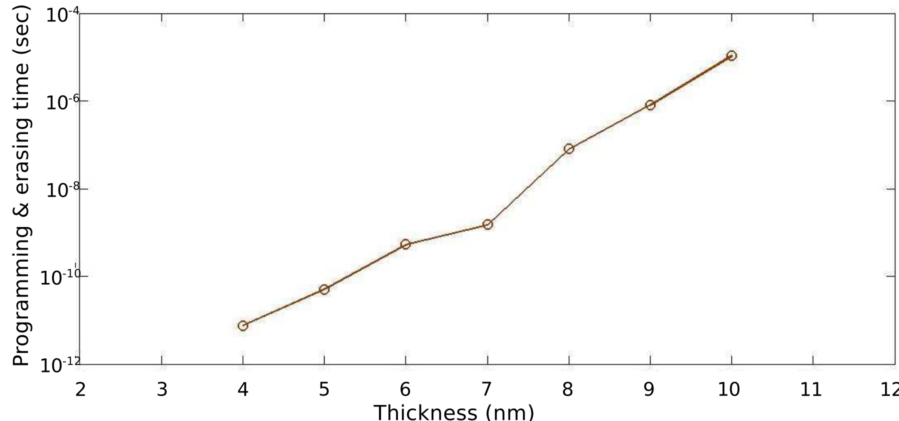


FIGURE 5.8: Charging/discharging time as function of the tunnel oxide layer thickness

This analysis shows that by maintaining a few fixed parameters, the charging/discharging time in the memory node is improved by a few tens of nanoseconds whilst the tunnel oxide layer becomes thinner (4 nm). Based on the programming and erasing time improvement and by properly scaling-down the suspended control gate structure, it is viable to reach even higher programming and erasing times. Figure 5.9 shows the programming and erasing times by combining the mechanical switching time and the charging time as function of the tunnel oxide thickness.

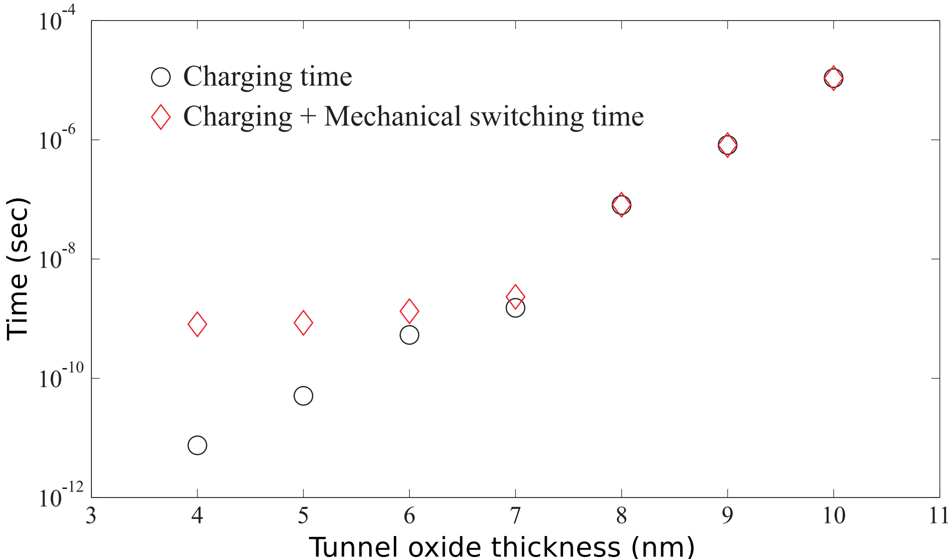


FIGURE 5.9: Programming and erasing time obtained as a combination of the mechanical switching time and the charging/discharging time as function of the tunnel oxide thickness

5.3 Energy analysis

The energy required to move the suspended control gate is a combination of several kinds of energies such as elastic energy E_m, kinetic energy E_k, damping loss energy E_d, electrostatic energy E_e and charging loss energy E_R [76]. These are defined as

$$E_m = \int_V \frac{1}{2}\epsilon\sigma dV \tag{5.1}$$

$$E_k = \int_V \frac{1}{2}\rho v^2 dV \tag{5.2}$$

$$E_e = \int_V \frac{1}{2}EDdV \tag{5.3}$$

$$E_d = \int_V \int_0^t F_d v dt dV \tag{5.4}$$

$$E_R = \int_0^t dE_e(t)dt \tag{5.5}$$

where ϵ and σ are defined as the stress and strain in the beam, ρ is the mass density, v is defined as the velocity, E & D are the electric and the displacement fields and F_d is the mechanical damping force. By implementing these energy equations within a 3D FEM analysis, it was found that the energy required to pull-in the suspended control gate is governed by just three energies, the electrostatic, mechanical and kinetic energies due to the others being rather small, which is why these do not contribute to the final energy consumption value. According to the transient analysis performed earlier, the suspended control gate shows that it pulls-in at 0.8 nsec, hence the associated switching energy at pull-in time: ≈ 8.21 fJ. The total energy required to move the suspended control gate is lower than the energy required for the self-buckled memory due to the SGSNM using a softer beam (aluminium) rather than a stiffer one (poly-Si) as for the other memory device [76]. The electrostatic, mechanic and kinetic energy curves are displayed in Fig. 5.10.

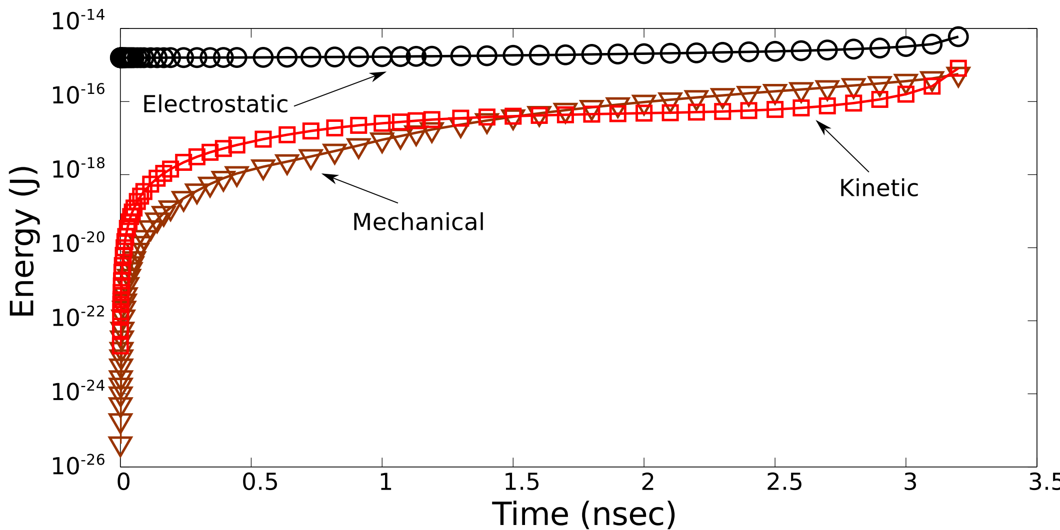


Figure 5.10: Energy consumption analysis in which the mechanical, electrostatic and kinetic as the principal energies

5.4 Short-Range Forces Analysis

The short range forces such as Casimir and van der Waals interact strongly at short distances, in particular when both layers *e.g.* suspended gate and tunnel oxide layer are in contact. The equation (3.7) shows the implementation of short-range forces as part of the pull-out analysis in which the Hamaker constant is introduced as part of the van der Waals force. The Hamaker constant defines the interaction between two or more molecules within a media [65]. In order to calculate the Hamaker constant for

the interaction between the suspended control gate/air-gap/tunnel oxide layer, a few parameters need to be analysed to obtain an accurate value. If the structure under analysis (three layer structure) is considered at an atomic level (see Fig. 5.11), the interaction between atoms creates multiple reflections as well as additivity, *i.e.* the influence of A induces a dipole in atom B parallel to the field of atom A polarises also another atom C. The induced dipole in atom C has an influence on B, as a consequence the field of A has a direct influence on B and as an effect of the reflection C is affected. To surpass the additivity issue, a continuous theory such as Lifshitz theory based on pairwise integration that neglects the atomic structure is used [37, 49]

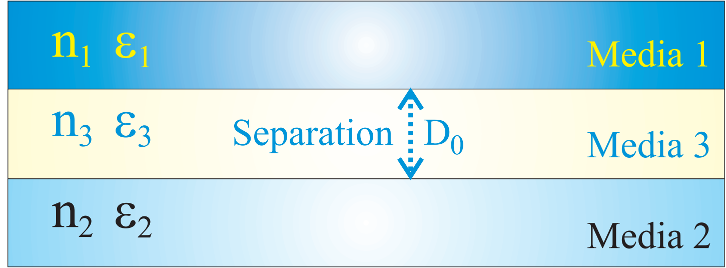


FIGURE 5.11: Schematic representation used to calculate the Hamaker constant by using the Lifshitz model

$$A_{132} = \frac{3kT}{2}\left(\frac{\epsilon_1 - \epsilon_3}{\epsilon_1 + \epsilon_3}\right)\left(\frac{\epsilon_2 - \epsilon_3}{\epsilon_2 + \epsilon_3}\right)\dots$$
$$\dots\frac{3\hbar\nu_e}{8\sqrt{2}}\frac{(n_1^2 - n_2^2)(n_2^2 - n_3^2)}{(\sqrt{n_1^2 + n_2^2}\sqrt{n_2^2 + n_3^2}))(\sqrt{n_1^2 + n_2^2} + \sqrt{n_2^2 + n_3^2})} \quad (5.6)$$

where ν_e is defined as the absorption frequency, n and ϵ are the refractive index and the permittivity for Aluminium, Air and SiO_2, respectively used in this analysis, k is the Boltzmann constant and T is the temperature in K. Table 5.3 shows the refractive index and permittivities of the materials under test.

TABLE 5.3: Set of parameters implemented within the Lifshitz model to obtain the Hamaker constant

Material	Refractive Index	Permittivity
Aluminium	1.43031	-56.149
Air	1.000273	1.0
SiO_2	1.54427	3.9

According to the Lifshitz model, the value obtained for the set of Aluminium/air/SiO_2 is 1.343×10^{-19} J. Compared with the electrostatic force van der Waals and Casimir forces are predominant at short-range. The definition of the van der Waals forces for two flat bodies is defined as [37, 48, 65, 77, 78, 79]

$$F_{vdW} = \frac{A_h}{6\pi D^3} \tag{5.7}$$

where A_h is defined as the Hamaker constant and D is the separation between layers.

The Casimir force is defined as

$$F_{Casimir} = \frac{\hbar c \pi^2}{240 D^4} \tag{5.8}$$

where $\hbar$ is the reduced Planck constant, c is the speed of light in vacuum and D is the initial air gap separation.

In order to study how the short-range forces affect the full behaviour of the suspended control gate whilst in contact, the short-range forces are implemented within CoventorWare as part of the internal forces that take place on the 3D mechanical and electromechanical analysis. Once implemented, the pull-out effect is analysed both with and without the short-range forces. Figure 5.12 shows the result obtained from this analysis.

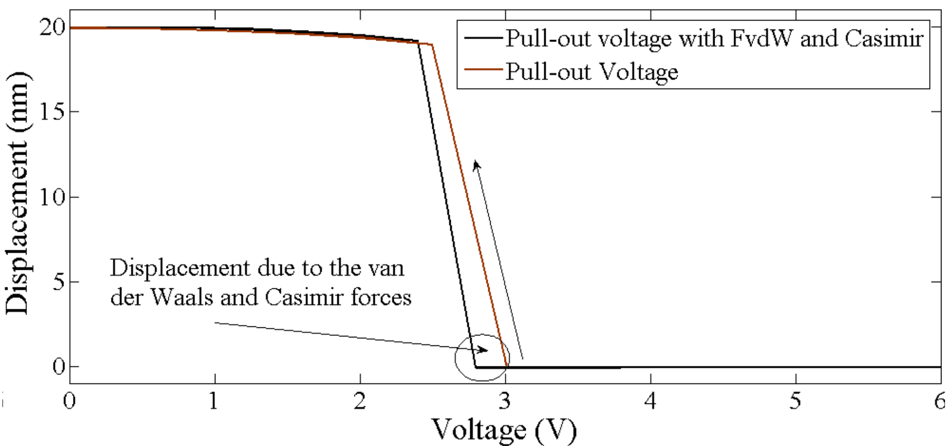


FIGURE 5.12: D - V characteristic curves for the pull-out analysis in which the short range forces are neglected in one analysis and included for the other

The pull-out analysis was analysed based on exactly the same dimensions as for the initial analysis of the doubly-clamped beam described in Chapter 3. As a result of the 3D FEM analyses, a small variation of $\approx$ 300 mV was found between both analyses. This variation becomes negligible at larger figures due to its characteristics. In contrast, at short dimensions the pull-out may never occur due to the fact that the interaction between the short-range forces and the mechanical restore force of the beam are comparatively similar to these set of forces, which is why they cannot overcome one another. In the case of the SGSNM cell, this scalability issue might arise when the air gap reaches $\approx$ 1 nm. Nevertheless, by properly scaling-down the other set of layers and by improving the suspended gate as well as the tunnel oxide layer materials, it is possible to overcome the short-range forces at this scale.

5.5 Comparison with other emerging non-volatile memories

The concept of non-volatile RAM memory has been discussed and put in front of the reflectors for some time. Different concepts have been explored in the last two decades. Major efforts have been focused on three main ideas: Ferromagnetic RAM (FeRAM), Magnetic RAM (MRAM) and Phase Change RAM (PCRAM) [13, 80, 29, 34]. In the last few years, two new concepts have been added to the list: Self-Buckled memory and the Suspended Gate Silicon Nanodot Memory. Each of these devices has its own characteristics and drawbacks, for instance FeRAM, which has the lowest power consumption but is the most difficult to scale. MRAM so far is the fastest, but it requires more current to program than the others, PCRAM shows the smallest and most scalable cell but it presents a big drawback when it comes to endurance. On the other hand, both the Self-Buckled and SGSNM memory devices present a relatively small feature compared with FeRAM and PCRAM. Self-Buckled memory requires low voltage to reach high switching speed between the On/Off states, besides it does not present oxide layer degradation but requires stiction control. In contrast, the SGSNM cell achieves very high programming and erasing speed operations at relatively low voltage and more importantly, enables us to avoid an unfavourable trade off between the On/Off current ratio and the programming and erasing voltages, which is not possible for the Self-Buckled NEM memory. The figures of merit of the emerging non-volatile memories are gathered in Table 5.13. Here are displayed the characteristics for each one of those emerging technologies.

Cell		Flash	Nano Crystal	FeRAM	PCRAM	MRAM	NEM Memory	SGSNM
Device Structure								
Mechanism of Non-Volatility		Charge in Floating Gate	Charge in Nano-dot	Polarisation	Phase Change	MR Charge	Mechanical Bistability	SG Airgap
Issues		Programm Volt. reduction	Programm Volt. reduction	H2 block CVD	Program. Curt. reduction	Program. Curt. reduction	Endurance stiction control	Endurance stiction control
Cell Size		4-8 F^2	4-8 F^2	10-20 F^2	8-15 F^2	8-15 F^2	6-10 F^2	6-10 F^2
Speed	Program	10 usec	10 usec	< 50 nsec	100 nsec	< 50 nsec	20 nsec	< 3 nsec
	Read	< 20 nsec	< 20 nsec	100 nsec	< 20 nsec	< 50 nsec	< 40 nsec	< 20 nsec
Voltage		< 12 V	< 12 V	5 V	3 V	3 V	4 - 6 V	4 - 6 V
P/E Cycle		10^8	10^8	10^{10}	10^{12}	10^{15}	$> 10^9$	$> 10^9$
Scalability		×	△	△	△	△	△	△
Silicon Compatibility		○	○	△	×	×	○	○

× No-possible △ Possible ○ Ok

F^2 - Minimum feature

FIGURE 5.13: Comparison between the emerging non-volatile memory devices with the SGSNM device (Courtesy of Prof. Mizuta)

Chapter 6

Fabrication and Characterisation of Suspended Gate Structures

The non-volatility as well as the fast programming and erasing operations of the Suspended Gate Silicon Nanodot Memory cell have been successfully demonstrated by means of numerical analyses in a hybrid-circuit simulation. Once those characteristics have been demonstrated, the next step that arises in the sequence is the fabrication of the memory device. Therefore, in this Chapter the blueprints to fabricate the SGSNM cell are introduced. The fabrication process of the memory device is divided into three main sections:

- Readout element (MOSFET)
- Floating Gate (SiNDs) Layer Deposition
- Suspended Control Gate

The fabrication process starts with the readout element (MOSFET) implementation as an entity, to later on continue with the fabrication of the floating gate layer by depositing silicon nanodots on the substrate (thick oxide layer). On top of this layer, the sacrificial layer is deposited (poly-Si) and on top of it, a thin aluminium layer as control gate is deposited by using e-beam deposition. The control gate is suspended by using wet and dry etching process.

6.1 Readout Element

The fabrication process of the readout element was performed in collaboration with the Hitachi Central Research Laboratory (CRL). Our counterparts provided us with several samples that contain two kinds of elements: transistor foundations and dummy samples to work with them. The dummy samples contain a high quality silicon dioxide layer and a poly-Si layer stacked over the silicon substrate. The poly-Si layer is used as a mechanical support as well as a sacrificial layer. On the other hand, the foundation samples, besides the stacked layers, include the highly-doped regions (source and drain) as well as the electrodes patterned on the substrate for the source, drain, control gate and back-contact terminals. The high quality SiO_2 layer and a rather thick sacrificial layer were deposited as depicted in Fig. 6.1.

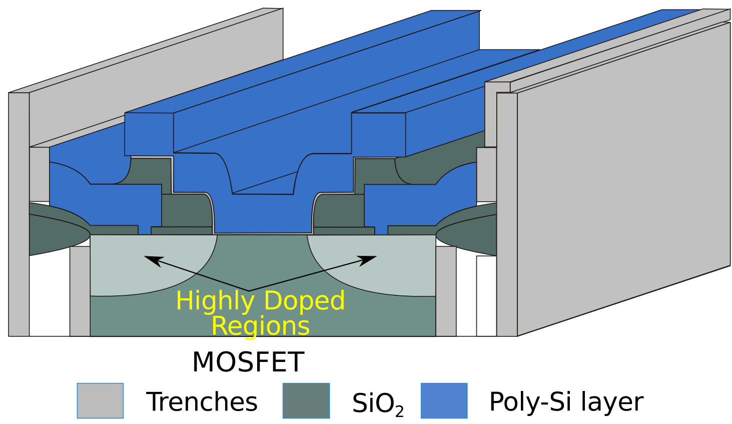


FIGURE 6.1: Cross-section of the transistor foundations provided by our counterparts at Hitachi CRL, which shows schematically the set of layers on the transistor

The dielectric layer on top of the foundations is a high quality oxide that is thermally grown by using a system that is composed of a furnace and a torch. The torch fed with O_2 and H_2 the furnace that creates water vapour that flows on it and reacts with the substrate made of silicon to create SiO_2 and volatile hydrogen. The chemical reaction that represents the above elements is described in Eq. (6.1).

$$\mathrm{Si} + 2\mathrm{H_2O} \Rightarrow \mathrm{SiO_2} + \mathrm{H_2} \uparrow \tag{6.1}$$

A thick poly-Si layer is deposited on top of the oxide layer by using the low pressure chemical vapour deposition (LPCDV) process [81]. Finally, on top of the sacrificial layer an aluminium layer is deposited by using electron beam deposition at high vacuum. This layer is responsible for performing the movement, as function of the applied voltage, of the pull-in and pull-out for the programming and erasing processes of the SGSNM. A set of schematic diagrams that represents the full fabrication process to suspend the control gate is displayed in Fig. 6.2.

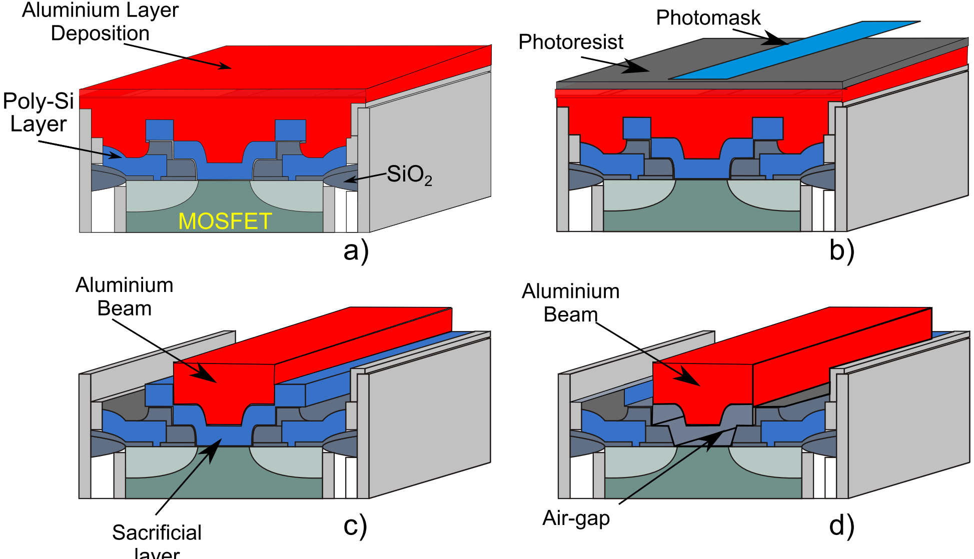


FIGURE 6.2: Schematic set of diagrams that describe the fabrication process of the suspended control gate over the dummy samples. a) e-beam deposition of the aluminium layer, b) photolithography process by using a double-clamped features mask, c) etching process for the aluminium layer and finally, d) isotropic dry etching process that suspend the double-clamped structure

6.2 Floating Gate (SiNDs) Layer Deposition

The fabrication of the floating gate on the Suspended Gate Silicon Nanodot Memory cell is performed by depositing a silicon nanodots monolayer over the thick SiO_2 surface. In the deposition process (due to be done), it is required to have high control over the diameter and density of the silicon nanodot monolayer. The technique used to deposit the silicon nanodot over the substrate (SiO_2) layer was developed by the Tokyo Institute of Technology. This technique uses a very high frequency (VHF) plasma decomposition of SiH_4 within an ultra high vacuum (UHV) chamber [2]. A schematic diagram of the equipment that is used to deposit the silicon nanodots is shown in Fig. 6.3.

The set of conditions implemented within the equipment, in order to obtain the silicon nanodots, takes into account two key parameters: the silicon nanodots grow larger in SiH_4 plasma and the nucleation of the silicon nanodots in the presence of H_2 is enhanced [2]. This analysis shows that the presence of hydrogen during this process increases nucleation rather than silicon nanodot growth [2]. Hence, the idea is to separate the growth and nucleation by controlling the H_2 fed into the chamber. When H_2 is fed into the chamber, nucleation of the silicon nanodot takes place and when off the nanodot grows in presence of pure SiH_4 plasma. As a result of the mix of gases shown elsewhere, the final product of this pulsed H_2 analysis is the fabrication of silicon nanodots with a diameter of 8 ± 1 nm. A histogram of the SiH_4 plasma with pulsed H_2 combination is depicted in Fig. 6.4.

Once the silicon nanodots have been deposited over the sample surface by using the

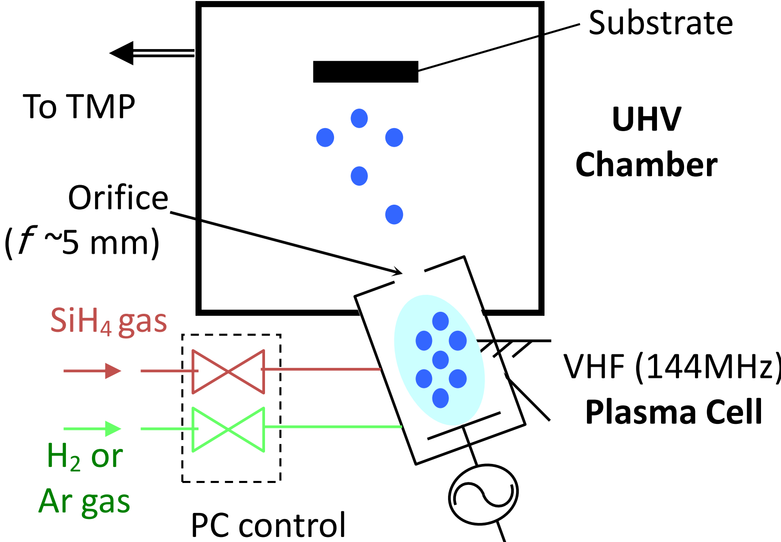


FIGURE 6.3: Schematic diagram of the apparatus used to deposit silicon nanodots by using SiH_4 plasma with pulsed H_2

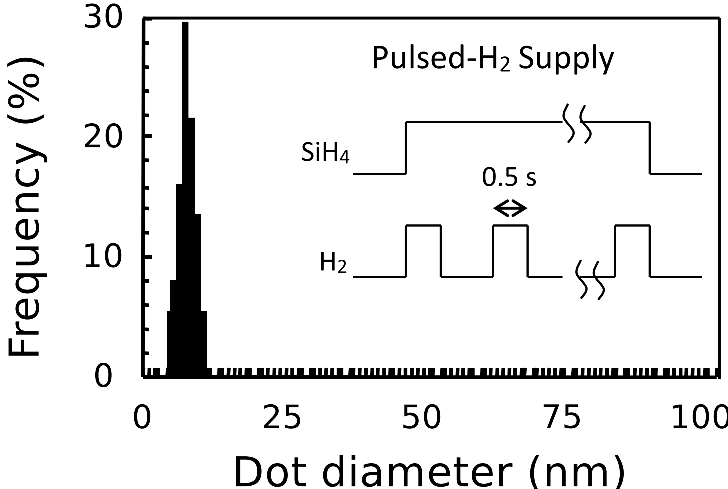


FIGURE 6.4: Distribution of the silicon nanodots size fabricated by pulsed-H_2 supply into SiH_4 plasma [2]

apparatus above mentioned, the technique known as Langmuir-Blodgett is used to assemble the silicon nanodots on the surface. This technique shows that it is possible to obtain a high density concentration of about 7.33×10^{11} dots/cm^2 when in presence of hexamethyldisilazane (HMDS) and chloroform as a suspension which is seven times larger than other methods such as deposited-film and twice compared to the film formed by the drop and evaporation method described elsewhere [2, 3, 26, 82]. A set of SEM images that shows the dot assembly process by using the Langmuir-Blodgett technique is depicted in Fig. 6.5.

The size of the silicon nanodots expected to be deposited on top of the substrate (SiO_2 layer) is $\approx 12 \pm 1$ nm. After the deposition, the samples are oxidised to form a SiO_2 shell to involve the silicon nanodot. The oxidising process will create an oxide layer of 4

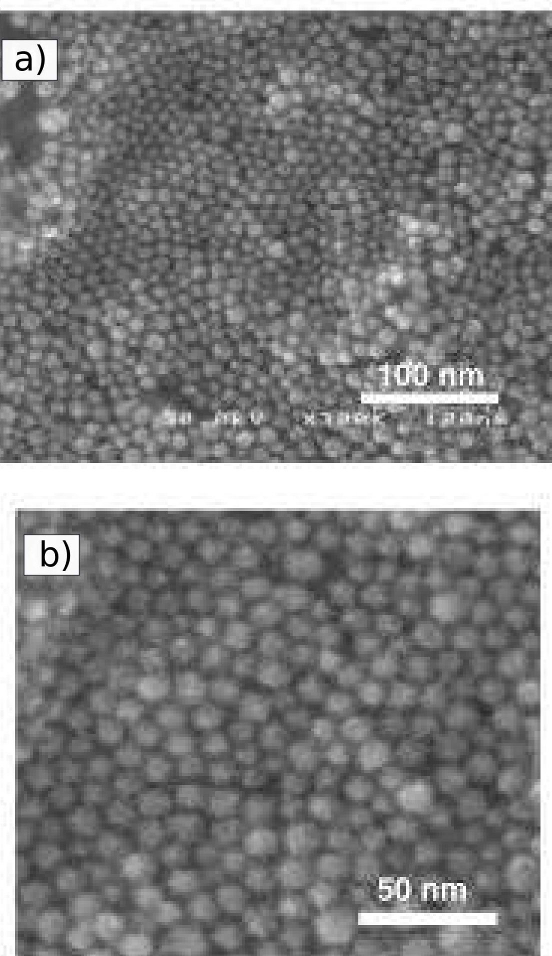


FIGURE 6.5: Set of SEM images that shows a silicon nanodot monolayer a) coated by HMDS with a density of 7.33×10^{11} dots/cm^2 and b) a zoom of the top view of the film in which a lattice-like structure is observed [3]

nm thick (2 nm thick for each one) that will isolate the silicon nanodots avoiding direct tunnelling and charge migration between them. On top of the floating gate layer will appear a rough surface created by the oxidising process. In order to smooth the rough SiO_2 surface a chemical-mechanical polishing method is used [83, 84]. By using this method, the surface of the floating gate layer is planarised and by using atomic layer deposition (ALD) the tunnel oxide layer is refined. It is possible to have large control on the deposition rate of the SiO_2 by using ALD due to the fact that the tunnel oxide layer is just a few nanometres thick (7 nm) [85]. A schematic diagram that displays the silicon nanodot thicknesses after oxidisation, the planing process of the floating gate layer after oxidising and the tunnel oxide layer deposition by using the ALD is depicted in Fig. 6.6.

As a consequence of the techniques mentioned above, it is possible to obtain a densely packaged silicon nanodots monolayer, in which the silicon nanodots are isolated between them by a SiO_2 shell that avoids lateral migration. The floating gate is also isolated by

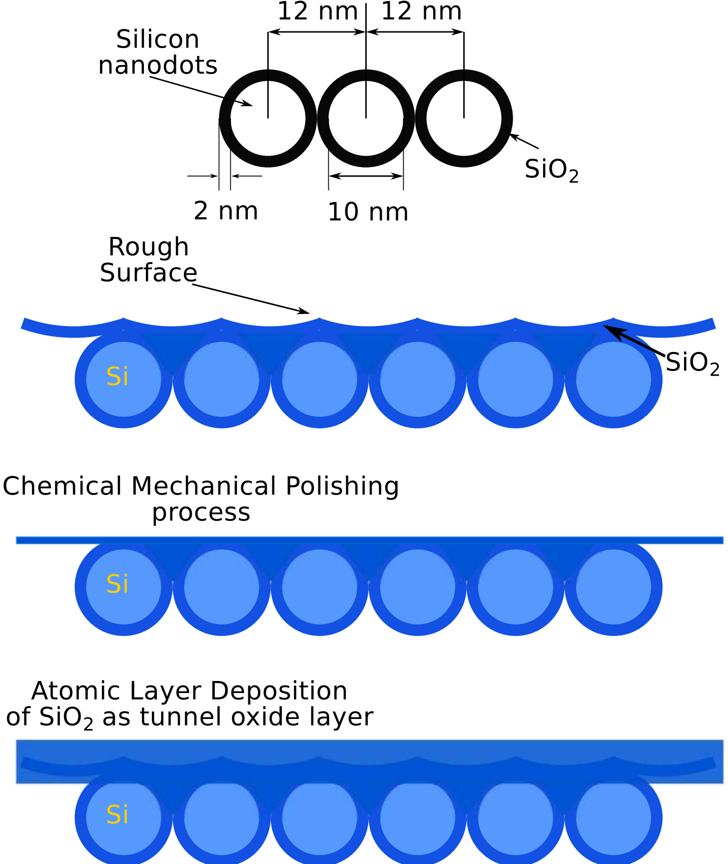


FIGURE 6.6: Set of schematic diagrams that a) display the thicknesses within the silicon nanodot, b) roughness of the surface after oxidising the silicon nanodot layer, c) after polishing the rough surface produced by the silicon nanodots oxidising process and d) atomic layer deposition process to deposit SiO_2 as tunnel oxide layer

a rather thick thermally grown oxide layer from the readout element which avoids the creation of leakage paths towards the substrate and towards the control gate by a thin tunnel oxide layer. Those oxide layers keep the charge stored within the silicon nanodots confined and the surface as smooth as possible for the tunnelling processes.

6.3 Suspended Control Gate Fabrication Process

The fabrication process of the suspended control gate requires the deposition of a sacrificial layer over which the control gate layer will be deposited. The sacrificial layer has a direct influence on the final behaviour of the suspended control gate due to the mechanical support provided initially for the control gate deposition and later on as it is suspended, and the remnant of the layer is used as the pillars over which the beam structure will be supported to perform the programming and erasing operations. The materials used as sacrificial layers within the fabrication processes are directly related with the capabilities of the facilities to get rid of them by using etching processes, *e.g.* dry etching (gases and equipment) or wet etching (acid and base benches). The mate-

rials commonly used as sacrificial layers are: poly-Si, germanium, amorphous-Si (α-Si), phosphosilicate glass (PSG), SiO_2, and polyimide, among others [39, 40, 55, 86]. For the SGSNM cell, poly-Si is used as sacrificial layer. This material is chosen due to the mechanical properties that it provides when the control gate layer is deposited as pillars once the sacrificial layer has been removed and because within our clean room at the University of Southampton we have the capabilities and facilities to etch this layer.

Once the sacrificial layer has been deposited over the sample as depicted in Fig. 6.1, the sample is measured by using two different methods such as ellipsometry and surface profiler. The information provided by these tools is analysed and used as a log for the sake of the device reproducibility and in the case that errors might occur during the fabrication process might occur. After being measured, it is possible that the sample is contaminated with organic or inorganic materials present on the measuring instruments. Hence, in order to avoid the cross-contamination, the sample is cleaned of both organic and inorganic materials. The cleaning process implemented is simple. The sample is soaked in a beaker with acetone. The beaker is put in an ultra-high frequency bath for 10 minutes, after which the sample is rinsed with isopropanol to remove the acetone and finally, it is dried with a N_2 gun. If the surface is quite contaminated and the substrate is Si-based, the heterodox option could be used which is to put the sample on fumic nitric acid (FNA) for 10 to 30 sec, rinse it with de-ionised (DI) water and dry it by using a N_2 gun. This process will clean the surface by removing any impurities on the Si-based substrate either organic or inorganic.

Aluminium Deposition

After performing the optical measurements and the associated cleaning process, the sample is glued into the evaporator chamber in order to deposit the aluminium layer. The evaporator system is an e-beam evaporator BAK600 from Leybold which has the capacity to handle 10 different materials 8 for electron beam coating through two crucibles (4 each) and 2 by using thermal coatings. The recipe for aluminium deposition (500 nm) states the use of several parameters to obtain a rather smooth surface. This recipe uses two pots as aluminium sources, low power (32 %), high voltage (8 kV), power (1.52 kW) and high vacuum ($<1\times10^{-5}$ mbars). These parameters gives a low rate deposition typically 2.1 to 2.7 Å/sec. By having this deposition rate, the final product shows a flat surface at the thickness required. Figure 6.7 show a set of images that depicts the sample coated with a thin aluminium layer. Once finished, the sample is measured by using ellipsometry and the stepper. The ellipsometer shows a thickness of $\approx$ 500 nm with a roughness of $\pm$ 2 nm. The conformal deposition shows the aluminium thickness on the walls created from previous structures such as MOSFET channel or pads. After those measurements are performed, the control gate layer is ready for the photolithography process.

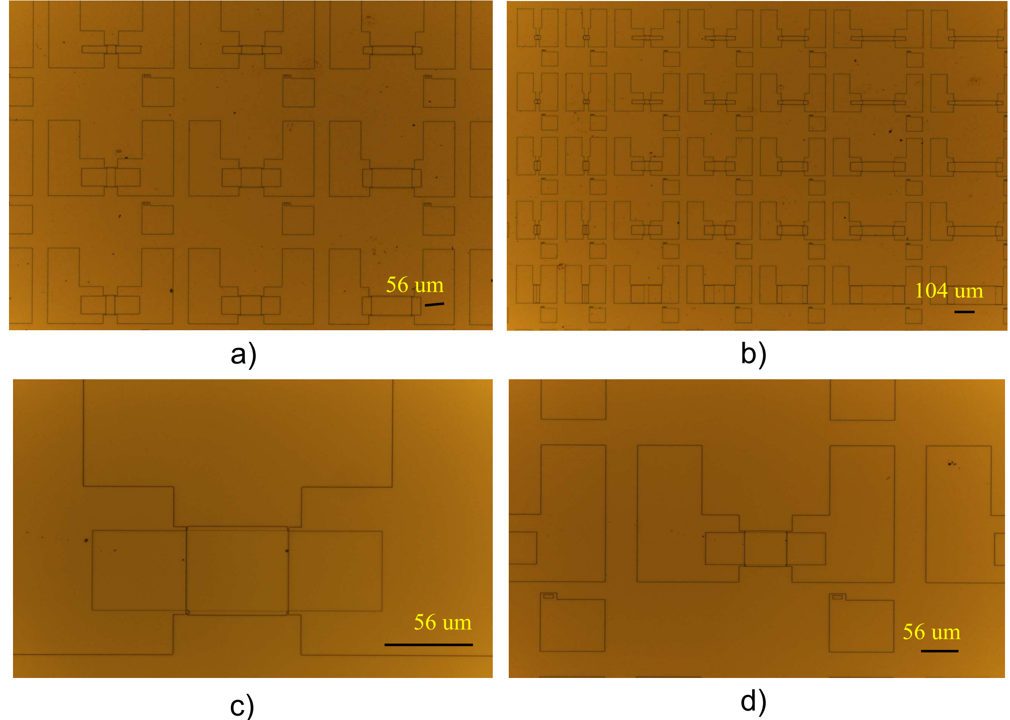


FIGURE 6.7: Set of optical images obtained after the sample, with MOSFET features, have been coated with aluminium. The images obtained are displayed with a zoom of a) 100×, b) 500×, c) 1000× and d) 10000×

Spin Coating Process

The imprint process starts by cleaning the sample surface as described elsewhere (not with FNA). The photoresist used to coat the samples is positive from the company Shipley series S1800. The photoresist used to pattern transfer the beams is S1813. To uniformly distribute the S1813 resist over the sample surface a Cee200 spin coater is used. According to the data-sheet provided by the photoresist company, it is possible to obtain a specific thickness in function of the acceleration and time that the photoresist is spined. In our case, we need the photoresist layer to have a thickness of 1 µm, and the recipe implemented on the Cee200 is optimised specifically for the S1813 photoresist. Table 6.1 shows the set of steps programmed within the spin coating machine, in order to obtain the 1 µm thickness.

TABLE 6.1: Spin parameters optimised to obtain 1 μm thick coating layer for a S1813 positive photoresist (Shipley)

Step	Speed (rpm)	Acceleration (rpm/s)	Time (sec)
1	5000	2500	2
2	5000	0	30
3	100	1000	2

After spin the photoresist on the sample, it is soft-baked on a hot plate for 60 sec at 115°C. The soft-baked process helps to improve the adhesion of the photoresist to the wafer or sample, as the case for the SGSNM, and also it hardens the resist layer. Right after those processes conclude, the sample is handled on other equipment to proceed with the pattern transfer.

Transfer of patterns

The system that is used to transfer the patterns from the photomask into the sample is a high-precision mask aligner AVG 620TB series. The pattern transfer process in this system follows a series of steps. As the first step, the mask is handled into the aligner and a set of parameters related to the mask and to the sample alignment process is introduced into the system. Those parameters includes thickness of the substrate, thickness of photoresist, area under exposition, separation distance between the surface of the sample (photoresist) and from the mask, exposure time, wavelength and contact pressure on the mask-sample interface. Once those parameters have been introduced to the system, the sample is positioned under the mask. At this point is when the manual aligning process starts by overlapping the aligning marks by using the x and y knobs and by tilting the sample between them through the θ knob. Once the aligning marks between both sample and photomask are overlapped, it is exposed by using UV light at 335 nm with an exposition time of 1.5 sec. As a result of the recipe developed for the small samples (2 × 2 cm) of the SGSNM, the maximum misalignment that the sample has is ≈ 500 nm over a 2 × 2 cm^2 sample. double-clamped beam features from 700 nm to 5 μm in width to several lengths are successfully obtained with a fully repeatable processes. Right after the exposition process finishes, the sample is soaked in developer. The solution used to develop the sample is the MF-319 developer. This chemical dissolves the photoresist not covered by the double-clamped features during the exposition step due to chain scission occurs after exposure. The sample is soaked in MF-319 for 45 to 47 sec, after that it is rinsed in DI water for a similar period of time and dried by using the N_2 gun. Figure 6.8 shows a set of images that contain the double-clamped features.

Wet and Dry Etching Processes

Once the aluminium layer has been patterned with double-clamped beam features, it is etched-out by using wet etching process. The aluminium etcher based on NPE 80/5/5/10 that has the following composition: H_3PO_4 : HNO_3 : CH_3COOH : H_2O [87]. It has a rate of ≈ 100 nm/min when heated at 30°C. Once the conditions are met, the sample is soaked in the solution by using a small sample basket. By following the etch-rate condition, the aluminium layer thickness (500 nm) is etched. It is possible to see how the aluminium layer is dissolved by the aluminium etcher by watching the opaque surface

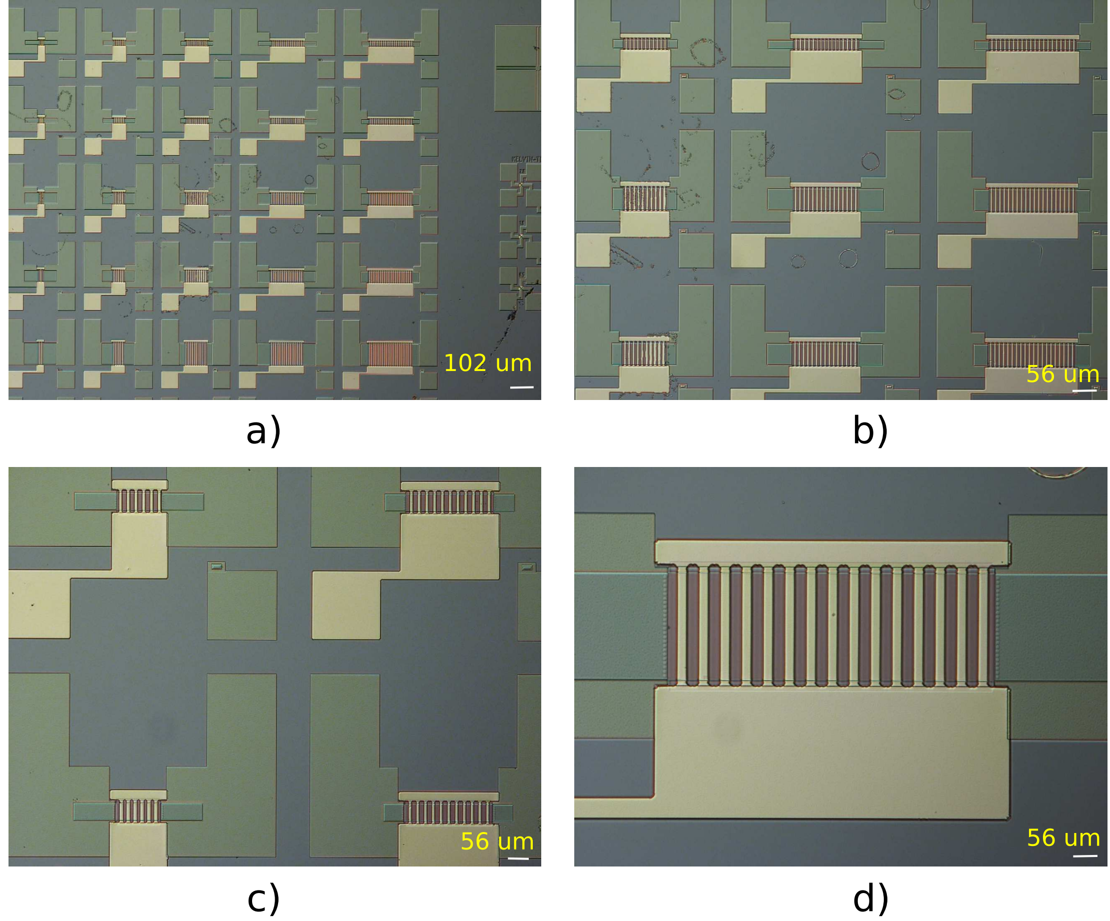


Figure 6.8: Set of optical images obtained after the photolithography process is performed. In here the aluminium beam features looks rather sharp and well defined as shown at different zooms such as a) 5×, b) 10×, c) 20× and d) 100×

of the poly-Si layer rather than the aluminium when time is ≈ 80% of total. According to the etching rate, after five minutes the process finishes and the sample is rinsed in DI water for ten minutes to remove any contaminants that might remain from the previous process. Once rinsed, the sample is dried by using the N_2 gun. To corroborate the etching process, the sample is analysed through the optical microscope and by using the scanning electron microscope (SEM). The images obtained by the optical microscope show that the aluminium layer uncovered by the photoresist has been removed from the top of the sample but the double-clamped beam features as they are displayed in the sequence of images depicted in Fig. 6.9.

A deep analysis of the aluminium beam features is performed by using SEM images. In these images (Fig. 6.10) the side walls, and the surface as well as the integrity of the beam are analysed. By spotting the aluminium structures on the sample surface, this shows that the walls of the beams were sharply etched without showing any damage on either the beam surface or the substrate. The set of images also show the lack of aluminium islands remanent on the beam surface or substrate. The implemented recipe shows an almost infinite selectivity between the aluminium and poly-Si layers.

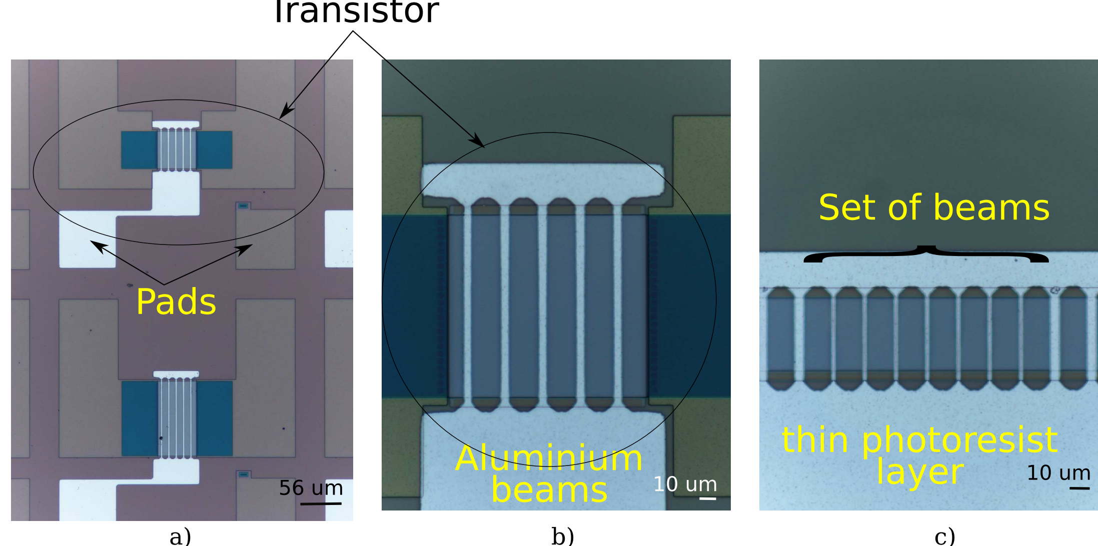


FIGURE 6.9: Set of optical images in which the double-clamped features are analysed after the aluminium wet-etching process took place. These images shows the aluminium beams sharpness at different zooms such as a) 5×, b) 10×, c) 20× and d) 100×

Once the aluminium double-clamped features are obtained, the next step within the fabrication process is to remove the sacrificial layer (poly-Si). The sacrificial layer is etched out by using a dry etching process that is divided into two steps such as anisotropic and isotropic etching processes [88]. This process is performed by using a RIE 80+ plasma equipment. The recipes used to perform the etching processes are tailored to react only with the sacrificial layer without damaging the substrate (silicons) or the tunnel oxide layer (SiO_2) due to both being Si-based. The gases that are used to realise each one of the steps above mentioned are sulphur hexafluoride (SF_6) and oxygen (O_2). The first step in the etching process uses the aluminium beam as a mask to remove anisotropically the poly-Si layer. Table 6.2 shows the rates and conditions used to obtain the anisotropic dry-etching processes without reacting with the oxide layer on the substrate.

TABLE 6.2: Mix of gases and conditions used within the RIE 80+ plasma system to remove anisotropically the sacrificial layer (poly-Si)

Step	Gases mix	Flow rate (sccm)	Time (sec)	Pressure (mTorr)	RF (Watts)	DC (V)
1	SF_6 & O_2	96 : 12	30	0	0	0
2	SF_6 & O_2	96 : 12	350	200	100	0
3	Ar	30	30	0	0	0

By implementing the set of parameters listed in Table 6.2 into the RIE 80+ equipment, the dry-etching process was performed obtaining as a result the removal anisotropically of the sacrificial layer as depicted in Fig. 6.11.

After removing anisotropically the poly-Si layer, a new recipe is implemented on the RIE 80+ equipment in order to remove the sacrificial layer isotropically. In Table 6.3 is

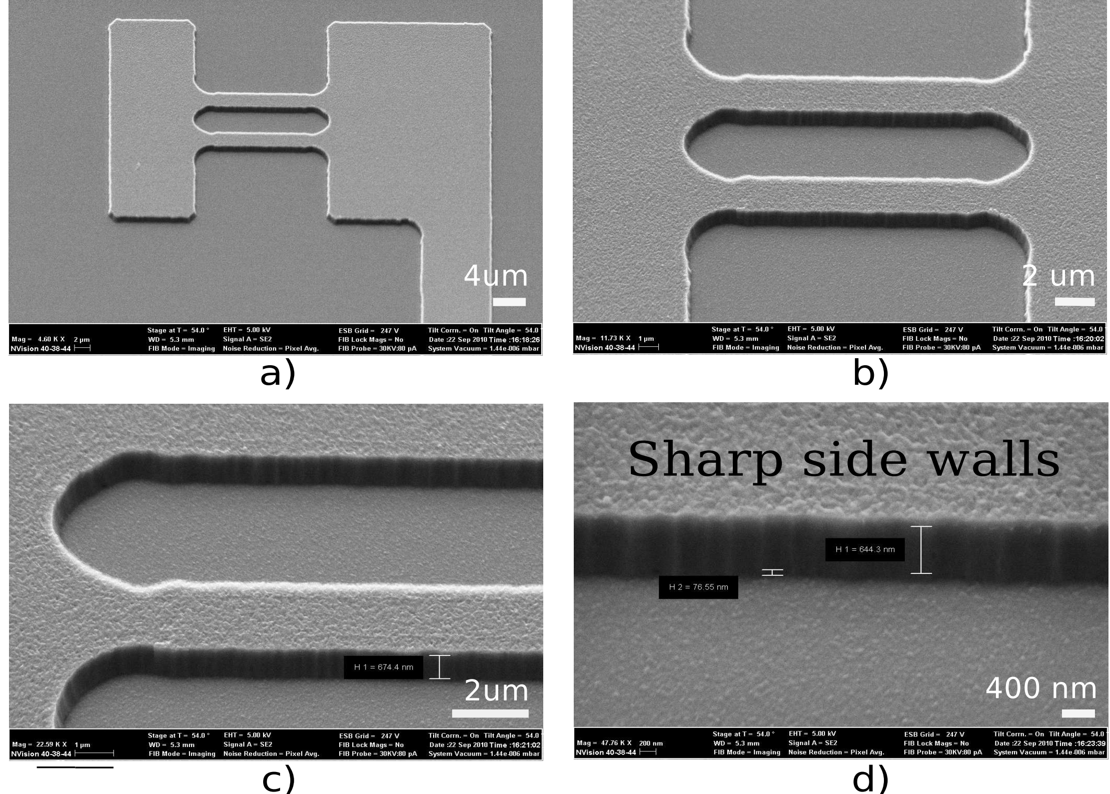


FIGURE 6.10: Set of SEM images tilted 54° shows the double-clamped beam structures after the aluminium wet etching process took place as shown in image a), image b) shows that by zooming on the two beams inner area, the side walls and the beam surface are not over-etched, in image c), the beam surface looks uniform and smooth and in image d) it shows that the sidewalls are sharp and the substrate do not show apparent damage

listed the set of parameters used to obtain the isotropic etching process of the poly-Si layer. As a result of the implementation of this recipe, a set of SEM images depicted in Fig. 6.12 shows the results obtained.

TABLE 6.3: Mix of gases and conditions implemented to isotropically remove the sacrificial layer by using

Step	Gases mix	Flow rate (sccm)	Time (sec)	Pressure (mTorr)	RF (Watts)	DC (V)
1	SF_6 & O_2	12 : 1.2	30	0	0	0
2	SF_6 & O_2	12 : 1.2	180	400	100	0
3	Ar	30	60	0	0	0

After the second round of the dry-etching process is performed, on the beam sidewalls as well as at the top of it, some damage is visible due to the double-plasma exposition. As a result of the double-exposition of plasma, the undercut is obtained affecting the beam integrity and the substrate. To solve this issue, a new recipe that can isotropically get rid of the sacrificial layer in a single plasma session is envisaged. The characteristic that are required for a single isotropic plasma exposition are: high selectivity, low power

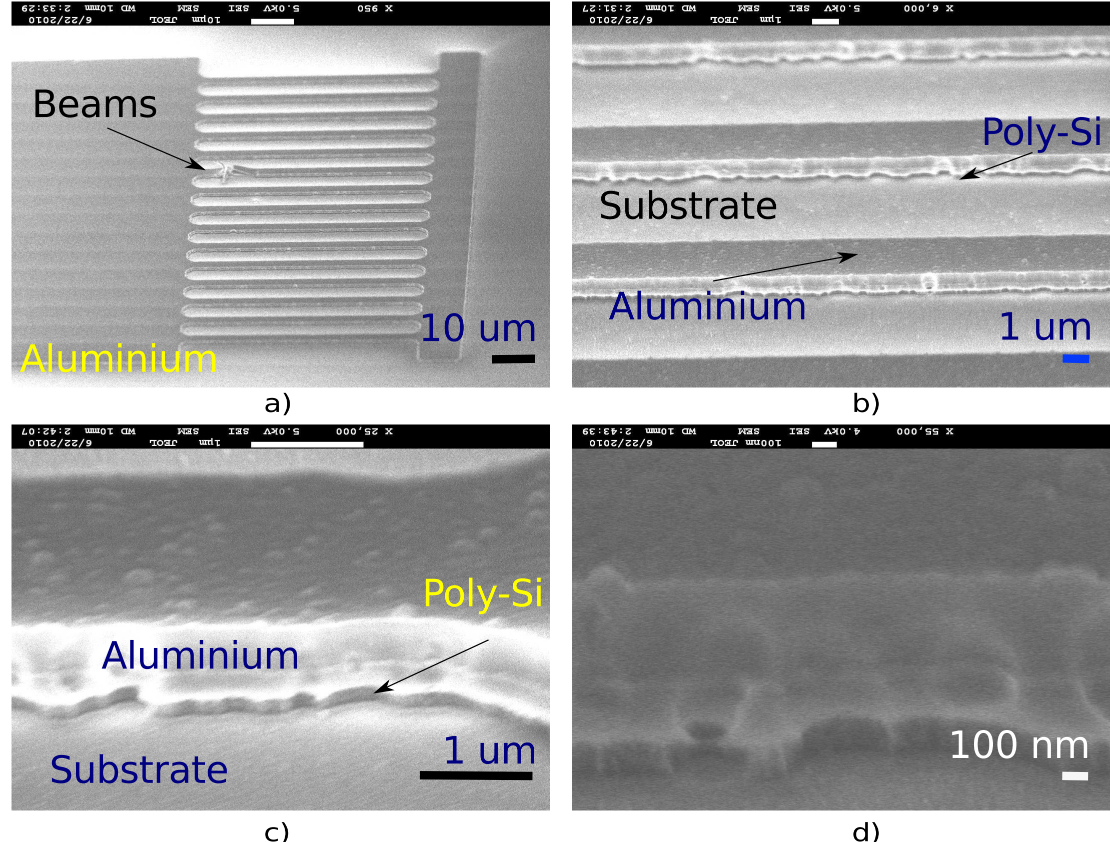
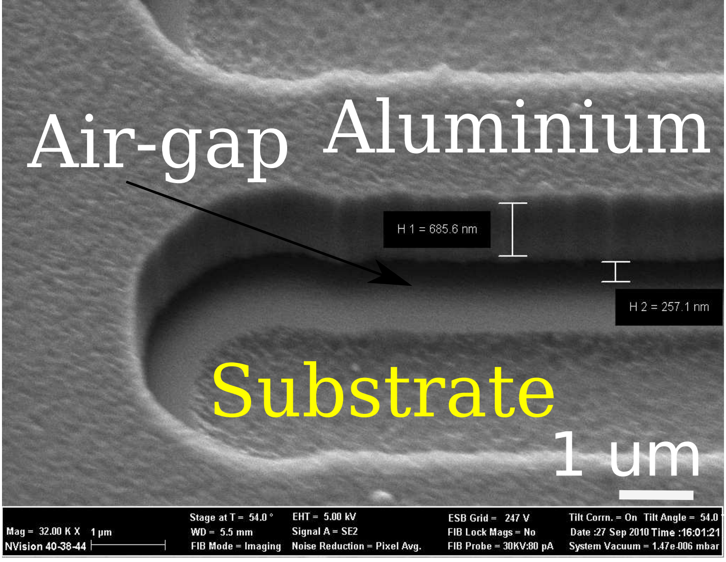


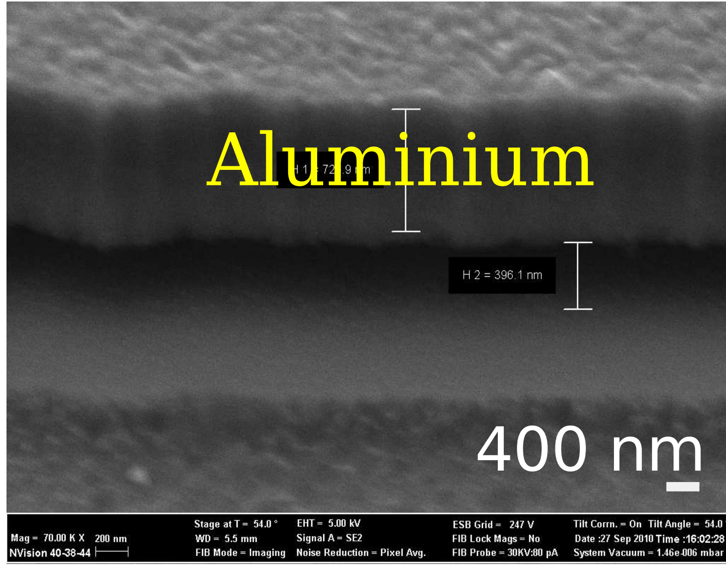
FIGURE 6.11: Set of SEM images obtained after the anisotropic dry etching process was performed, a) shows the set of images as an entire beam array, b) by zooming on the beams, it is possible to see the bi-layer beam in which the aluminium and the poly-Si layer are visible, c) and d) show a zoom over the central beam area, which shows a sidewall not as smooth as in previous image

and a chemical reaction rather than a physical one as for anisotropic processes such as the previous one. In order to obtain the single plasma exposition, several fluorine based compounds such as SF_6, tetrafluomethane (CF_4), trifluoromethane (CHF_3) and Xenon difluoride (XeF_2) are capable of isotropically etching the poly-Si layer. Nevertheless, the best candidate is SF_6. Under plasma conditions, SF_6 is degraded to SF_4 that is a stable molecule in which the extra F_2 radicals react with the Si-based layer in the presence of O_2. The excess of this mix is swept out from the camber by Ar. Considering those factors and by modifying the conditions within the RIE 80+, the chemical reaction that is required to isotropically etch-out the sacrificial layer is tested. Table 6.4 list the parameters used to obtain the one-step plasma isotropic process.

The recipe that contain the set of parameters used to isotropically remove the sacrificial layer was tailored to produce a chemical interaction between the sacrificial layer and the set of gases. The three step recipe process states an homogeneous atmosphere (SF_6 & O_2) prior to the plasma process, the chemical process and finally, the purge of the chamber by the inert gas. The flow of gases that are pumped into the chamber of the RIE 80+ are rather low, this flow is needed due to the reaction that is required is not



a)



b)

FIGURE 6.12: Set of SEM images that shows the double-clamped beams successfully suspended when two dry-etching plasma processes are used. a) Shows the air-gap at the corner of the set of beams and b) shows the under-cut at the centre of the double-clamped beam structure

physical as in most fabrication processes but chemical. As part of this chemical process is the DC bias, this 0 V DC bias expressed in Table 6.4 indicates that there is not acceleration voltage within the chamber produced by the bottom electrode due to the the mix of gases and pressure within the chamber. As a result of the single-step-plasma process, the sacrificial layer (poly-Si) was remove isotropically without affecting severely the substrate or other layers as depicted in the set of SEM images in Fig. 6.13.

TABLE 6.4: Mix of gases and parameters used within the RIE 80+ plasma system to remove isotropically the sacrificial layer by using a one-step plasma process

Step	Gases mix	Flow rate (sccm)	Time (sec)	Pressure (mTorr)	RF (Watts)	DC (V)
1	SF_6 & O_2	12 : 1.2	30	0	0	0
2	SF_6 & O_2	12 : 1.2	420	400	100	0
3	Ar	30	60	0	0	0

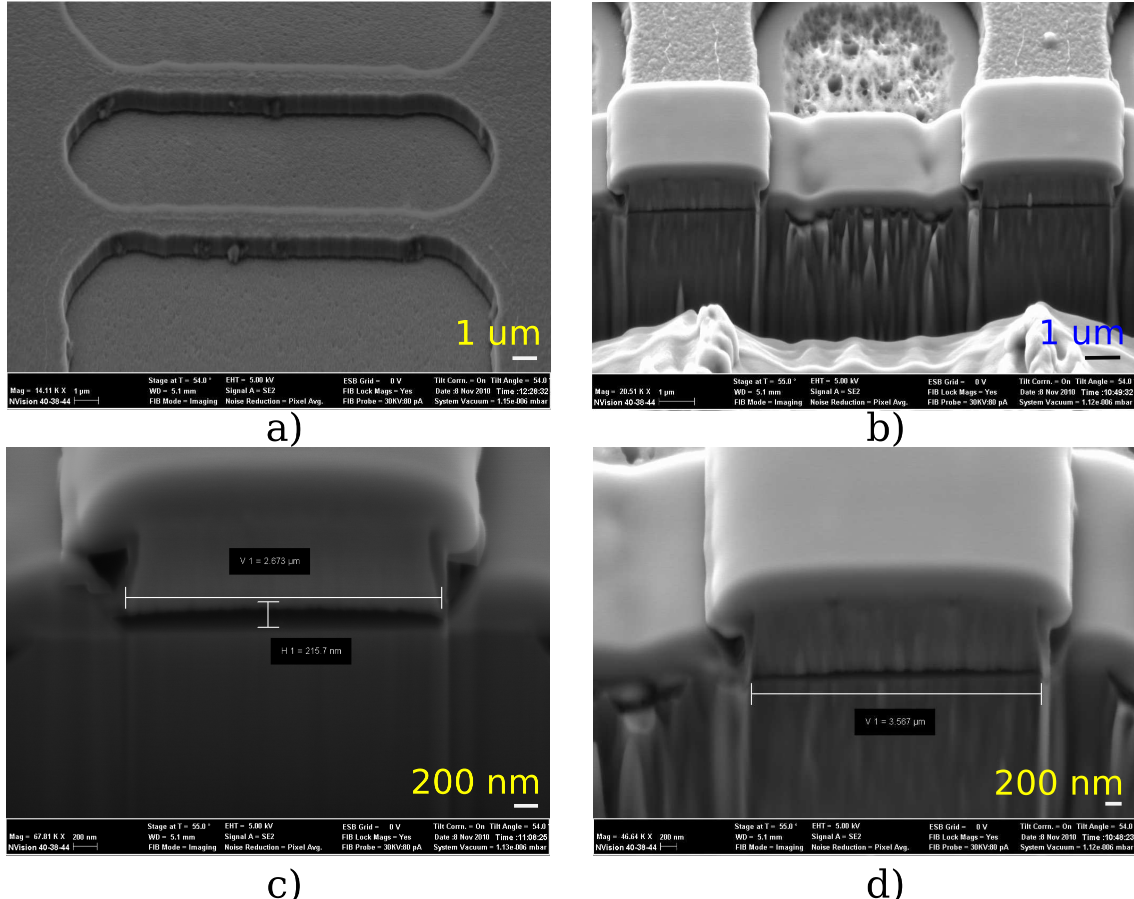


FIGURE 6.13: Set of SEM images that represents the one-step-plasma process that isotropically removes the sacrificial layer of the double-clamped beam features without affecting the other layers. Image a) shows two double-clamped beam structures before the cross-section process, b) SEM cross-section of the beams showing successfully the undercut in both structures, c) and d) images display individually a zoom over each beam

The chemical interaction produced by the interaction between the set of gases used to isotropically remove the sacrificial layer, allows us to, by one-step-plasma, get rid of the poly-Si layer without affecting the substrate (SiO_2) nor the suspended control gate. As a result of the dry etching process, the cross-section analysis through the set of SEM images, shows that the sacrificial layer was successfully removed. The uniformity of the air-gap displayed on the SEM images shows the effectiveness of the recipe as well as the repeatability of the entire process. Continuing with the trend, the next step within the

process is to test the movement of the suspended beam features through capacitance variation by using a C - V meter.

As a result of the dry etching process, the cross-section analysis through the set of SEM images, shows that the sacrificial layer was successfully removed. The uniformity of the air-gap displayed on the SEM images shows the effectiveness of the recipe as well as the repeatability of the entire process. Continuing with the trend, the next step within the process is to test the movement of the suspended beam features through capacitance variation by using a C - V meter.

6.4 Electro-Mechanical Beam Measurements

After the double-clamped beam structures have been successfully suspended, their movable capabilities are investigated through the pull-in and the pull-out effects by performing capacitance - voltage measurements on the beam structures. The capacitance - voltage measurements are performed by using a C - V meter Agilent 4279A @ 1 MHz, which is linked to a test probe system. The measurement starts by handling the sample over the chuck in the test-probe. As soon as the sample is in place, the calibration process of the equipment takes place (see AppendixA). Once the calibration process has finished, the test probes are positioned on the substrate (chuck) and on the control gate electrode, respectively, whilst in the C - V meter, a set of parameters such as initial and final voltages, steps and the sensitivity required are introduced. It is necessary to point out that the C - V meter has the capacity to operate at full cycle from $\pm$ 100 V. However, at the moment it is restricted to operate at $\pm$ 38 V. There were performed several analyses by using a single beam, double beams as well as several arrays of double-clamped beam structures. There were performed several C - Vmeasurements, from which just few of them were successful. From these analyses, we present as successful results a single and an array of double-clamped beam structures. The first successful structure is a single double-clamped structure that has 58.3 µm in length at 4.8 µm width, with a beam thickness of $\approx$ 600 nm and an air gap of $\approx$ 320 nm. In here, it is necessary to pointed out that the initial dimensions of the beam were modified mainly due to the fabrication process which involved an increment in length and a reduction in width as pointed out. Moreover, the variation in thickness of the suspended control gate was produced by the single-step plasma process due to the generation of aluminium oxide on top of the structure. Figure 6.14 shows the double-clamped beam structure under test.

Once the structure displayed in Fig. 6.14 was handled on the C - V meter, the conditions to characterise the structure were stated. The initial step was to define the range 0 voltage in which the apparatus will sweep as well as the step and the integration rate. The set-up include a span of voltage from 0 to 20 V with a step of 0.5 V and a low integration rate. As a result of the electro-mechanical analysis a C - V curve was

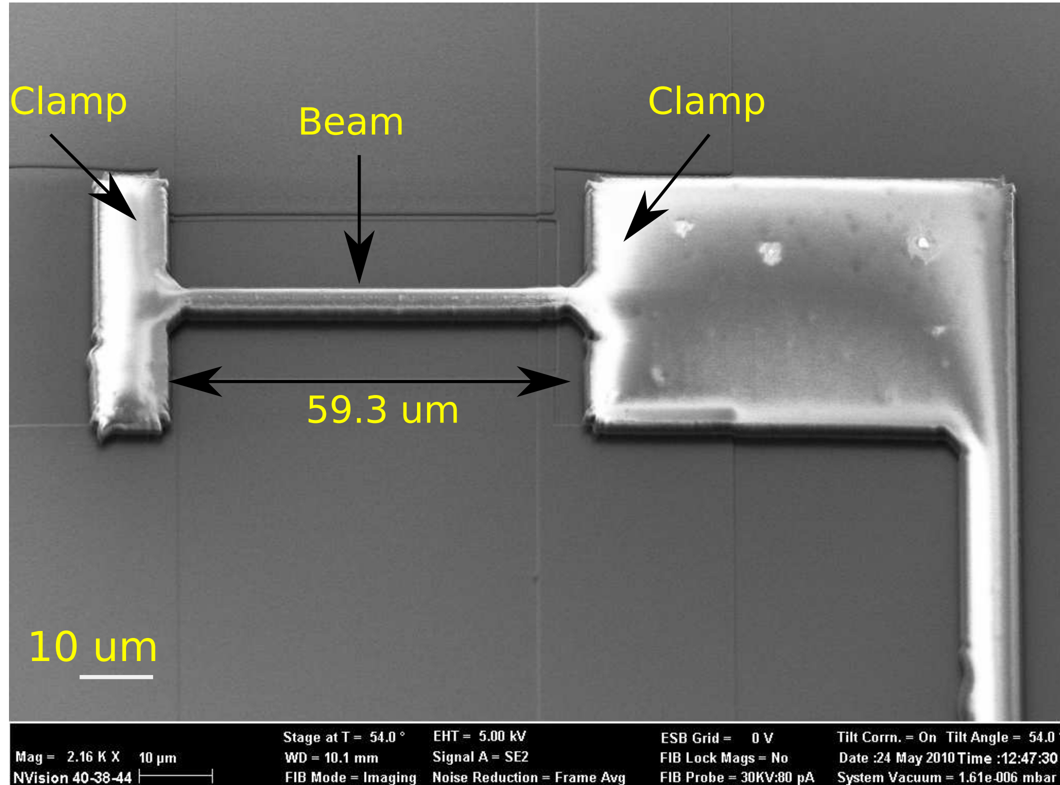


FIGURE 6.14: Double-clamped suspended beam structure under test in which the beam electrode is visible over the substrate

obtained (Fig. 6.15). In here, the biased structure shows how the beam is bending when 6 V are reached. By increasing the applied voltage till 9 V, it is possible to see how the beam is at the pull-in voltage. Further increment in the applied voltage makes the beam to collapse on the substrate due to the electrostatic force that is acting on both layers. The intermediate set of points are due to the un-homogeneous etching process of the sacrificial layer. This variation in the etching process produces that some sections of the beam bends at different time rather than in an uniform and smooth way as in the numerical analysis.

By using a capacitance - voltage approximation function, the displacement - voltage characteristic curve was obtained as shown in a small upper-left corner on Fig. 6.15. This curve shows us how the beam bend downwards by action of the applied voltage. In this curve, it is possible to see clearly that the pull-in point was reached at 9 V and the intermediate points displayed on the C - V curve are due to the stiffness on the beam when bending. This extra stiffness can be due to the top oxide layer generated after the oxygen plasma process and due to the non-uniform etching process of the sacrificial layer. Once the control gate has collapsed on the tunnel oxide layer and the applied voltage still increasing on it, the contact area increases until the maximum voltage is reached, as shown in the numerical simulation. When the applied voltage reached the maximum voltage programmed, the sweeping process changes and the second part of the sequence takes place from 20 V to 0 V. When the applied voltage was considerably lower than

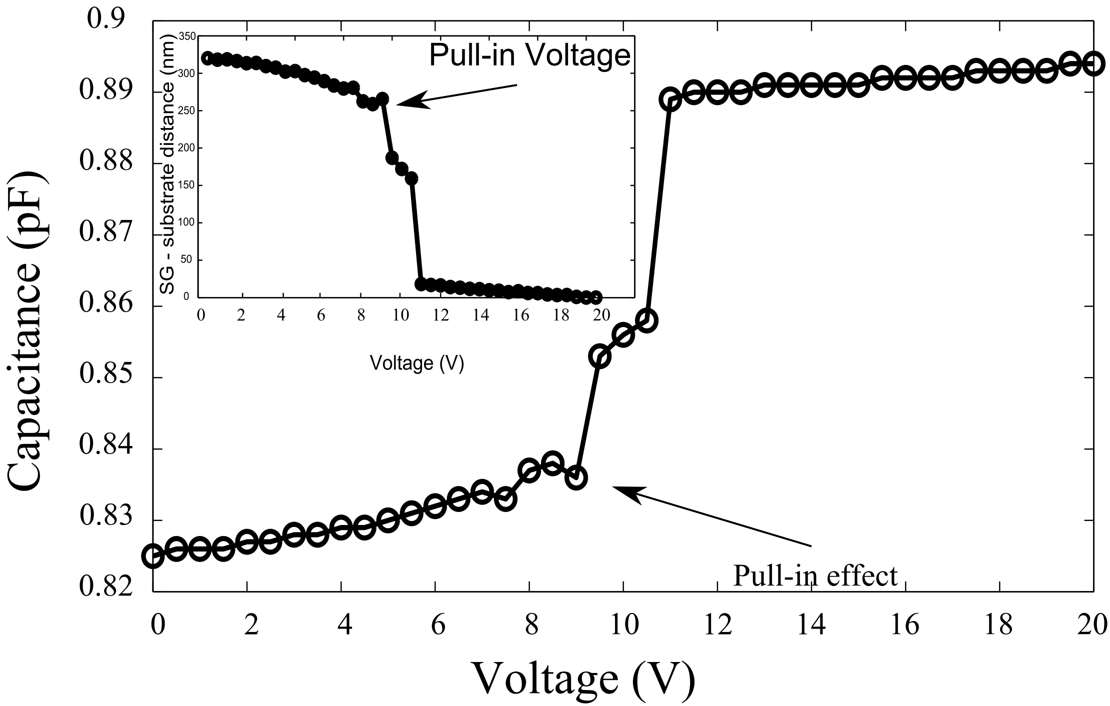


FIGURE 6.15: Pull-in effect obtained as a result of the electrical C - V measurements of a single double-clamped beam of 59.3 µm in length at 4.8 µm width, with a beam thickness of 500 nm and an air-gap of 320 nm. By using the information obtained from the C - V characterisation, the pull-in effect and the displacement - voltage characteristic curve as displayed within the main figure in the upper-left part

the pull-in voltage, the control gate remain attached to the substrate. Further reduction in the applied voltage did not show any change in the position of the two layers, this indicates that the short-range forces are stronger that we considered them and further study is needed to overcome those forces in order to obtain the full hysteresis cycle.

A complementary transient analysis was performed to calculate the time at which the suspended control gate collapses on the tunnel oxide layer. In this numerical analysis, the dimensions used were those obtained from the double-clamped beam. The analysis include a 3D FEM and were implemented in Comsol. As a result of this implementation, it was found that the double-clamped structure took $\approx$ 26 nsec to pull-in into the tunnel oxide layer.

In a similar way, an array of five beams were characterised obtaining a rather interesting pull-in curve. Once the set of beams were handled on the C - V meter, a similar process as for the single beam analysis was follow. In this analysis we used the maximum span available in the equipment. The input data included a span of voltage form 0 to 38 V, a step of 0.5 V and a high integration rate. The beams under analysis were longer (102 µm) than the previous single beam. As a result of the array under test, a C - V pull-in curve was obtained (Fig. 6.16).

While the array is biased, there is no-displacement present from the set of beams till

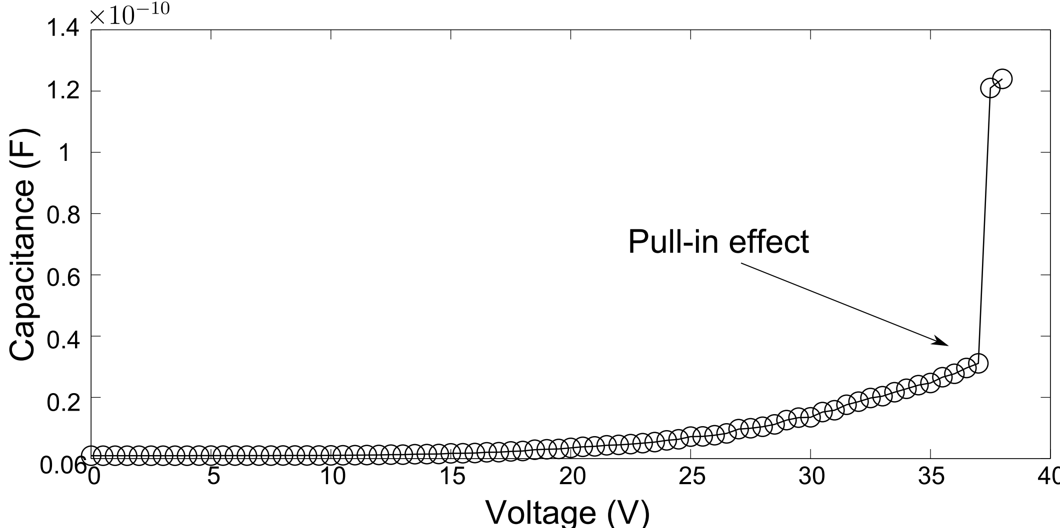


FIGURE 6.16: Pull-in effect obtained as a result of the electrical C - V measurements of a set of double-clamped beams

the applied voltage is considerably large (> 20 V). From this voltage till 35 V, the set of beams bend downwards until the pull-in point is reached. Once the pull-in point has been reached ($\approx$ 36.5 V) by the array of beams, the set of beams collapsed at once as clearly stated in Fig. 6.16. After the maximum voltage, allowed by the equipment, was reached, the applied voltage switch the bias of the voltage and start to sweep from 38 V to 0 V. After the applied voltage was reduced below the pull-in voltage, the set of beams did not show any variation in the capacitance value which indicates that the beam array did not pull-out in a similar way as in the single beam. Further work is needed to find the reason why the short-range forces are so strong that the beams did not return to its initial isolated position.

According to Fig. 6.15 we can see that the double-clamped beam structure start bending downwards at $\approx$ 6 V. When the applied voltage is increased, the beam bends further and the pull-in effect by action of the electrostatic force is produced at $\approx$ 9 V. In order to test the accuracy of the 3D FEM numerical analysis, the same structure is introduced. For the sake of comparison between these analysis, Eq. (3.5) is also used.

Recalling that Eq. (3.5) is based on the simplest model to represent the double-clamped structure (double-plate capacitor model), which only considers the upper plate material, thickness and the initial air-gap as main conditions. The voltage obtained by using this model is far from reality. Nevertheless, by analysing the structure with a more complex and reliable numerical analysis such as the 3D FEM, the result obtained by using this analysis shows an improved result closer to the value measured but not quite the same. This occurs due to the 3D FEM analysis which considers the double-clamped beam as an homogeneous structure, whilst in the real structure several variations occur which are produced by the different steps in which the structure is involved. Those variations are localised mainly in the fabrication process such as beam thickness, clamp position of

the beam and width. These perturbations in the structure modify the original double-clamped structure. Taking into account these variations, these are fed into the 3D FEM analysis. As a result of the numerical analysis (see Fig. 6.17), the pull-in voltage obtained is ≈ 9 V.

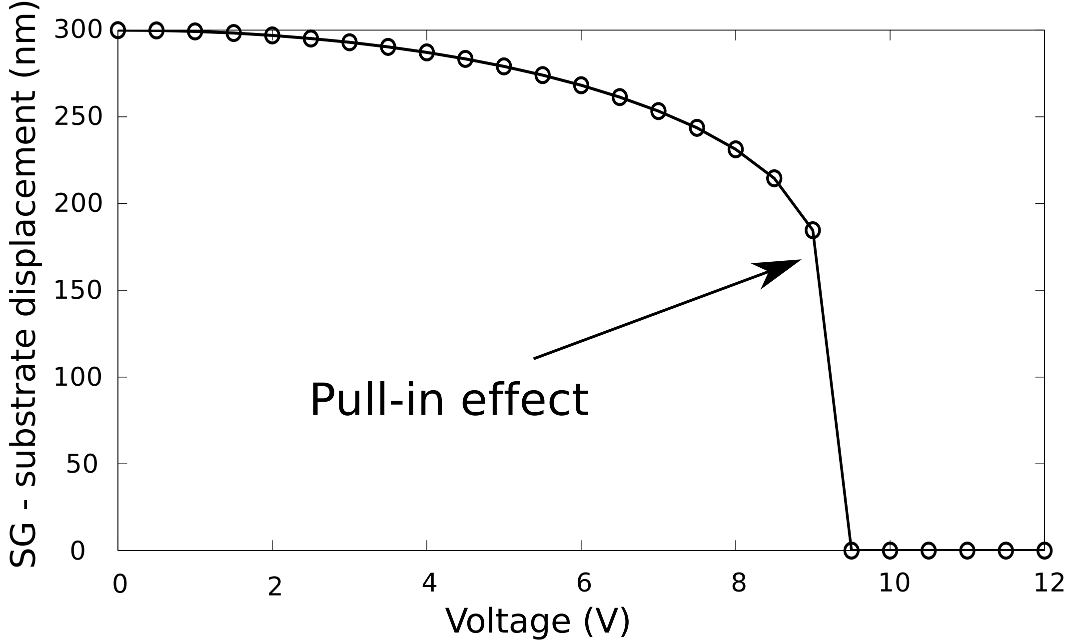


FIGURE 6.17: Pull-in analysis obtained by a numerical simulation once the structural parameters of the double-clamped beam structure have been updated

As result of the numerical processes mentioned above, Table 6.5 summarises the set of pull-in voltages as well as the value measured from the single beam structure.

TABLE 6.5: Comparison of the double-clamped beam structure while analysed by using a simple model, 3D FEM, *in situ* measured and an up-dated 3D FEM analysis

Method	Length (μm)	Width (μm)	Thickness (μm)	Pull-in (V)
Eq. (3.5)	58	5	0.5	3.76
3D FEM	58	5	0.5	5.82
Measurement	58	5	0.5	9
Improved 3D FEM	59.3	4.8	0.6	≈ 9

The set of pull-in results listed in Table 6.5 describes the accuracy obtained by using different set of analysis in comparison to the measurements. The pull-in voltage obtained from the double-plate capacitor model only considers the top plate material (Young's modulus), thickness, area and the initial air-gap as key parameters, while for the 3D FEM analysis, the software uses libraries that contain electro-mechanical properties for plenty of materials based on *in situ* measurements such as Al, Si, poly-Si, SiO_2, Si_3N_4, Cu, etc. When both analyses are performed, the 3D FEM analysis shows a better approximation of the measurements than the double-plate capacitor model. The

pull-in voltage on the FEM analysis is obtained due to a set of conditions such as the libraries, material analysis and the electro-mechanical analysis of the double-clamped beam structure overall, by taking into account the fabrication process involved to create the suspended control gate, modify the original set of parameters such as thickness of the suspended gate, dimensions of the suspended beam and the initial air-gap. The changes that occur in the double-clamped beam structure listed in Table 6.5, show that the beam length increases by more than a micron and the width is reduced by 200 nm. As a consequence of the multiple etching processes, a thin layer of aluminium oxide was formed on top of the aluminium beam. Also, by considering that the substrate influence the functionality of the device, the variations on the double-clamped beam structure as well as the inclusion of the substrate were introduced in a new numerical analysis by using ConventorWare and Comsol. The 3D FEM analysis shows that the double-clamped beam structure collapses over 9 V, similar to the value obtained in the measurement. Moreover, by implementing the entire structure composed by the readout element with the suspended control gate in Comsol, the numerical analysis shows that the suspended control gate bends downwards, thus generating the pull-in effect. The double-clamped beam structure cannot collapse on the tunnel oxide layer due to inherent software drawbacks, nevertheless, the pull-in effect is depicted and occurs at ≈ 9 V as depicted in Fig. 6.18

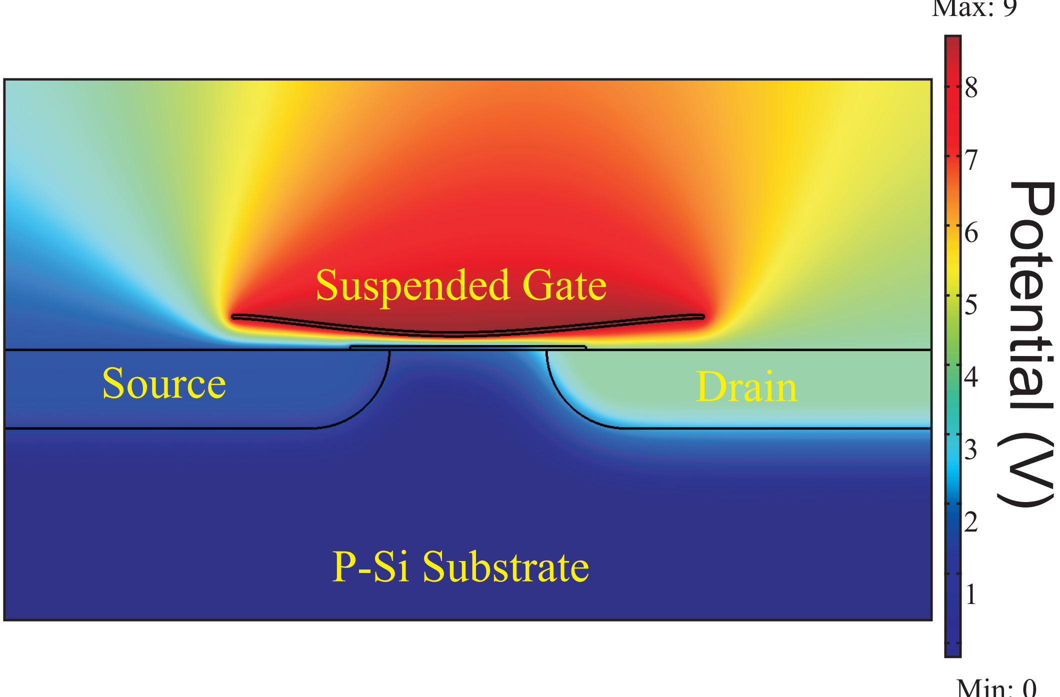


FIGURE 6.18: 2D FEM analysis in which the readout element is included for more accurate pull-in voltage estimate

On the other hand, once the double-clamped beam structure has collapsed on the tunnel oxide layer and the applied voltage was reduced, the beam structure did not return to its isolated flat position but remains attached to the tunnel oxide layer even when the ap-

plied voltage was lower than the pull-in voltage. This effect occurred due to the stiction generated was stronger than expected between the beam material and the tunnel oxide layer. This unexpected large stiction was caused by the short range forces such as the van der Waals and Casimir forces. As a consequence of this large stiction, no reasonable pull-out effect was obtained by using this particular double-clamped beam structure. Nevertheless, we are working on the improvement of the recipes in the fabrication process, in order to obtain the pull-out effect and as a consequence the capacitance - voltage hysteresis cycle.

Chapter 7

Conclusions and Future Work

In this thesis, the co-integration of two technologies such as the nano-electromechanical systems with the metal-oxide-semiconductor technology were successfully implemented within a numerical circuit simulation in order to create a new hybrid non-volatile memory device known as the Suspended Gate Silicon Nanodot Memory. The numerical implementation of both technologies was divided between the suspended control gate and the quantum mechanical tunnelling process. The first analysis was used to investigate the electro-mechanical characteristics of the suspended control gate such as the pull-in and pull-out effects, C - V hysteresis curve, power consumption and transient analysis. These analyses were implemented as a 2D and 3D FEM analysis via a commercial software such as Comsol and CoventorWare. The second analysis was used to analyse the quantum-mechanical tunnelling process for the charging and discharging processes of the memory device. To perform this analysis, a homemade code whose foundation is defined by the Tsu-Esaki equation and the transfer matrix method was used to calculate the current density curve. This code uses FEM as the main kernel. The curves obtained from these analyses were transformed into analog voltage dependent functions and implemented as behavioural models within an electric circuit simulator such as SmartSpice. As a result of these analyses, it was found that the time at which the suspended control gate collapses on the tunnel oxide layer occurs at 0.8 nsec. The time that it takes for charging and discharging the memory node was found to be 1.7 nsec. Therefore, the total time that the Suspended Gate Silicon Nanodot Memory takes for programming and erasing the memory node is 2.5 nsec. This time is at least three magnitude orders faster than nowadays' flash memory and ten times faster than the "fastest" emerging non-volatile memory.

In order to fabricate the Suspended Gate Silicon Nanodot Memory device, the suspended control gate was designed and fabricated by using the substrate provided by our counterparts at Hitachi CRL. On top of the stack of layers (Si/SiO_2/poly-Si) provided, a layer of aluminium (500 nm) was deposited by using electron beam deposition. The control layer was patterned with doubly-clamped beam features through a photolithographic process.

As a consequence of the recipes developed for this process, features 700 nm in width and 26 μm in length were successfully obtained. A wet etching process was implemented to remove the aluminium surplus on the sample surface. To remove the sacrificial layer without severely affecting the aluminium beam features nor the substrate, a single-step plasma process was developed to dry etch isotropically the sacrificial layer. The recipe developed offers high selectivity between several Si-based materials such as poly-Si, SiO_2 and Si. Once the doubly-clamped beam structures are suspended, the pull-in effect is investigated through the capacitance variation by using a C - V meter. For a structure of 26 μm at 5 μm with an air-gap of 300 nm, the structure shows a pull-in effect of 9 V. This value was corroborated through a numerical 2D/3D FEM analysis showing a good agreement with the value measured.

7.1 Future Work

To fabricate the floating gate layer and the memory device, it is necessary to firstly get rid of the sacrificial layer present on the foundation samples. Once removed, the process developed by the Tokyo Institute of Technology is implemented and the silicon nanodots are ordered in a monolayer as previously shown. Once the silicon nanodot monolayer has been obtained, the sample is oxidised to form an isolation shell around the dots. On top of it, a thin oxide layer is grown which will work as the tunnel oxide layer. After this deposition, the sacrificial layer is deposited, in which several materials discussed above are suggested. Finally, by following the fabrication process of the suspended gate explained in this Thesis, a thin aluminium layer is deposited by using e-beam, it is patterned by using the photolithographic process (described elsewhere) and etched out by using the wet etching process. It is here that the trenches are opened to make the holes and contact pads with source, drain and back contact electrodes by using the lift-off process. Once the electrodes have been fabricated, the sacrificial layer is etched out. To remove the sacrificial layer (poly-Si), the recipe developed in this Thesis is reliable, repeatable and easy to implement.

In order to improve the fabrication of the Suspended Gate Silicon Nanodot Memory device, several parameters such as sacrificial layer material, etching processes and beam material must be considered to obtain better results during the fabrication of the non-volatility memory device. These parameters are:

Sacrificial Layer Material

In order to avoid any damage to the doubly-clamped beam features and tunnel oxide layer, the sacrificial layer has to be easy to etch away. Poly-Si is not a standard sacrificial material and it is not easy to remove without compromising other Si-based material such

as SiO_2. Sacrificial materials that can be used within the SGSNM are: Si-Ge, α-Si or Ge. To etch those materials, there are several recipes which have already been developed in our facilities which show high selectivity between Al and SiO_2.

Etching Processes

The etching processes are linked to the material considered as a sacrificial layer. The single-step dry-etching plasma process developed to remove the poly-Si layer shows that it is possible to tune the already developed recipes for a particular sacrificial layer material or develop another one based on a different set of gases such as XeF_2. This compose is effective for Si-based materials such as SiO_2 or α-Si. In our cleanroom, there is an equipment that can be used to etch-out those kinds of sacrificial layer materials if new recipes are required.

Doubly-Clamped Beam Design

By analysing the set of doubly-clamped beam features of the first photomask, we learnt that by designing beams with squared clamps, these became semi-circular due to the etching processes that are performed during the fabrication process. Therefore, a new batch of clamps is required in order to avoid the roundness of the clamps. It was also found from the previous photomask that thin beams are easy to remove when etch processes are performed. Hence, the widths of the beams are linked to their lengths as depicted in Table 7.1

Table 7.1: Set of parameters required to design the new photomask

Feature	Length (µm)	Width (µm)	Thickness (nm)	Air-gap (nm)	Sacrificial layer
1	105	5, 4, 3.5, 2.5	500, 300	300	α-Si, Ge, SiO_2
2	58	5, 4, 3.5, 2.5, 1.5	500, 300	300	α-Si, Ge, SiO_2
3	25	3, 2.5, 1.5, 0.8	500, 300	300	α-Si, Ge, SiO_2

Pull-out effect

After the single-plasma etching process has been performed and the characterisation process has occurred, the pul-out effect is foreseen. In order to obtain this effect, a deep analysis is required y using the material composition by using SIMS. By using this method, it is possible to find the materials that conform the substrate (tunnel oxide layer/sacrificial layer) and by means of this analysis we can infer how to overcome the short-range forces that retain the tunnel oxide layer attached to the tunnel oxide layer and allows the pull-out effect.

Once the Suspended Gate Silicon Nanodot Memory device has been fabricated, the characterisation of the memory device is foreseen. The set of analysis proposed to characterise the memory device are:

1. Electromechanical analysis such as C - V hysteresis curve
2. Electrical measurements such as I - V curve
3. Mechanical switching analysis
4. Endurance analysis
5. Non-volatility analysis

Through the improvement of the fabrication process and the characterisation of the SGSNM, it is expected to extract a model to compare the numerical analyses (2D/3D FEM) with the curves obtained from the characterisation of the memory device. Here, the stiction caused by the short range forces due to the interaction of the doubly-clamped beam structure, the sacrificial layer material residues and the tunnel oxide layer are the key factors in the analysis.

Appendix A

Calibration for the C - V Meter

In the front panel of the Agilent 4279A @ 1 MHz C - V meter, there are four BNC output connectors which are arranged into two "T" connectors. Those connectors are plugged into the interconnection panel which are interconnected to the probes on the probe stage system. In order to calibrate the equipment, the sample is handled on the central chuck of the holder which has a central hole for central vacuum. Once the sample has ben handled on the chuck, the C - V meter is calibrated by using a state-of-the-art kit. Calibration process eliminates the capacitance, inductance and resistivity inherent to the connectors, probes & wires from the C - V meter till the test probes which also involves the chuck holder. Figure A.1 shows the interconnection between the C - V meter and the test probe system through the interconnection panel.

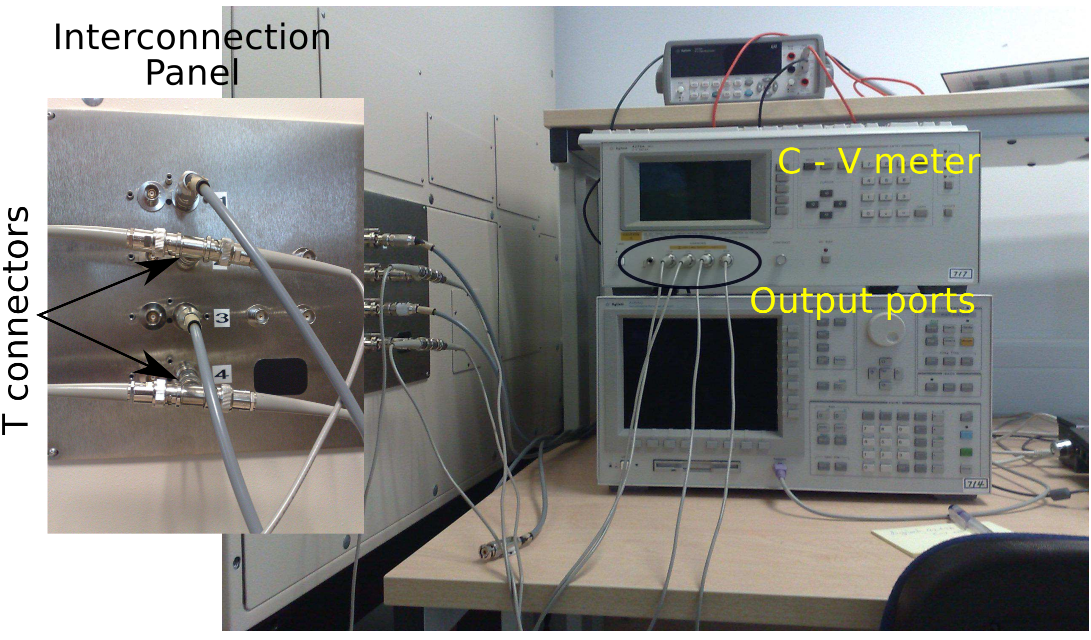


FIGURE A.1: Interconnection between the C - V meter with the test probe system through the interconnection panel

In the rear part of the C - V meter, a GPIB connector is plugged to control the equipment via PC. In here, the calibration sequence is executed. Calibration requires to perform measurements in short mode and in open circuit mode. In the probe station, one of the probes is positioned on the sample under test and the other on the substrate (chuck). For the short analysis, both are positioned over the chuck surface and for the open loop, both are isolated from the stage and between them. After the calibration process has been performed, the probes are positioned on the chuck and over the device under test (a beam or a set of beams). Figure A.2 shows the stage in which the sample is handled. In here, one of the probes is used as the ground on the device under test (100 μm) and the other is energised on the chuck. Once in position, the analysis starts by defining a set of parameters in the system such as initial and end voltages, step, tune, among others. Once those parameters have been specified, the C - V measurements are performed.

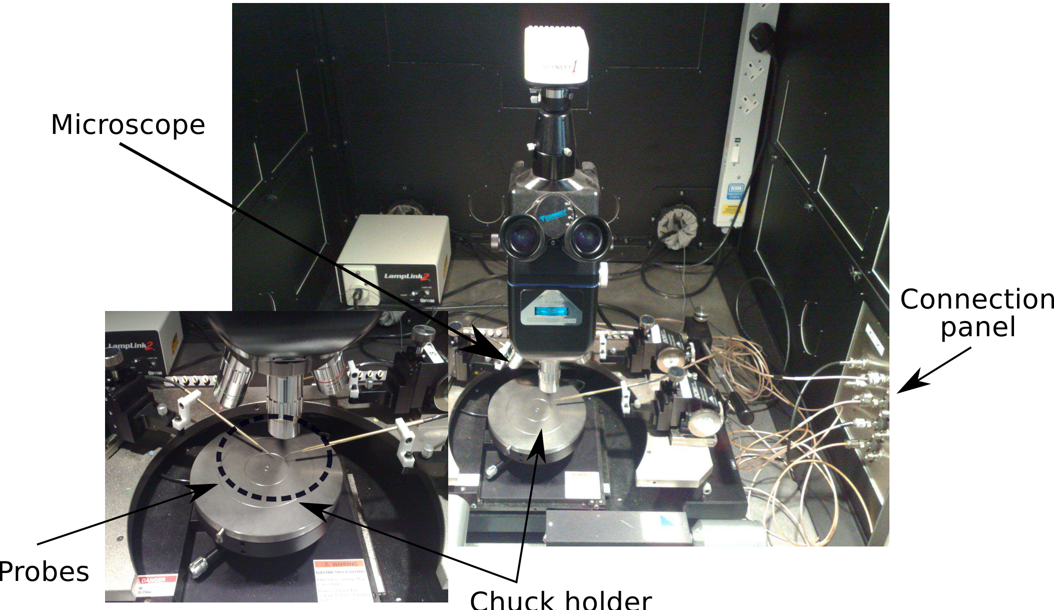


Figure A.2: Stage probe system and the array used to perform the C - V measurements

Appendix B

Photomask Design

The fabrication process of the Suspended Gate Silicon Nanodot Memory cell uses three photomasks to fabricate the suspended control gate, as it is depicted in Fig. B.1.

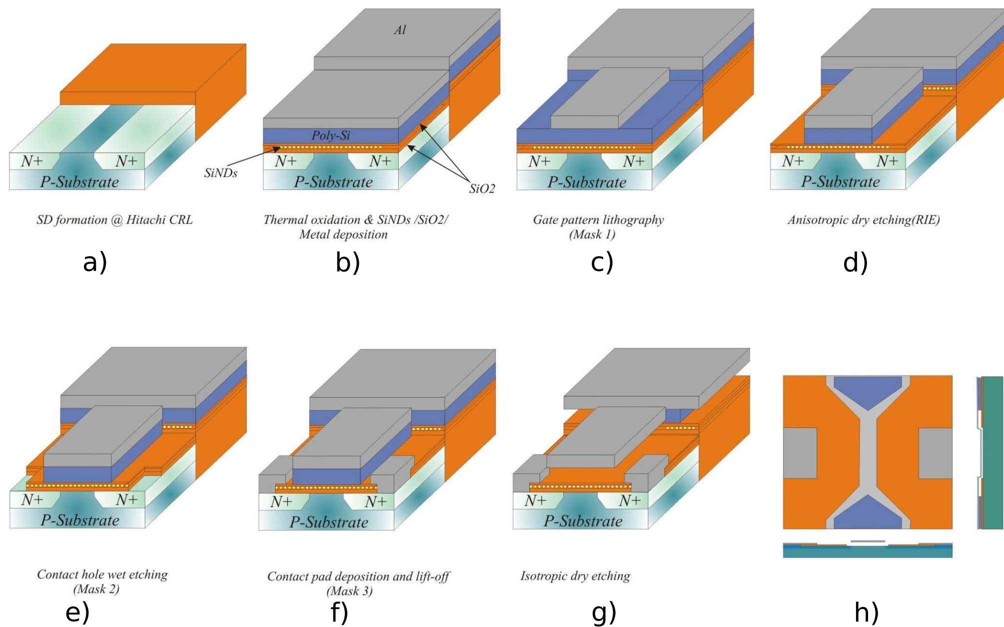


FIGURE B.1: Suspended Gate Silicon Nanodot Memory cell fabrication process. a) Memory device foundation, b)readout element, silicon nanodot floating gate, sacrificial layer and metal deposition layers; c) suspended gate patterned and wet etched (mask 1); d) by using anisotropic dry etching process, the sacrificial layer is anisotropically removed; e) Source-Drain contact holes are created using wet etching (mask 2); f) Contact pad deposition and lift-off (mask 3); g) finally, by using isotropic dry etching the sacrificial layer is removed and the beam is suspended, h) top, side and lateral views of the SGNM cell

The photomasks contains features to fabricate the doubly-clamped beam structures such as:

- Doubly-clamped beam features
- Holes creation
- Pads

The photomask that includes the doubly-clamped beam structures has a broad range of beam elements. The length of the beam elements starts from 25 μm until 200 μm and the width of those elements starts at 700 nm until 5 μm as listed on Table B.1.

TABLE B.1: Dimensions used in the photomasks for the length and width of the beams, respectively, as well as the pad dimensions

Set of Beams	Length (μm)	Width (μm)	Pads (μm^2)
1	25	0.8, 1.5, 2.5 & 3	80 × 80
2	58	1.5, 2.5, 3.5, 4 & 5	80 × 80
3	105	2.5, 3.5, 4 & 5	80 × 80

The set of dimensions listed in Table B.1 are implemented in a layout editor based on windows. Within the photomask designed in L - edit, it contains the set of beam dimensions, the pads for source, drain and back contact and the holes to interconnect the pads with the highly-doped regions (see Fig. 6.1). A schematic diagram of the Suspended Gate Silicon Nanodot Memory cell in which the three photomasks are included is depicted in Fig. B.2.

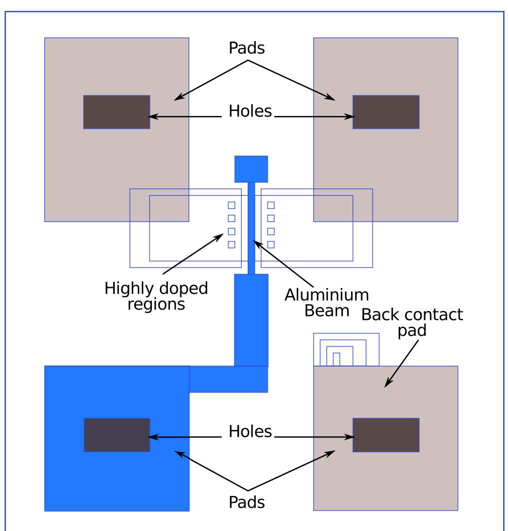


FIGURE B.2: Schematic diagram of the SGSNM cell that shows the three photomask such as pads, doubly-clamped beam features and the holes for the highly-doped regions

The photomask is divided as 1 × 1 cm^2 samples. Figure B.4 shows the photomask designed that includes the beam features and the control gate pad.

Dimensions
Width
Length
1.25 µm
2.5 µm
4 µm
5 µm
700 nm
26 µm
58 µm
105 µm

FIGURE B.3: Set of parameters that shows the length and width of the doubly-clamped beam features

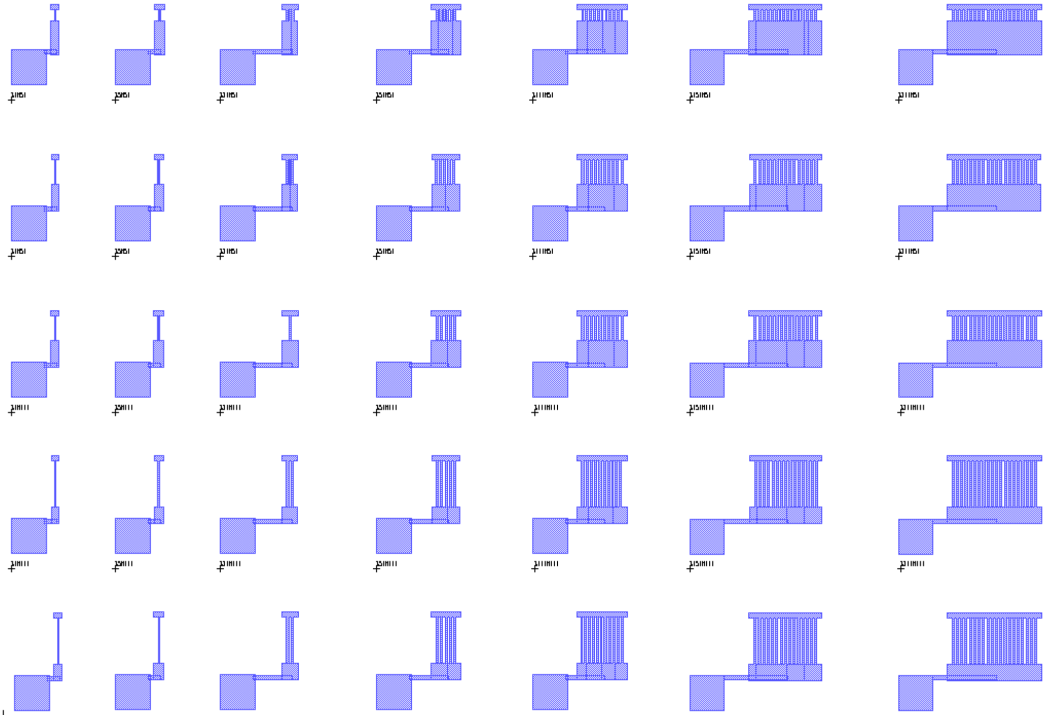


FIGURE B.4: Mask of the doubly-clamped beam features

Bibliography

[1] International Technology Roadmap for Semiconductors, ITRS2010 Edition. Technical report, ITRS, 2010.

[2] T Ifuku, M Otobe, A Itoh, and S Oda. Fabrication of Nanocrystalline Silicon with Small Spread of Particle Size by Pulsed Gas Plasma. *Jpn. J. Appl. Phys*, 36:4031, 1997.

[3] Atsushi Tanaka, Yoshishige Tsuchiya, Koichi Usami, Shin-ichi Saito, Tadashi Arai, Hiroshi Mizuta, and Shunri Oda. Synthesis of Assembled Nanocrystalline Si Dots Film by the Langmuir–Blodgett Technique. *Japanese Journal of Applied Physics*, 47(No. 5):3731–3734, 2008.

[4] Ashok K Sharma. *Advanced Semiconductor Memories Architectures, Designs, and Applications*. 2003.

[5] Chengming Hu, editor. *NonVolatile Semiconductor Memories, technologies, design and applications*. IEEE Press, 1991.

[6] Tegze P Haraszti. *CMOS Memory circuits*. Kluwer Academic Publishers, 2000.

[7] Murray H Woods. An EPROM's integrity starts with its cell structure, August 1980.

[8] M Momodomi, R Kirisawa, R Nakayama, S Aritome, T Endoh, Y Itoh, Y Iwata, H Oodaira, T Tanaka, M Chiba, R Shirota, and F. Masuoka. New device technologies for 5V-only 4Mb EEPROM with NAND structure cell. *IEDM*, pages 412–415, 1988.

[9] F. Masuoka, M Asano, H Iwahashi, T Komuro, N Tozawa, and S Tanaka. a 256-kbit Flash E2PROM Using Triple-polysilicon Technology. *IEEE J. Solid-State Circuits*, SC-22:548–552, 1987.

[10] Http://www.arlabs.com Labs, Andromeda Research. The EPROM+, 2010.

[11] P Pavan, R Bez, P Olivo, and E Zanoni. Flash Memory cells - An overview. *Proceedings IEEE*, 85:1248, 1997.

[12] F. Masuoka. Are you ready for next-generation dynamic RAM chips? *IEEE Spectrum*, 27(11):110–112, 1990.

[13] B. Prince. *Emerging memories: technologies and trends*. Springer Netherlands, 2002.

[14] G Groeseneken, HE Maes, J. Van Houdt, and JS Witters. *Basics of nonvolatile semiconductor memory devices*. 1971.

[15] S Tam, P. K. Ko, and C Hu. Lucky-Electron Model of Channel Hot Electron injection in MOSFET's. *IEEE Trans. on Elect. Dev.*, 31:1116–1125, 1984.

[16] K Sonada, M Tanizawa, S Shimiza, Y Araki, S Kawai, T Ogura, S Kobayashi, K Ishikawa, and Y Inoue. Compact Modeling of Flash Memory Cells Including Substrate-Bias-Dependednt Hot-Electrons Gate Current. In *SISPAD*, pages 215–218, 2003.

[17] M. Lenzlinger and EH Snow. Fowler-Nordheim tunneling into thermally grown Si02. *Journal of Applied Physics*, 40(1):278–283, 1969.

[18] M Geller, a Marent, T Nowozin, and D Bimberg. Self-organized quantum dots for future semiconductor memories. *Journal of Physics: Condensed Matter*, 20(45):454202, 2008.

[19] M. Depas, B. Vermeire, PW Mertens, RL Van Meirhaeghe, and MM Heyns. Determination of tunnelling parameters in ultra-thin oxide layer poly-Si/SiO2/Si structures. *Solid State Electronics*, 38(8):1465–1471, 1995.

[20] S Haddad, C Chang, B Swaminathan, and J Lien. Degradations due to hole trapping in Flash Memory cells. *IEEE Electron Device Letters*, 10:117–119, 1989.

[21] Stefan Lai. Non-volatile memory technologies: The quest for ever lower cost. *2008 IEEE International Electron Devices Meeting*, 2007(Figure 2):1–6, December 2008.

[22] Chih-Hao Cheng and Joseph Ya-Min Lee. Metal-high-κ-high-κ-oxide-semiconductor capacitors and field effect transistors using Al La_2O3 Ta_2O_5 SiO_2 Si structure for nonvolatile memory applications. *Applied Physics Letters*, 91(19):192903, 2007.

[23] J.J. Welser, S. Tiwari, S. Rishton, K.Y. Lee, and Y. Lee. Room temperature operation of a quantum-dot flash memory. *IEEE Electron Device Letters*, 18(6):278–280, June 1997.

[24] G D Wilk, R M Wallace, and J M Anthony. high k-gate dielectrics: Current status and material properties considerations. *Applied Physics Review*, 89:5243, 2001.

[25] S. Tiwari, F. Rana, K. Chan, H. Hanafi, W. Chan, and D. Buchanan. Volatile and Non-Volatile Memories in Silicon with Nano-Crystal Storage. In *IEDM*, pages 521–524. IEEE, 1995.

[26] A Tanaka, Yoshishige Tsuchiya, G Yamahata, Hiroshi Mizuta, Shunri Oda, and K Usami. High-density assembly of nanocrystalline silicon quantum dots. *Current Applied Physics*, 6(3):344–347, 2006.

[27] R Rao, R Steimle, M Sadd, C Swift, B Hradsky, S Straub, T Merchant, M Stoker, S Anderson, and M Rossow. Silicon nanocrystal based memory devices for NVM and DRAM applications. *Solid-State Electronics*, 48(9):1463–1473, 2004.

[28] S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E.F. Crabbe, and K. Chan. A silicon nanocrystals based memory. *Applied Physics Letters*, 68(March):1377, 1996.

[29] G. Muller, N. Nagel, C.-U. Pinnow, and T. Rohr. Emerging non-volatile memory technologies. *ESSCIRC 2004 - 29th European Solid-State Circuits Conference (IEEE Cat. No.03EX705)*, pages 37–44, 1995.

[30] Andy Chung, Jamal Deen, Jeong-Soo Lee, and M Meyyappan. Nanoscale memory devices. *Nanotechnology*, 21(41):412001, October 2010.

[31] Faezeh Arab Hassani, C Cobianu, S Armini, V Petrscu, P Merken, Dimitrios Tsamados, Adrian Mihai Ionescu, Yoshishige Tsuchiya, and Hiroshi Mizuta. Design and Analysis of an In-Plane Resonant Nano-Electro-Mechanical Sensor for Sub-Attogram-Level Molecular Mass-Detection. In *2009 International Conference on Solid State Devices and Materials (SSDM 2009)*, 2009.

[32] Tasuku Nagami, Hiroshi Mizuta, N. Momo, Y. Tsuchiya, S. Saito, Tadashi Arai, T. Shimada, and Shunri Oda. Three-dimensional numerical analysis of switching properties of high-speed and non-volatile nanoelectromechanical memory. *IEEE Trans. on Elect. Dev.*, 54:1132–1139, 2007.

[33] Hiroshi Mizuta, Tasuku Nagami, Mario A Garcia-Ramirez, Jun Ogi, Benjamin Pruvost, H. Yoshimura, Yoshishige Tsuchiya, and Shunri Oda. Co-integration of silicon nanodevices and NEMS for advanced information processing. *2008 9th International Conference on Solid-State and Integrated-Circuit Technology*, pages 2375–2378, October 2008.

[34] E.J. Prinz. The zen of nonvolatile memories. *2006 43rd ACM/IEEE Design Automation Conference*, pages 815–820, 2006.

[35] G. Klimchitskaya, U. Mohideen, and V. Mostepanenko. Casimir and van der Waals forces between two plates or a sphere (lens) above a plate made of real metals. *Physical Review A*, 61(6), May 2000.

[36] K A Milton. *The Casimir Effect, Physical Manifestations of Zero-Point Energy.* World Scientific Publishing, 2001.

[37] R. H. S. Winterton. Van der Waals forces. *Contemporary Physics*, 11(6):559–574, 1970.

[38] Benjamin Pruvost, Hiroshi Mizuta, and Shunri Oda. 3-D Design and Analysis of Functional NEMS-gate MOSFETs and SETs. *IEEE Transactions On Nanotechnology*, 6(2):218–224, March 2007.

[39] Nicolas Abelé, R Fritschi, K Boucart, F Casset, P Ancey, and Adrian Mihai Ionescu. Suspended-Gate MOSFET: bringing new MEMS functionality into solid-state MOS transistor. In *IEDM, Late news*, volume 00, pages 479–481, 2005.

[40] Y. Nishi and R. Doering. *Handbook of semiconductor manufacturing technology*. CRC Press, 2000.

[41] Mohamed Gad-el Hak, editor. *The MEMS Handbook, MEMS: Design and Fabrication*. Taylor & Francis Group, 2006.

[42] S Senturia D. *Microsystem Design*. Kluwer Academis, 2001.

[43] Comsol Multiphysics, 2008.

[44] CoventorWare, Designer Reference, 2008.

[45] D. FitzPatrick and I. Miller. *Analog behavioral modeling with the Verilog-A language*. Kluwer Academic Pub, 1998.

[46] Keneth S Kundert and Olaf Zinke. *The Designer's Guide to Verilog AMS*. Kluwer Academic Publishers, 2004.

[47] Mario A Garcia-Ramirez, Yoshishige Tsuchiya, and Hiroshi Mizuta. Suspended Gate Silicon Nanodots Memory. In *ESSDERC/CIRC Fringe (ESS-Fringe)*, Edinburgh, 2008.

[48] Jozua Laven and Johannes P C Vissers. The Hamaker and the Lifshitz approaches for the Van der Waals interaction between particles of composite materials dispersed in a medium. *Colloids and Surfaces*, 152:345 – 355, 1999.

[49] H. Hamaker. The London-van del Waals attraction between spherical particles, October 1937.

[50] a Anandarajah. Single Correction Function for Computing Retarded van der Waals Attraction. *Journal of Colloid and Interface Science*, 176(2):293–300, December 1995.

[51] R K Gupta. *Electrostatic pull-in test structure design for in-situ mechanical property measurements of microelectromechanical systems (MEMS)*. PhD thesis, Massachusetts Institute of Technology, 1997.

[52] Saeed Maoveni. *Finite Element Analysis: Theory and Application with Ansys*. Prentice Hall; 1st edition, 1999.

[53] Gunnar Backstrom. *Simple Field of Physics*.

[54] S. M. Sze. *Physics of Semiconductor Devices*. Wiley, 1981.

[55] R Singh, DL Harame, and M.M. Oprysko. *Silicon germanium: Technology, modeling, and design*. Wiley-IEEE Press, 2004.

[56] S. Inc. *Star-HSPICE Manual*. Number June. Avant, 1999.

[57] *SmartSpice, Manual*. Silvaco, 2008.

[58] Cadence, PSpice A/D and Advanced Analysis.

[59] R. van de Haar, J. Hoekstra, and R.H. Klunder. A SPICE model for single electronics. *2002 IEEE International Symposium on Circuits and Systems. Proceedings (Cat. No.02CH37353)*, pages V–868–V–871, 2002.

[60] Arturo Librado Sarmiento Reyes. *A partition method for the determination of multiple DC operating points*. PhD thesis, DELFT, 1994.

[61] S Saha, U Hanke, G U Junsen, and T Saether. Modelling of Spring constant and pull-down Voltage of Non-uniform RF MEMS Cantilever Incorporating Stress Gradient. *Sensors and Transducers Journal*, 98:54, 2008.

[62] S Chowdhury, M Ahmadi, and W Miller. A closed-form model for the pull-in voltage of electrostatically actuated cantilever beams. *J. Micromech. Microeng.*, 15:756, 2005.

[63] Using Verilog-A in Advanced Design System. Technical Report September, 2004.

[64] Verilog-A Reference Manual, Agilent Technologies. Technical Report December, Agilent Technologies, 2003.

[65] V. Adrian Parsegian. *Van der Waals forces: A handbook for Biologists, chemists, engineers and physicists*, volume 131. Cambridge University Press, October 2006.

[66] Hiroshi Mizuta and Tomonori Tanoue. *The physics and applications of resonant tunnelling diodes*. Cambridge University Press, 1995.

[67] R. Tsu and L. Esaki. Tunneling in a finite superlattice. *Applied Physics Letters*, 22(11):562–564, 1973.

[68] L. L. Chang, L Esaki, and R. Tsu. Resonant tunneling in semiconductor double barriers. *Applied Physics Letters*, 24(12):593, 1974.

[69] Wei Li. Generalized Free Wave Transfer Matrix Method for Solving the Schrödinger Equation With an Arbitrary Potential Profile. *IEEE Journal of Quantum Electronics*, 46(6):970–975, June 2010.

[70] C F Huang, S D Chao, D R Hang, and Y C Lee. Chinese journal of physics vol. 46 , no. 3 june 2008. *Chinese Journal of Physics*, 46(3):14–17, 2008.

[71] William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. *Numercial Recipes, Fortran77*. Cambridge University Press, 1992.

[72] Yusuke Shuto, Shuu'ichirou Yamamoto, and Satoshi Sugahara. Nonvolatile static random access memory based on spin-transistor architecture. *Journal of Applied Physics*, 105(7):07C933, 2009.

[73] M. Ryzhii and V. Ryzhii, editors. *Physics and Modeling of Tera-and Nano-devices*. World Scientific Pub Co Inc, 2008.

[74] G. Reimbold. Modified 1/f trapping noise theory and experiments in mos transistors biased from weak to strong inversion - influence of interface states. *IEEE Trans. Electron Devices*, ED-31(9):1190–1198, 1984.

[75] G. Krause, K. R. Hofman, M. F. Beug, T. Muller, and T. Mikolajick. 1/f noise analysis of a 75 nm twin-flash technology non-volatile memory cell. *Non-volatile Memory Symposium*, NVMTS2006:12–15, 2006.

[76] Tasuku Nagami, Yoshishige Tsuchiya, Ken Uchida, Hiroshi Mizuta, and Shunri Oda. Scaling Analysis of Nanoelectromechanical Memory Devices. *Japanese Journal of Applied Physics*, 49(4):044304, April 2010.

[77] Kellar Autumn, Metin Sitti, Yiching a Liang, Anne M Peattie, Wendy R Hansen, Simon Sponberg, Thomas W Kenny, Ronald Fearing, Jacob N Israelachvili, and Robert J Full. Evidence for van der Waals adhesion in gecko setae. *Proceedings of the National Academy of Sciences of the United States of America*, 99(19):12252–6, September 2002.

[78] J Thus and Polystyrene Polystynene. London-van der Waals Force. pages 1–7.

[79] Yu Tian, Noshir Pesika, Hongbo Zeng, Kenny Rosenberg, Boxin Zhao, Patricia McGuiggan, Kellar Autumn, and Jacob Israelachvili. Adhesion and friction in gecko toe attachment and detachment. *Proceedings of the National Academy of Sciences of the United States of America*, 103(51):19320–5, December 2006.

[80] Kosmas Galatsis, Kang Wang, Youssry Botros, Yang Yang, Ya-Hong Xie, J F Stoddart, R B Kaner, Mihri Ozkan, Chongwu Zhou, and Ki Wook Kim. Emerging Memory Devices, nontraditional possibilities based on nanomaterials and nanostructures. *IEEE Circuits and Devices*, 22(3):12–21, 2006.

[81] Chia-Wen Chang, Chih-Kang Deng, Che-Lun Chang, Ta-Chuan Liao, and Tan-Fu Lei. High-Performance Solid-Phase Crystallized Polycrystalline Silicon Thin-Film Transistors with Floating-Channel Structure. *Jpn. J. Appl. Phys*, 47:3024, 2008.

[82] Y. Tsuchiya, T. Iwasa, A. Tanaka, K. Usami, Hiroshi Mizuta, and S. Oda. Formation of an Ordered Array of nc-Si Dots by Using a Solution Droplet Evaporation Method. In *MRS Spring Meeting*, 2004.

[83] Wang Zhengfeng, Yin Ling, Ng Sum Huan, and Teo Phaik Luan. Chemical Mechanical Planarization. Technical report, Singapore Institute of Manufacturing Technology, 2001.

[84] www.entegrisfluidhandling.com Entegris. Chemical Mechanical Polishing (CMP) Slurry: Considerations for liquid flow control technologies, 2004.

[85] J. D. Ferguson, E. R. Smith, a. W. Weimer, and S. M. George. ALD of SiO_2 at Room Temperature Using TEOS and H_2O with NH_3 as the Catalyst. *Journal of The Electrochemical Society*, 151(8):G528, 2004.

[86] S Wolf and R N Tauber. *Silicon processing for the VLSI Era: Vol. 1 Process Technology 2nd Ed.* Lattice Press, 2000.

[87] Aluminium etching, 2009.

[88] A J van Roosmalen, J A G Baggerman, and S J H Brader. *Dry Etching for VLSI.* Plenum Press, updates in edition, 1991.