

UNIVERSITY OF LEEDS

DOCTORAL THESIS

---

**Towards the Mass Fabrication of Single  
Electron Transistors for Biosensing  
Applications**

---

*Author:*

Louis FRY-BOURIAUX

*Supervisor:*

Prof. Christoph WÄLTI

*A thesis submitted in fulfilment of the requirements  
for the degree of Doctor of Philosophy*

*in the*

Bioelectronics Group

Department of Electronic and Electrical Engineering

November 2017

# Declaration of Authorship

I, Louis FRY-BOURIAUX, declare that this thesis titled, 'Towards the Mass Fabrication of Single Electron Transistors for Biosensing Applications' and the work presented in it are my own. I confirm that:

- This work was 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.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

---

Date:

---

# *Abstract*

Doctor of Philosophy

by Louis FRY-BOURIAUX

The development of ultra sensitive charge sensing devices such as single-electron transistors (SETs) for next-generation biomedical applications has received considerable attention in the past few years. In this thesis, a potential approach for the mass-fabrication of metallic SETs for ultra-sensitive biosensing applications — an important prerequisite for early diagnosis of many serious diseases — is investigated. Using the orthodox theory of Coulomb blockade it is shown that it is possible to engineer an SET system that can satisfy the requirements for a highly sensitive charge sensor operating at room temperature while using metallic electrodes rather than semiconductor structures. In this configuration, the SET design and fabrication process is simplified greatly by lifting the dependence of the system on the confinement energy of electrons in the quantum dots (QDs), as is the case in semiconductor SETs. In return, this makes the tunnel junction properties and the geometrical arrangement of the islands and electrodes far more critical in determining the maximum operating temperature of the device. Here, the geometrical requirements for such a sensitive device are studied theoretically whilst the tunnel junction properties are studied experimentally and then theoretically to provide a thorough assessment of the abilities of the proposed SET system.

Atomic-layer deposition (ALD) has proven to be a highly reliable technique for depositing uniform thickness and reproducible thin metal-oxide films and particularly the  $\text{Al}_2\text{O}_3$  ALD process is known to be ‘ideal’ with highly reproducible properties. Here, a systematic study of the electronic properties of ALD deposited  $\text{Al}_2\text{O}_3$  thin films in MIM structures was performed to assess the ALD techniques applicability to the mass fabrication of quantum tunneling junctions for metallic SET structures. The two most crucial material parameters relevant to the design of metallic SET tunnel barriers are studied in detail; the dielectric constant of the film that determines the junction capacitance, and the properties of the potential barrier that mediates electron tunneling. Photolithographic techniques were used to create electrodes with a wide range of characteristic lengths as to provide a wide range of impedances. Measurements and subsequent analysis show that a high consistency can be attained over large surface areas in the film properties, and that electrode coverage is very effective, showing promise for mass-fabrication applications. Further analysis of the measurements shows that small static distortions in the barrier can affect the symmetry of MIM diode IV characteristics operating in the direct tunneling regime and that under certain circumstances the effect of surface states can be observed in the tunneling conductance.

# *Acknowledgements*

First I would like to thank my family for being so supportive throughout my life.

I am greatly indebted to Mark Rosamond for providing great support in the early days of this project, and his consistent and thorough approach to fabrication has been of great value in guiding the project to success.

I cannot express my gratitude enough for my project supervisor Christoph Wälti for always believing in me, supporting my work, and guiding me through the intellectual challenges faced during this project.

Finally I would like to express my gratitude to my friends and colleagues for always being open to listen and discuss the present project.

# Table of Contents

<b>Declaration of Authorship</b>	<b>ii</b>
<b>Abstract</b>	<b>iii</b>
<b>Acknowledgements</b>	<b>iv</b>
<b>Table of Contents</b>	<b>v</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>ix</b>
<b>Abbreviations</b>	<b>xi</b>
<b>Physical Constants</b>	<b>xv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 High sensitivity biosensors . . . . .	1
1.2 Aims and Objectives . . . . .	13
1.3 Thesis Structure . . . . .	15
<b>2 SET biosensors, design and limitations</b>	<b>17</b>
2.1 Introduction . . . . .	17
2.2 Theory of metallic SEDs . . . . .	17
2.2.1 Orthodox theory of Coulomb blockade . . . . .	17
2.2.2 Single and multi-island SET devices . . . . .	27
2.2.3 Charge sensitivity . . . . .	38
2.3 SET biosensor design . . . . .	43
2.3.1 Proposed system architecture . . . . .	43
2.3.2 SET optimisation for operating temperature . . . . .	48
2.3.3 SET optimisation for charge sensitivity . . . . .	50
2.4 Mass fabrication of metallic SETs with ALD . . . . .	58
2.5 Conclusion . . . . .	61
<b>3 Experimental Details</b>	<b>65</b>
3.1 Introduction . . . . .	65
3.2 Fabrication techniques and apparatus . . . . .	65
3.2.1 UV Photolithography . . . . .	65
3.2.2 Metallic thin film deposition . . . . .	70
3.2.3 Atomic layer deposition . . . . .	72
3.2.4 Packaging techniques . . . . .	78
3.3 Imaging techniques and apparatus . . . . .	79
3.3.1 Optical microscopy . . . . .	79

---

3.3.2	Electron microscopy techniques . . . . .	80
3.3.3	Scanning probe microscopy techniques . . . . .	84
3.3.4	Compositional imaging and mapping . . . . .	86
3.4	Electrical measurement techniques and apparatus . . . . .	89
3.4.1	Temperature control . . . . .	89
3.4.2	Electronic measurement systems . . . . .	92
<b>4</b>	<b>Fabrication Results and Characterisation</b>	<b>99</b>
4.1	Introduction . . . . .	99
4.2	MIM capacitor design and fabrication . . . . .	99
4.2.1	Design details . . . . .	100
4.2.2	Fabrication process . . . . .	103
4.3	Electrodes and ALD film characterisation . . . . .	109
4.3.1	AFM ALD-film characterisation . . . . .	109
4.3.2	TEM cross section characterisation . . . . .	112
4.3.3	AFM bottom electrode characterisation . . . . .	117
4.4	Compositional characterisation . . . . .	119
4.5	Device summary and conclusion . . . . .	121
<b>5</b>	<b>Measurement Results and Analysis</b>	<b>123</b>
5.1	Introduction . . . . .	123
5.2	Capacitance measurements . . . . .	123
5.2.1	Low-frequency measurements . . . . .	124
5.2.2	Capacitance-voltage measurements . . . . .	133
5.3	DC conductance measurements . . . . .	140
5.3.1	Tunneling thickness and area dependence . . . . .	140
5.3.2	Tunneling temperature dependence . . . . .	150
5.4	Conclusion . . . . .	153
<b>6</b>	<b>System Modelling</b>	<b>155</b>
6.1	Introduction . . . . .	155
6.2	Direct tunneling field enhancement . . . . .	155
6.3	Temperature dependence of the barrier parameters . . . . .	172
6.4	Dielectric constant . . . . .	176
6.5	Conclusion . . . . .	187
<b>7</b>	<b>Conclusion</b>	<b>189</b>
7.1	Future work . . . . .	192
7.2	Outlook . . . . .	194
	<b>Bibliography</b>	<b>197</b>

# List of Figures

1.1	Carcinogenesis diagram . . . . .	2
1.2	Point of care device concept . . . . .	3
1.3	Example ELISA . . . . .	5
1.4	Surface plasmon resonance biosensor . . . . .	6
1.5	SEM image of a SiNW array biosensor . . . . .	8
1.6	SiNW biosensor response . . . . .	9
1.7	FET vs SET transconductance . . . . .	10
1.8	Metallic SET device and shadow angle evaporation . . . . .	11
1.9	Si multi-island SET . . . . .	12
2.1	Elementary components . . . . .	18
2.2	Single-electron addition energy . . . . .	20
2.3	Single junction Coulomb blockade . . . . .	24
2.4	Single junction charge states . . . . .	26
2.5	Single electron transistor and band diagram . . . . .	28
2.6	Coulomb diamonds and oscillations . . . . .	30
2.7	Multi-island SET diagram . . . . .	32
2.8	Potential profile in a multi-island SET . . . . .	35
2.9	Coulomb blockade in a multi-island SET . . . . .	37
2.10	External charge noise model . . . . .	41
2.11	SET biosensor system architecture . . . . .	44
2.12	Structure of an alkylthiol SAM . . . . .	45
2.13	Multiplexing biosensor schematic . . . . .	47
2.14	Multi-island SET temperature optimisation . . . . .	49
2.15	Debye length in a buffer solution . . . . .	53
2.16	Trapped ions and defects in molecular layers . . . . .	54
2.17	Direct and floating gate coupling . . . . .	56
2.18	Floating gate with a control gate . . . . .	58
2.19	CMOS gate length versus technology node . . . . .	59
2.20	SET mass fabrication example with EUV resolution . . . . .	61
2.21	< 20 nm linewidth by EUV lithography . . . . .	62
3.1	Mask aligner system schematic . . . . .	66
3.2	Electron beam evaporation system schematic . . . . .	71
3.3	Adsorption kinetics in ALD reaction . . . . .	74
3.4	ALD system schematic . . . . .	75
3.5	Device packaging . . . . .	78
3.6	Scanning electron microscope diagram . . . . .	81
3.7	Transmission electron microscopy techniques . . . . .	83
3.8	Atomic force microscope schematic . . . . .	85
3.9	EELS and EDX core level spectroscopy . . . . .	87
3.10	4K cryostat system schematic . . . . .	90

---

3.11	Sample mount schematic . . . . .	90
3.12	High temperature measurement system schematic . . . . .	91
3.13	DC I–V measurement circuit diagram . . . . .	93
3.14	Lock-in circuit used for AC measurements . . . . .	95
3.15	AC and C–V measurement circuit diagram . . . . .	96
4.1	Device electrode mask patterns (inverted) . . . . .	100
4.2	Device etch mask patterns . . . . .	101
4.3	Plot of the range of capacitance values . . . . .	102
4.4	Edge-bead removal process schematic . . . . .	104
4.5	Optical images of completed devices . . . . .	108
4.6	AFM images of Al <sub>2</sub> O <sub>3</sub> on Au and SiO <sub>2</sub> surfaces . . . . .	110
4.7	AFM images of 5 nm and 3 nm ALD-Al <sub>2</sub> O <sub>3</sub> monitor substrates . . . . .	112
4.8	SEM and TEM image of the MIM device and sample . . . . .	113
4.9	TEM image analysis . . . . .	115
4.10	Summary of the device geometries . . . . .	116
4.11	High resolution AFM roughness analysis . . . . .	117
4.12	Statistical distributions of the slope and inclination angle . . . . .	119
4.13	STEM/EDX compositional mapping . . . . .	120
5.1	Impedance frequency dependence of Au-Al <sub>2</sub> O <sub>3</sub> -Cr devices . . . . .	125
5.2	Impedance frequency dependence of Au-Al <sub>2</sub> O <sub>3</sub> -Au devices . . . . .	126
5.3	Summary of measured capacitances . . . . .	128
5.4	Sample fitting results of Au-Al <sub>2</sub> O <sub>3</sub> -Au devices with different thickness . . . . .	129
5.5	Extracted permittivity and junction resistance . . . . .	131
5.6	C–V measurement results for 2 and 3 nm films . . . . .	135
5.7	Voltage coefficient of capacitance for 2,3,10,20, and 30 nm films . . . . .	136
5.8	Electric field dependence of the permittivity . . . . .	138
5.9	Summary of the J–E curves of each batch . . . . .	142
5.10	Summary of the $\rho$ –V curves of each batch . . . . .	143
5.11	MIM structure energy band diagram . . . . .	145
5.12	Fowler-Nordheim plot of 5 nm MIM device . . . . .	147
5.13	Asymmetry of 5 nm device and low bias resistivity summary . . . . .	148
5.14	2 nm device conductance temperature dependence . . . . .	151
6.1	MIM structure schematic and image . . . . .	156
6.2	Consistency in the current density of 2nm devices . . . . .	158
6.3	MIM morphology and energy band diagrams . . . . .	159
6.4	Field enhancement modelling of the potential . . . . .	161
6.5	Fitting results using the proposed model . . . . .	164
6.6	Relation of $n$ parameter to the surface geometry . . . . .	167
6.7	Schematic representations of asymmetric roughness . . . . .	169
6.8	Fitting results of I–V curves at 4 and 290 K . . . . .	174
6.9	Temperature dependence of the electron affinity . . . . .	175
6.10	Screening model of the bulk permittivity . . . . .	179
6.11	Ionic susceptibility electric field dependence . . . . .	183

# List of Tables

3.1	List of photolithography equipment . . . . .	69
3.2	Fiji F200 ALD system parameters . . . . .	77
4.1	ALD cycles and thickness . . . . .	106
4.2	Summary of fabricated devices . . . . .	121
5.1	Estimated errors in the AC analysis of the junctions . . . . .	130
6.1	Tunneling model fitting parameters . . . . .	165
6.2	Parameter correlations . . . . .	165
6.3	Phonon modes detected in Al <sub>2</sub> O <sub>3</sub> . . . . .	187



# Abbreviations

<b>2DEG</b>	<b>2-Dimensional Electron-Gas</b>
<b>AC</b>	<b>Alternating-Current</b>
<b>ADC</b>	<b>Analog-to-Digital Converter</b>
<b>ADF</b>	<b>Annular Dark-Field</b>
<b>AFM</b>	<b>Atomic-Force Microscopy</b>
<b>ALD</b>	<b>Atomic-Layer Deposition</b>
<b>APTMS</b>	<b>Aldehyde PropylTriMethoxySilane</b>
<b>CMOS</b>	<b>Complementary Metal Oxide Semiconductor</b>
<b>CNC</b>	<b>Computer-Numerical-Control</b>
<b>CVD</b>	<b>Chemical-Vapour Deposition</b>
<b>DC</b>	<b>Direct-Current</b>
<b>DFPT</b>	<b>Density-Functional Perturbation-Theory</b>
<b>DFT</b>	<b>Density-Functional Theory</b>
<b>DNQ</b>	<b>DiazoNaphthoQuinone</b>
<b>DSP</b>	<b>Digital-Signal-Processing</b>
<b>DUV</b>	<b>Deep Ultra-Violet</b>
<b>EBE</b>	<b>Electron-Beam Evaporation</b>
<b>EBL</b>	<b>Electron-Beam Lithography</b>
<b>EBR</b>	<b>Edge-Bead Removal</b>
<b>EDX</b>	<b>Energy-Dispersive X-ray</b>
<b>EELS</b>	<b>Electron-Energy-Loss Spectroscopy</b>
<b>EIS</b>	<b>Electrochemical Impedance Spectroscopy</b>
<b>ELISA</b>	<b>Enzyme-Linked ImmunoSorbent Assay</b>
<b>ENEPIG</b>	<b>Electroless-Nickel Electroless-Palladium Immersion-Gold</b>
<b>EUV</b>	<b>Extreme Ultra-Violet</b>
<b>FEG</b>	<b>Field-Emission Gun</b>
<b>FET</b>	<b>Field-Effect Transistor</b>
<b>FIB</b>	<b>Focused Ion-Beam</b>
<b>FNT</b>	<b>Fowler-Nordheim Tunneling</b>
<b>GPC</b>	<b>Growth-Per-Cycle</b>

---

<b>HAADF</b>	<b>H</b> igh- <b>A</b> ngle <b>A</b> nnular <b>D</b> ark- <b>F</b> ield
<b>HFA</b>	<b>H</b> uman <b>F</b> etuin <b>A</b>
<b>HRP</b>	<b>H</b> orse <b>R</b> adish- <b>P</b> eroxidase
<b>ICP</b>	<b>I</b> nductively- <b>C</b> oupled <b>P</b> lasma
<b>IPA</b>	<b>I</b> so <b>P</b> ropyl <b>A</b> lcohol
<b>IR</b>	<b>I</b> nfra- <b>R</b> ed
<b>LO</b>	<b>L</b> ongitudinal- <b>O</b> ptical
<b>LST</b>	<b>L</b> yddane- <b>S</b> achs- <b>T</b> eller
<b>MBE</b>	<b>M</b> olecular- <b>B</b> eam <b>E</b> pitaxy
<b>MG</b>	<b>M</b> ie- <b>G</b> rüneisen
<b>MIM</b>	<b>M</b> etal- <b>I</b> nsulator- <b>M</b> etal
<b>MIS</b>	<b>M</b> etal- <b>I</b> nsulator- <b>S</b> emiconductor
<b>MOCVD</b>	<b>M</b> etal- <b>O</b> rganic <b>C</b> hemical- <b>V</b> apour- <b>D</b> eposition
<b>MOSFET</b>	<b>M</b> etal- <b>O</b> xide- <b>S</b> emiconductor <b>F</b> ield- <b>E</b> ffect <b>T</b> ransistor
<b>MRI</b>	<b>M</b> agnetic- <b>R</b> esonance- <b>I</b> maging
<b>NMR</b>	<b>N</b> uclear- <b>M</b> agnetic- <b>R</b> esonance
<b>NSCLC</b>	<b>N</b> on- <b>S</b> mall- <b>C</b> ell <b>L</b> ung <b>C</b> ancer
<b>NVM</b>	<b>N</b> on- <b>V</b> olatile <b>M</b> emory
<b>NW</b>	<b>N</b> ano <b>W</b> ire
<b>OLED</b>	<b>O</b> rganic <b>L</b> ight- <b>E</b> mitting- <b>D</b> iode
<b>PAC</b>	<b>P</b> hoto <b>A</b> ctive <b>C</b> ompound
<b>PBS</b>	<b>P</b> hosphate- <b>B</b> uffered <b>S</b> aline
<b>PDF</b>	<b>P</b> robability- <b>D</b> istribution <b>F</b> unction
<b>PEALD</b>	<b>P</b> lasma- <b>E</b> nhanced <b>A</b> tomical <b>L</b> ayer <b>D</b> eposition
<b>PEG</b>	<b>P</b> oly <b>E</b> thylene <b>G</b> lycol
<b>PID</b>	<b>P</b> roportional/ <b>I</b> ntegral/ <b>D</b> erivative
<b>POC</b>	<b>P</b> oint <b>O</b> f <b>C</b> are
<b>PSA</b>	<b>P</b> rostate- <b>S</b> pecific <b>A</b> ntigen
<b>QD</b>	<b>Q</b> uantum <b>D</b> ot
<b>RF</b>	<b>R</b> adio- <b>F</b> requency
<b>SAM</b>	<b>S</b> elf- <b>A</b> ssembled <b>M</b> onolayer
<b>SE</b>	<b>S</b> econdary <b>E</b> lectron
<b>SED</b>	<b>S</b> ingle- <b>E</b> lectron <b>D</b> evice

---

<b>SEM</b>	<b>S</b> canning <b>E</b> lectron- <b>M</b> icroscopy
<b>SELDI-TOF</b>	<b>S</b> urface- <b>E</b> nhanced <b>L</b> aser <b>D</b> esorption/ <b>I</b> onization- <b>T</b> ime <b>O</b> f <b>F</b> light
<b>SET</b>	<b>S</b> ingle- <b>E</b> lectron <b>T</b> ransistor
<b>SPR</b>	<b>S</b> urface <b>P</b> lasmon <b>R</b> esonance
<b>STEM</b>	<b>S</b> canning <b>T</b> ransmission <b>E</b> lectron- <b>M</b> icroscope
<b>STM</b>	<b>S</b> canning <b>T</b> unneling- <b>M</b> icroscopy
<b>STS</b>	<b>S</b> canning <b>T</b> unneling- <b>S</b> pectroscopy
<b>TEM</b>	<b>T</b> ransmission <b>E</b> lectron- <b>M</b> icroscopy
<b>TM</b>	<b>T</b> ransverse- <b>M</b> agnetic
<b>TMA</b>	<b>T</b> ri <b>M</b> ethyl <b>A</b> luminium
<b>TMAH</b>	<b>T</b> etra <b>M</b> ethyl <b>A</b> mmonium <b>H</b> ydroxide
<b>TO</b>	<b>T</b> ransverse- <b>O</b> ptical
<b>UV</b>	<b>U</b> ltra- <b>V</b> iolet
<b>VCC</b>	<b>V</b> oltage- <b>C</b> oefficient of <b>C</b> apacitance
<b>VTI</b>	<b>V</b> ariable- <b>T</b> emperature <b>I</b> nsert
<b>WKBJ</b>	<b>W</b> entzel- <b>K</b> ramers- <b>B</b> rillouin- <b>J</b> effreys



# Physical Constants

Elementary Charge	$e$	$=$	$1.60217662 \times 10^{-19}$	C
Avogadro's Number	$N_A$	$=$	$6.02214086 \times 10^{23}$	mol <sup>-1</sup>
Permittivity of Free Space	$\epsilon_0$	$=$	$8.85418782 \times 10^{-12}$	A <sup>2</sup> .s <sup>4</sup> .kg <sup>-1</sup> .m <sup>-3</sup>
Planck's Constant	$h$	$=$	$6.62607004 \times 10^{-34}$	m <sup>2</sup> .kg.s <sup>-1</sup>
Planck's Reduced Constant	$\hbar$	$=$	$1.05457180 \times 10^{-34}$	m <sup>2</sup> .kg.s <sup>-1</sup>
Electron Rest Mass	$m_e$	$=$	$9.10938356 \times 10^{-31}$	kg
Boltzmann Constant	$k_B$	$=$	$1.38064852 \times 10^{-23}$	m <sup>2</sup> .kg.s <sup>-2</sup> .K <sup>-1</sup>



*For Science.*



# Chapter 1

---

## Introduction

---

Diseases in general are a blight on society that impacts the well-being and productivity of families and individuals, and many of these diseases are deadly, such as cancer, which is one of the most studied. New cases of cancer are diagnosed at a rate of more than 0.1% of the global population every year[1], corresponding to at least 7 million people diagnosed with cancer globally every year. Currently non-small-cell lung cancer (NSCLC) constitutes the highest mortality rate worldwide amongst other common types of cancer, where the life expectancy is less than 5 years in 85% of cases[1–3]. Cancer research has been ongoing for more than two centuries since the first surgeries were attempted to remove cancerous growths[4] and the types of cancers and the conditions under which they occur are now understood much better. Despite these efforts, the significant death toll and the burden on society imposed by these diseases is very high, and there are significant challenges facing the effective treatment and diagnosis of most cancer types.

Of course, an affordable and effective medical treatment that can cure or reverse a particular cancer would be the best solution, and would solve the immediate desperately needed health-care solution, and the large economic burden on society. Unfortunately, despite the world-wide effort, no solutions are on the horizon. However a significant improvement in outcome, i.e. in reducing mortality rates and improving the quality of life of the patient, using existing medical treatments could be achieved in many cases if the disease could be detected earlier. Currently there are many years between the occurrence of cancer and the diagnosis. Hence an effective method of detecting cancers as early as possible is paramount[5]. This is not just the case for cancer and in fact, earlier detection of deadly diseases in general would significantly reduce mortality rates and associated economic costs[6]. However, novel approaches and technological solutions are required to enable effective early diagnostics.

### 1.1 High sensitivity biosensors

As a potential solution, high sensitivity biosensors are being developed to fulfil the need for earlier diagnosis of deadly diseases. Medical practitioners use biomarkers to diagnose specific conditions in a patient. The concept of a biomarker refers to a means of identifying the state of health or of a disease through the measurement of unique indicators[7], often increased concentration of proteins in a patient's blood or other

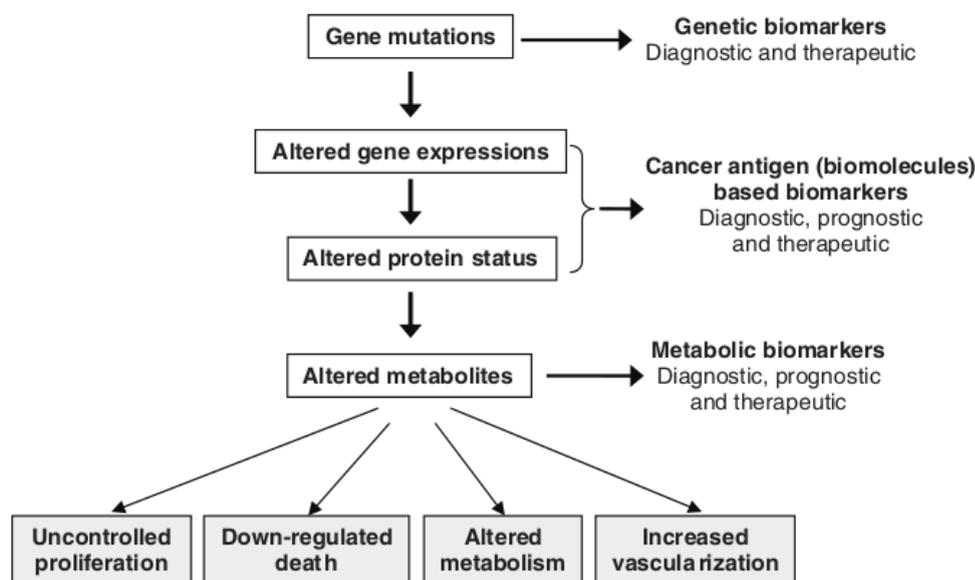


FIGURE 1.1: Simplified time-flow diagram of the process of carcinogenesis, showing opportunities for identifying biomarkers. Adapted from Bhatt *et al.*[7].

bodily fluid. In the context of a disease like cancer, biomarker molecules are released by cancerous cells and can fall under a variety of types such as DNA and proteins amongst other biomolecular entities, that can be used to diagnose, monitor and improve prognoses of the disease[8]. Figure 1.1 shows a flow diagram of the carcinogenesis process, the process by which cells become cancerous, and details the types of biomarker molecules produced at various stages of genesis. The search for biomarkers related to specific cancers constitutes an international scientific effort that guides the development of sensing technology with concrete sensitivity and specificity requirements. Using a series of biomarkers rather than just a single biomarker during an assay increases the accuracy of the diagnosis, through the minimisation of false positives. Many biomarkers that are NSCLC-specific have been discovered, and due to the complexity of lung cancer and many other diseases, biomarker ‘panels’ are now proposed to increase precision and specificity in diagnosis[7, 9], whereby a large number of biomarkers are used to minimise false positive as much as possible.

Serum-based biomarkers are most compatible with integrated lab-on-chip sensing platforms, which are discussed in Section 2.3.1, as significant advances in microfluidic technology have enabled liquid flows to be controlled precisely using very small sample volumes[10, 11]. Biomarkers of this type include DNA and proteins released by tumours into the bloodstream[12], that can subsequently be detected using ultra-sensitive detection techniques that combine high sensitivity and specificity[13]. In general biomarkers may be produced in connection with a tumour in various ways and can be classed depending on their origin[14]. For instance, risk biomarkers describe molecules associated with,

or that contribute to carcinogenesis, and response biomarkers that describe molecules produced by the body in response to the tumour. The use of biomarker panels for diagnosis in point-of-care applications can potentially be integrated into one device by simultaneously detecting multiple proteins, as has been shown for example with silicon nano-wire (SiNW) based systems[15, 16] amongst others. These new types of devices and sensors are set to revolutionise healthcare by replacing the current costly and time-consuming state-of-the-art equipment with hand-held devices that can be used similarly to common glucose sensors today. Figure 1.2 illustrates the different aspects of the point-of-care (POC) device concept, which could be used to quickly perform diagnoses without resorting to expensive and time consuming techniques that require specific expertise to operate, such as magnetic resonance imaging (MRI) and enzyme-linked immunosorbent assays (ELISAs), the latter of which are discussed in further paragraphs.

Tumours in early stages of growth constitute a very small number of cells, thus the concentrations of certain biomarkers released into the bloodstream are significantly reduced compared to advanced tumours[18]. A good example of a biomarker that has gained wide acceptance amongst medical practitioners is the prostate-specific antigen (PSA) biomarker for screening prostate cancer[19], that is also widely used as a model biomarker for testing novel biosensing platforms[13, 15, 20]. Although there has been controversy over its use alone due to correlations with other diseases and the risk of

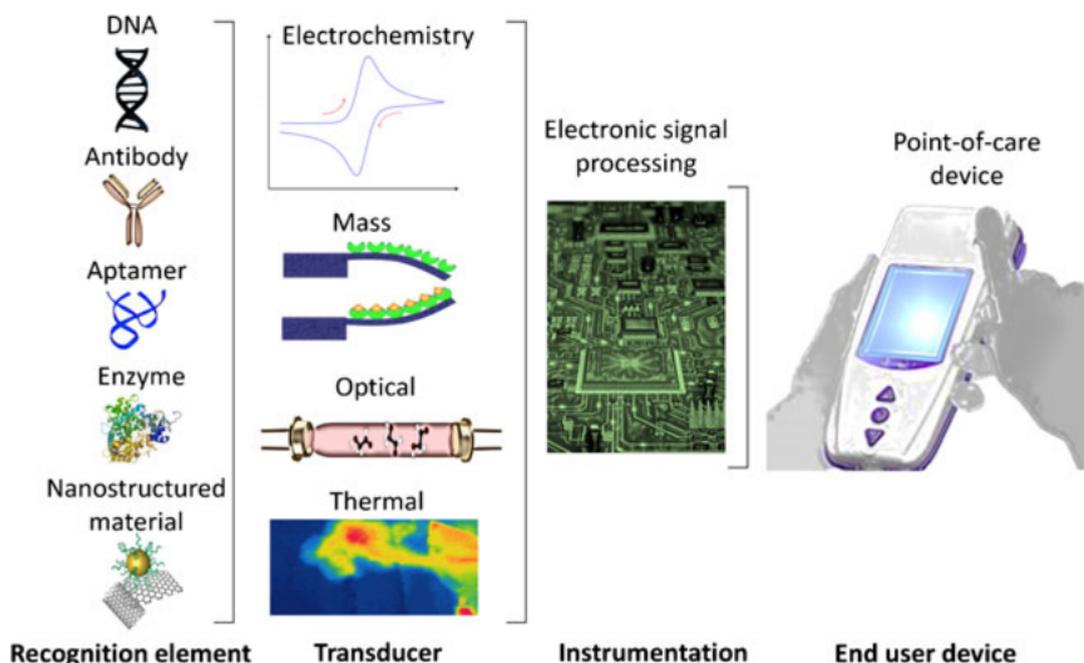


FIGURE 1.2: Flow diagram of the diagnostic detection of biomarkers. From left to right, a multitude of biomarkers is presented, alongside alternative detection strategies. Upon detection an electronic signal is produced and processed leading to a diagnostic readout on the POC device. Adapted from Lopez-Barbosa *et al.* [17].

overdiagnosis[21], which highlights the need for multiple biomarkers to be assayed at the same time, and specifically the use of biomarkers with higher specificity. The detection guidelines for PSA screening are as follows: Normal serum PSA levels are currently considered to be less than  $2.5 \text{ ng.mL}^{-1}$ , above which individuals are considered at risk, and it is recommended to monitor PSA levels every year. The accepted threshold for discerning high risk individuals is currently  $4 \text{ ng/mL}$  which constitutes the earliest possible precise detection of the signs of prostate cancer by way of this single biomarker. Advanced prostate cancers can yield PSA concentrations between 20 and  $1000 \text{ ng.mL}^{-1}$ [18, 19] which highlights the difference in orders of magnitude of the biological signal. With regard to pre-clinical detection, PSA alone cannot be used to conclusively diagnose prostate cancer as the normal baseline value is only increased slightly.

To date there is not a consensus on a standardised set of biomarkers for many of the known cancers, due to the many sub-types, therefore the medical community is preparing for significant reforms that involve the collection and archiving of biomarker measurement data on a previously unseen scale to begin personalising detection and cure strategies[8]. Under this initiative termed ‘precision medicine’, all potential biomarkers are considered and particularly those that when used alongside others drastically increase the sensitivity, specificity and therefore the accuracy of a diagnostic assay. There has already been good progress in the discovery and characterisation of novel biomarkers of smaller molecular mass[22, 23], due to the observation that both small proteins and fragments of larger proteins in serum must hold valuable diagnostic information[24]. To this end mass spectroscopy techniques are being used to determine the molecular mass of proteins associated with different cancers[25]. Progress is such that new potential serum-biomarkers with higher sensitivity and specificity have been identified by mass spectroscopy for breast, prostate, ovarian, renal, gastric and liver cancers. Further details are summarised by Muthu *et al.*[26], and Yang *et al.*[27] discuss promising candidates for NSCLC diagnosis. In addition to efforts to look for low molecular weight proteins, known large molecular weight proteins are still being identified as potential lung cancer biomarkers: A recently proposed serum-based protein biomarker panel has shown high specificity towards late and early stage lung cancer[16]. The physical properties and chemical structure of proteins in general are discussed in more detail in Section 2.3.3.

The precision medicine era that we are entering, as described by Vargas *et al.* [8], is predicted to drastically improve research outcomes if the database management systems can be realised, and with more contributions globally this development is expected to accelerate. With this infrastructure and the developments in sensor technology, it is thought that a long term solution to the early detection problem is through the use of highly sensitive sensors that can detect single, or at least small numbers of molecules or DNA fragments at the very earliest stages of cancer development[28] provided these

molecules and DNA fragments are specific enough alongside others to identify a specific condition in a patient. As nanofabrication methods improve, technology that can be integrated into smaller and less costly devices will enable point of care applications sooner. The development of ultra-sensitive biomedical sensors is underpinned by the anticipation of all of these results[23]. The application of these sensors is not relevant only to the detection and subsequent diagnosis of cancer, but to any potentially useful medical information relating to therapeutic reponse, prognoses and risk assessment[14].

Currently the most sensitive assays that are used extensively in the medical practise are performed using ELISAs, that have been shown in some cases to achieve sensitivities of the order of  $10 \text{ pg.mL}^{-1}$ , for example in detecting human fetuin A (HFA)[30], a biomarker specific to a few types of cancer, and the HIV-1 p24 antibody[31], used in the diagnosis of the HIV-1 infection. The measurements they produce are used as benchmarks in many studies that use known biomarker proteins[16, 32]. Commercially available kits are typically limited in sensitivity to concentrations on the order of  $1 \text{ ng.mL}^{-1}$ [31]. To illustrate the scale, consider the atomic mass, measured in standard atomic units (Da), of the HFA biomolecules, which is on the order of 40 kDa, corresponding to roughly  $6.6 \times 10^{-23} \text{ kg}$ . Then a concentration of  $1 \text{ ng.mL}^{-1}$  represents roughly one HFA molecule per  $0.066 \mu\text{m}^3$  volume, which converted to molar concentration gives  $25 \text{ nM}$  ( $\text{nmol.L}^{-1}$ ). A typical sandwich-type ELISA process is schematically described in Figure 1.3, where a substrate is first coated with a capture antibody specific to the target antigen. Treatments to prevent non-specific binding and interactions are then usually applied, after which the diluted samples containing the target can be added to the solution. The targets then bind to the surface-bound antibodies over an incubation period. After washing the sample to remove unbound antigen, a detection antibody

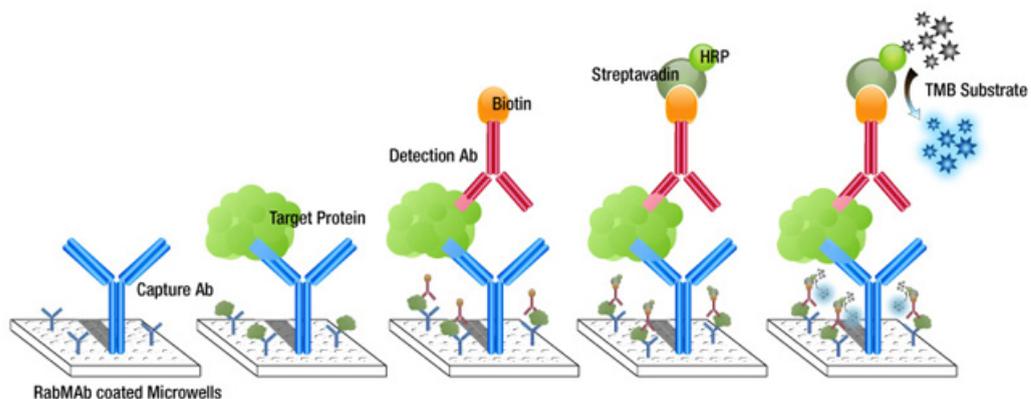


FIGURE 1.3: An example sandwich ELISA assay, commonly used for measurement of low concentrations of target antigens. Image obtained from [29]

functionalised with a reliable binding protein is specifically bound to the target antigen. Typically the avidin-biotin interaction is used in applications where strong binding affinity is required between a target protein and another target such as another protein, functionalised surface or nanoparticle. To amplify the enzymatic signal a biotinylated signalling protein such as horseradish peroxidase (HRP) for colorimetric signal creation can be used, that binds strongly to the avidin functionalised detection antibody. There are many ELISA protocols that vary in sensitivity depending on the target proteins and the quantification method[33], and many commercially available kits can be used to assay many different proteins of relevance to cancer. Therefore ELISA is a widely applicable and versatile method for detecting specific biomolecules in a laboratory setting.

Despite the wide success of ELISAs in determining the concentration of important cancer biomarkers to clinically relevant levels[20], the sensitivity limitations and heavy dependence on protocols limits the applicability of the technique to the early detection of biomarker panels. It is also difficult to develop new ELISA techniques for emerging biomarkers as these new molecules must each be first characterised to develop an applicable process. In this regard, mass spectrometric techniques, particularly surface-enhanced laser desorption/ionisation time-of-flight SELDI-TOF mass spectroscopy have been used to find biomarkers in serum of importance to early detection biomarkers and has helped to confirm that low-molecular weight proteins must be helpful biomarkers[26]. However

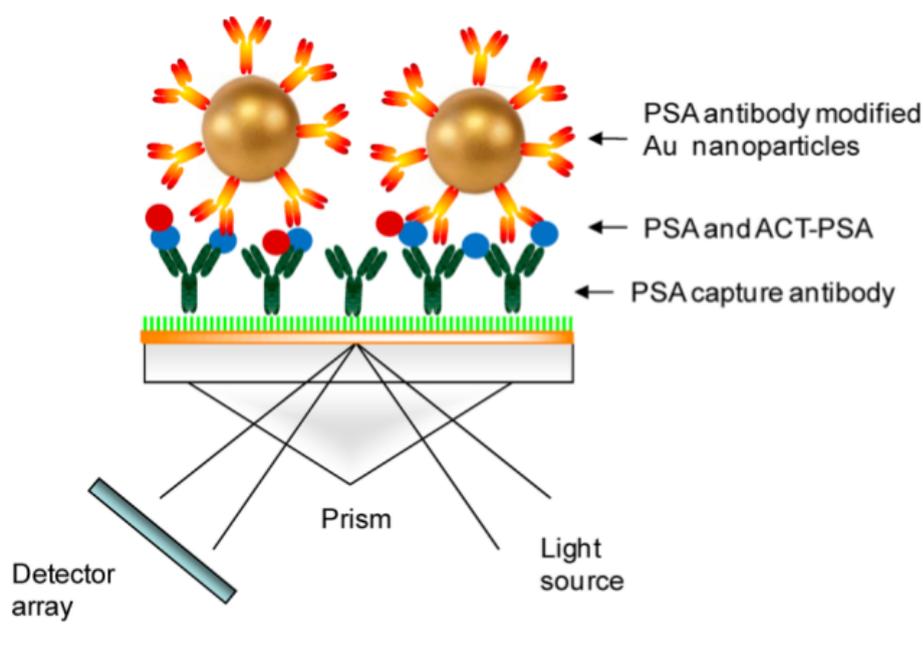


FIGURE 1.4: Example of an SPR biosensor, where PSA detection is amplified by the inclusion of anti-PSA functionalised Au nanoparticles that bind to surface bound PSA, and consequently change the dielectric properties of the local medium to a greater extent, resulting in an amplified signal. Adapted from Uludag *et al.*[34].

the technique suffers due to its highly qualitative nature and due to this doesn't provide much progress in new assay processes and technology for studying specific proteins. ELISA is challenging to scale down as it is a laboratory-based technique that requires expensive scientific apparatus to operate, extensive reagents and highly qualified technicians and experts to perform. Due to the wide variety of measurement methods, the dependence of results on particular protocols and the time required to develop these protocols, it is very costly to provide a universal way of screening for certain types of cancer based on this method. Furthermore the protocols are time-consuming, resource-intensive and vary from assay to assay which makes the technique highly unfavourable for integration into point of care devices[35].

Another widely used biosensing platform is based on surface plasmon resonance (SPR), illustrated in Figure 1.4, where electron oscillations at the interface between a metal and a dielectric material are affected by changes in the permittivity of the dielectric medium near the interface where biomolecules are adsorbed[36]. In this technique, a laser is shone onto the back of a metal at a high angle from the surface normal, so that the light is reflected and captured by a detector that measures the light intensity. In very thin films (on the order of 50 nm) of metals that show SPR, such as Au, surface plasmons can be created on the top side of the metal, which contains the biological dielectric medium, when the light is incident at a precise angle. At this angle, the optical absorption of the substrate becomes greater, which can be detected, and the angle can then be recorded. The adsorption of new species onto the top surface affects the dielectric properties of the medium and the conditions for SPR to occur are then slightly changed, which results in a shift in the angle at which the optical absorption peak occurs. This technique has proven invaluable in molecular biology research, as it allows the monitoring of reactions with the surface, and can be used to make quantitative estimates of concentrations and chemical properties. However this technique requires complex equipment including a precisely aligned optical system, and thus presents challenges with regards to integration in POC applications. SPR biosensors are classed as optical biosensors, since the detection principle is based on the dielectric properties of the system. As the dielectric medium is influenced by aggregation of mass onto the surface, the angular resolution of the absorption detector limits the smallest detectable change in mass. The minimum detectable concentration is reported to be on the order of  $10 \text{ pg.mL}^{-1}$ [37]. Protocols based on SPR biosensing platforms that use labelling to amplify the signal, such as represented in Figure 1.4, offer more sensitive and faster assays than the equivalent ELISA tests[34], and thus SPR provides a higher throughput laboratory technique than ELISA in some specific cases, although it is more costly due the equipment required and limited in sensitivity for label-free detection. Despite these challenges, it has been shown that smart phone technology can be used to create a portable SPR biosensor platform[38], from which they determined a concentration

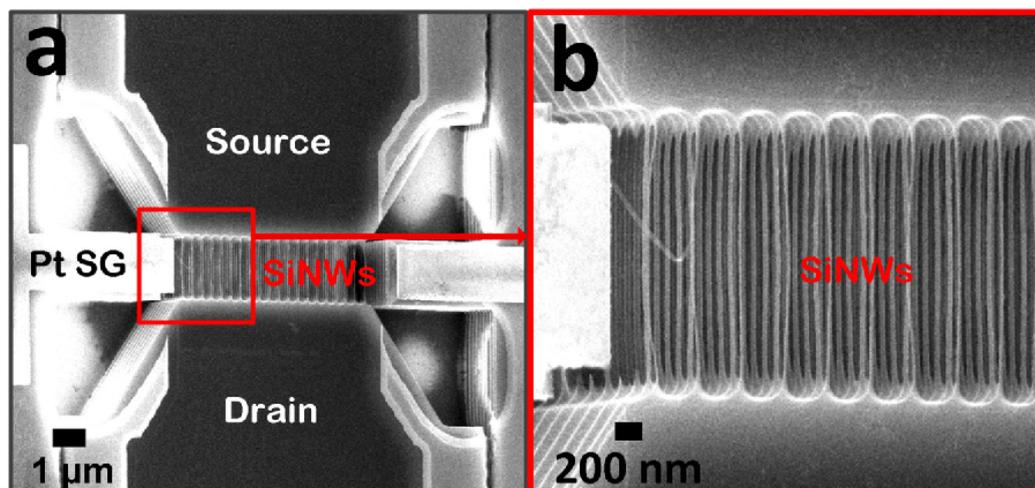


FIGURE 1.5: Scanning electron microscope (SEM) images of a stacked SiNW structure fabricated using CMOS processing, used to maximize the surface area for charge sensing. A voltage is applied across the source and the drain, and the resulting current is a function of charges near the wires. Adapted from Buitrago *et al.*[39], where (a) shows the whole transistor structure at a slight angle from above, and (b) shows a magnified image of the SiNW array.

resolution on the order of 50 nM, which corresponds to approximately  $7.5 \mu\text{g}\cdot\text{mL}^{-1}$  for the 150 kDa protein used in their work.

The most significant challenges facing the development of electronic transducers to detect biomarker panels are the sensitivity and throughput limitations, as molecules from the very few cancerous cells that exist at an early stage need to be captured to increase survival rates[18]. Part of this challenge is due to the need for large biomarker panels to be assayed to yield higher accuracy, requiring devices that can multiplex sensing. SiNW systems have led the development of such sensors where continuous developments in CMOS processing enable the mass fabrication of devices that can be densely integrated alongside microfluidic systems and other peripherals. Sensors that use the field-effect as a transducer have the advantage that the detection method can be label-free, where the binding event is directly converted to a measurable signal. Many conventional biomolecule detection strategies, such as ELISA, utilize labelling techniques to identify the analyte. These technologies are limited in that the detection is a function of the labelling technique and the labelling molecule has properties of its own that must be considered. The development of label-free techniques implies a significant reduction in the cost associated with preparing the sample and the equipment used to conduct experiments. Figure 1.5 shows scanning electron microscope (SEM) images of a SiNW array system fabricated using a CMOS process, showing that the sensor surface area can be greatly increased using arrays. The SiNWs shown in these images were coated with  $\text{Al}_2\text{O}_3$  deposited by atomic layer deposition (ALD) to passivate the wires from the sample. This form of technology currently offers the most economical solution for mass

fabrication, since an entire biosensing system can in principle be integrated into a single chip without the need for detection optics[40], and studies that focus on the fabrication of arrayed SiNW biosensors with CMOS compatible processes[39] and microfluidics on a large scale[41] are already underway.

SiNW devices have been applied as pH sensors[43] and as biomolecule sensors[44–46]. The presence of charges external to the semiconducting NW have the same effect as a gate electrode held at a given potential that causes accumulation or depletion of the charge carriers in the wire. As the wire becomes smaller in diameter, a greater surface-to-volume ratio becomes available, meaning that a smaller number of carriers will be affected by the same external charge, increasing the effect on the conductance of the wire. The interface between the semiconductor and the ligand used for biomolecule capture and the microfluidic environment around the NWs are crucial factors in the practical application of the technology, as induced charges along the NW depend on the electrostatic environment and the geometry of the system. For instance, if the NW's surrounding permittivity is high, the field lines from a charge placed a certain distance away will smear out along the NW, causing a small depletion or accumulation region all along the NW. Whereas in a smaller permittivity dielectric, the field lines will be more concentrated in the region it's closest to, causing a much larger and localised depletion or accumulation region that is in fact much more effective at changing the conductance of the wire[47]. Figure 1.6 shows the change in current along a NW as a function of the potential applied to a nearby gate electrode, and the effect of local external charge that shifts the response[42]. The curve shown corresponds to a SiNW system fabricated

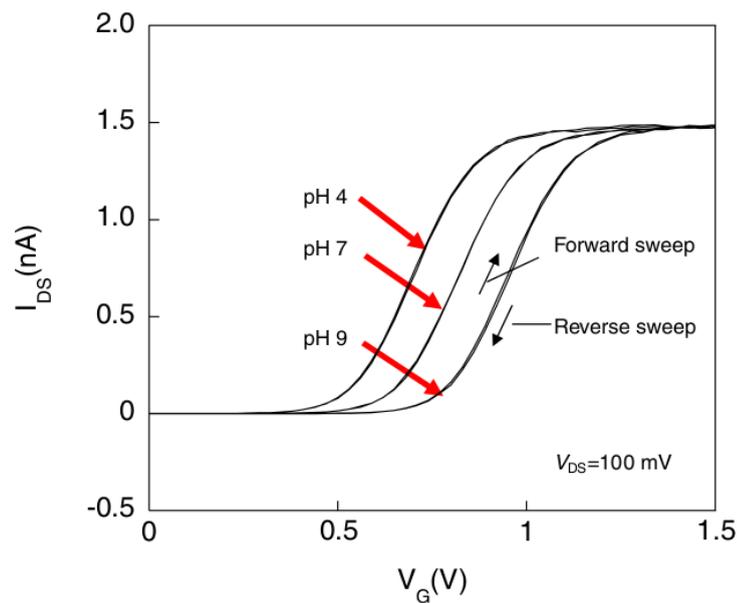


FIGURE 1.6: Response of a SiNW to an external gate voltage, under immersion in solutions of varying pH. Adapted from Kudo *et al.*[42].

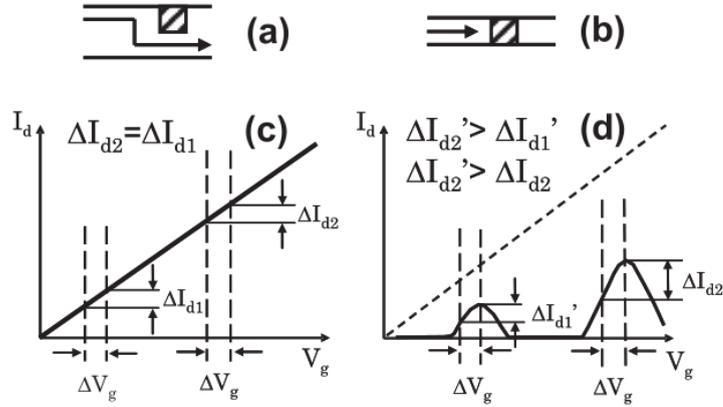


FIGURE 1.7: (a) and (c) show the change in transconductance of an FET system when the external charge is changed, and (c) and (d) show the response of an SET system, where the change in transconductance could be engineered to be greater than in an FET system. Adapted from Nakajima *et al.*[49].

using electron beam lithography (EBL) and wet etching techniques, with a final NW width of 80 nm and an oxide passivation layer 19 nm thick. Femto-molar (fM) range detection can be achieved in SiNW systems, as has been shown in an electrostatically formed SiNW using a number of side and back gates to form and move the NW closer to the sensing surface[48]. While SiNWs already provide very high sensitivity, there is interest to find even more sensitive systems that could be used to detect single molecules, and perhaps even to study the dynamics of the molecular structure.

Single electron transistors (SETs) are exceptionally sensitive charge devices, where by analogy to a field-effect transistor (FET), the conductance of the channel can be strongly modulated by the presence of external charges. Theoretical treatments show that within a small bandwidth, sub-electronic charge sensitivity, i.e. a small fraction of  $e$ , of the conductance can be achieved[50]. Figure 1.7 (a)–(d) shows a comparison of the transconductance of an FET and an SET device, which is the conductance of the channel measured as a function of changes in the gate electrode potential. Figure 1.7 (d) shows the Coulomb oscillations of the transconductance[51], that could provide a greater change in conductance due to external charges. The discovery of electronic charging effects in granular metallic thin-films[52] and the realisation that electron transfer in nanoscale metallic grains involves a significant change in the free electrostatic energy of the system compared to the thermal energy[53, 54] initiated research on single electron charging phenomena. It was predicted theoretically by Averin *et al.* that when a nanoscale junction is biased by a current source at low temperature, oscillations of the junction voltage should arise due to the correlated discrete transfer of single electrons[55], which occur due to the significant energetic barrier presented to tunneling electrons, termed the Coulomb blockade. The Coulomb blockade effect will be discussed in detail in Chapter 2. Improvements in the resolution of EBL technology allowed the

fabrication of very small metal-insulator-metal (MIM) junctions and circuits in which Coulomb blockade could be observed at practical temperatures[56], and since, the most versatile and useful structure was found to be the SET. An SET is composed of a very small metallic island tunnel-coupled to two very small metallic electrodes through which a current may flow, and is capacitively coupled to a gate electrode that is used to modulate the free energy of the island. Figure 1.8 (b) shows an SEM image of a metallic SET, fabricated using shadow angle evaporation[57], illustrated in panel (a), whereby a suspended mask is used with an angled Al evaporation source to produce islands and electrodes, and where an  $\text{Al}_2\text{O}_3$  thin film can be grown by controlled, *in-situ* exposure to oxygen atmosphere[58].

Much SET research was first performed using aluminium devices fabricated using metal evaporation techniques[58–62] such as shown in Figure 1.8 (a), and still is today[63, 64] where metallic SETs find applications as radiation[65, 66] and charge detectors[67, 68]. Since the discovery of Coulomb blockade and single-electron effects in nanoscale circuits, many potentially ground breaking applications have been proposed, such as the use of SET circuits for CMOS logic circuits[69] and quantum bit read-out stages in quantum computers[68, 70]. However as the Coulomb blockade is a fundamental effect linked to the classical interaction of the electron charges, SET devices can be produced in a large number of material systems besides the traditional Al– $\text{Al}_2\text{O}_3$ –Al structures. It has been shown for instance, that charging effects can be produced in Cr– $\text{Cr}_2\text{O}_3$ –Cr junctions[71], which show unique transport properties[72, 73]. Semiconducting SET systems have been produced, and a wide variety of quantum transport properties, electronic confinement effects and magnetic phenomena have been observed and explained theoretically as a result[74, 75]. Due to the fundamental nature of single electron charging effects, interest in the relation of these effects to atomic structure have led to the creation of single-atom, dopant or molecule SET devices[76–82], which show great potential for room temperature operation. SETs can also be fabricated or formed using alternatives to conventional lithography techniques used in CMOS and EBL processes. For example, AuNW SETs

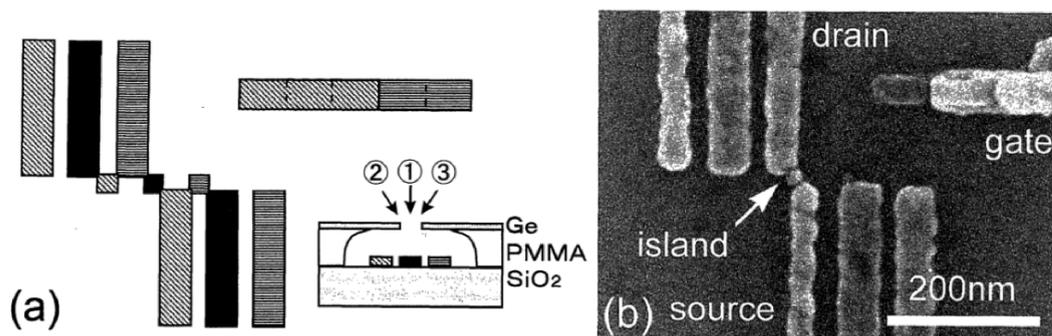


FIGURE 1.8: (a) Diagram showing the shadow angle evaporation technique for fabricating metallic SETs. (b) SEM image of an SET fabricated using this method. Adapted from Nakamura *et al.*[58].

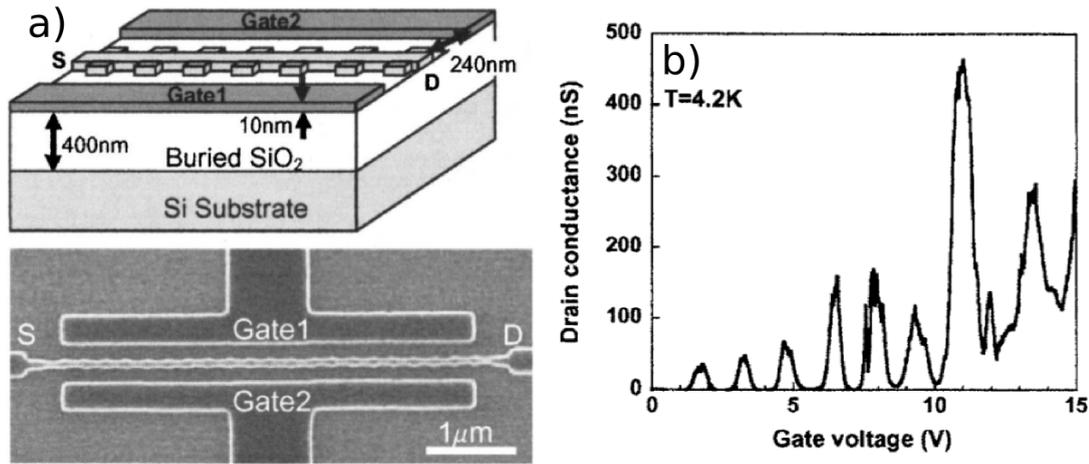


FIGURE 1.9: (a) Schematic and associated SEM image of a multi-island SET system comprised of 22 islands. (b) Transconductance of the 22-island device shown panel in (a), measured at 4.2 K, which shows the well known Coulomb oscillations[51] of the transconductance when the gate potential is varied. Adapted from Ohkura *et al.*[89].

grown chemically from the ground-up[83] and Au nano-particles (AuNP) deposited onto nanoscaled electrode arrays[84] have been demonstrated, where new innovative avenues are being explored using biomolecular technology[76, 85]. Furthermore, tunnel barriers can be formed electrostatically using very small electrodes to deplete a two-dimensional electron gas (2DEG) in semiconductor heterostructures[75, 86, 87], which have also been demonstrated in sub-electronic charge sensor applications[88].

Due to the significant confinement effects on electrons in nanoscale silicon SET systems, Coulomb blockade has frequently been observed at room temperature[90–92], which has revived efforts to focus on the CMOS compatible silicon material system. For example, room temperature Coulomb blockade effects can be achieved in down-scaled tri-gate metal-oxide-semiconductor FET (MOSFET) devices[93]. These device cannot be easily mass-produced however, as only devices with specific roughness features were found to work. Similarly, SETs can be formed from SiNWs, using a fabrication process called pattern-dependent oxidation[94], where the NW is patterned with constrictions, which after oxidation lead to tunnel barrier formation in these regions. This technique was effectively exploited to form multi-island SETs[89], a schematic and associated SEM image of which is shown in Figure 1.9 (a), which can be used to raise the operating temperature of single electron transistors, and will be discussed in detail in Chapter 2. Figure 1.9 (b) shows the transconductance associated with the device shown in panel (a), which similarly to single island systems, shows Coulomb oscillations. This technology has led to the first experiments using room temperature silicon multi-island SETs as biosensors, where it has been shown that pH changes[95] and biomolecule binding events[49, 96, 97] can be detected. Nakajima *et al.* report a minimum sensitivity of  $1 \text{ ng.mL}^{-1}$  for the detection of the protein streptavidin[49], which shows the potential for

SETs to rival other high sensitivity biosensors. Although these initial results are highly promising and do prove the SET biosensor concept is achievable, the yield of devices manufactured this way is very low. Furthermore it is not clear whether the devices presented in these works function as intended, i.e. that the use of multiple islands is the reason for room temperature operation. This uncertainty is associated with the oxidation fabrication step that forms the tunnel barriers in these devices, since the oxidation process is very difficult to control, and is used in this design not only as a crucial step in forming the tunnel barriers, but is crucial in forming the islands also.

Semiconducting SET systems provide the important benefit that confinement effects contribute significantly to the Coulomb blockade[74] compared to metallic systems of the same scale, however the details of the confinement and the coupling with the electrodes are strongly dependent on the geometry and the material properties of the system[68], and thus it is currently too difficult to produce arrays of islands with identical confinement properties. Ohkura *et al.*'s results[89], which presents a clearly defined structure (Figure 1.9 (a)), can just as well be explained as due to just a single or a few islands, rather than due to the combined properties of the whole chain of 22 islands, which was assumed homogeneous in their work. The images of the structures presented by Nakajima *et al.* show a much less homogeneous structure[49], where islands are not clearly defined, therefore it is likely that the working devices they identified are due to cases where exceptionally small islands are present with suitably formed tunnel barriers connected them together. Confinement effects in the context of SET devices are currently too difficult to reproduce on an industrial scale, therefore alternative solutions to the Si/SiO<sub>2</sub> material system are sought that could be used to mass manufacture SETs.

## 1.2 Aims and Objectives

The aim of this project is to investigate the potential of metallic SETs applied as biosensors, where the unprecedented charge sensitivity could potentially be exploited in a biosensor platform, and could eventually provide a route to the detection of single molecules. To address the issues faced with semiconductor SET systems already reported in the literature, the use of alternative material systems that are compatible with mass fabrication processes are required. It is proposed in this work that if electronic confinement effects can be minimized, the geometrical tolerance of the fabrication processes will not be as important. In this case, the SET system depends more on the electrostatics of the system, which is also less sensitive to geometrical variations. The electrostatics of the SET system, discussed at length in Chapter 2, are determined in part by the properties of the insulating thin film that separates the island(s) from the electrodes, which is required to function as both a high quality and precise capacitor and an equally precise electron tunneling resistance. Atomic layer deposition (ALD) is a widely

used self-regulating chemical vapour deposition process, described in detail in Chapter 3, which has been used extensively to deposit conformal, and highly uniform thickness films of  $\text{Al}_2\text{O}_3$  [98] in applications such as the 45 nm CMOS FET gate technology and MIM capacitors[99] and tunnel diodes[100]. Conventional silicon SETs fabricated using EBL techniques have typically used variations of an oxidation step and specific patterns to form tunnel junctions, called pattern dependent oxidation[94]. Although this method has been shown to work, the device yield is poor due to the variability of the tunneling properties, and the SET parameters often change during voltage and thermal cycling, due to the drift of dopants and impurities[68, 101], which is worsened at higher temperatures. Although the resolution of CMOS processes is adequate to form nanoscale electrodes and islands, the same cannot be said for the formation of tunnel barriers with a high yield using these oxidation techniques. Thus, the primary objective of this project is to take advantage of the high uniformity, reliability and mass fabricability of thin  $\text{Al}_2\text{O}_3$  films formed by ALD to mass fabricate room temperature operable SETs, since  $\text{Al}_2\text{O}_3$  has been used extensively throughout the history of SET research, and is known to work in these applications.

There are few reports that discuss the fabrication of SETs using ALD in the literature, as efforts in SET research have focused primarily on silicon devices in low-temperature applications. The few reports that exist on the subject have focused on the use of noble metals such as Au and Pt as the island and electrode materials[102–104]. Noble metals do not readily oxidise under an oxygen atmosphere, and as long as ALD growth occurs linearly and covers the surface, the thickness of the oxide can be precisely controlled. Alternative ALD processes that employ different gas precursors can be used to prevent oxidation of the substrate surface, however these systems are much less available, and much less experimented with. Metallic SETs are simpler to design than semiconducting SETs, as the electrostatic properties can be more easily determined and reproduced, as will be discussed in detail in Chapter 2. The secondary objective of this project is to target the use of noble metals for the SET source and drain electrodes and the islands, which can be used to take advantage of the benefits of widely available and studied ALD processes. The use of such metals may also circumvent the issue of providing electrostatically stable electrodes and islands at room temperature, in that the effect of voltage and thermal cycling is not so important as with semiconductor devices. Additionally, it is not clear why metallic SETs fabricated with ALD- $\text{Al}_2\text{O}_3$  tunnel barriers show a poor yield, and only a small number of devices are usually investigated in the literature. To address this issue, the properties of an ALD deposited film need to be understood over a larger scale, and the scaling relationship of the film properties with the geometry must be established, which should aid in the practical design of mass fabricable SETs based on ALD- $\text{Al}_2\text{O}_3$ . Therefore, a third objective is to determine the geometrical scaling

behaviour of the properties of the ALD- $\text{Al}_2\text{O}_3$  film relevant to the electrostatics and operation on an SET, notably the capacitance and electron tunneling properties.

### 1.3 Thesis Structure

In this work, a large number of MIM devices with varying geometrical parameters were characterised electronically in detail to assess the properties of  $\text{Al}_2\text{O}_3$  ultra-thin films, that is thicknesses  $\leq 5$  nm, over large surface areas relevant to industrial scale manufacturing. Using this methodology, a better knowledge of the material properties and their effect on the SET system can be obtained through a top-down approach, similarly to the way that down-scaling of devices has driven CMOS research to better understand the geometrical and material parameters of transistors to continue down-scaling. The knowledge gained through this approach should then serve to prescribe the SET system behaviour more accurately. The literature that treats the use of ALD deposited  $\text{Al}_2\text{O}_3$  to create SET devices is few and far between, and the few reports that do exist present a bottom up approach, and thus the SET parameters are determined when a working device is found. In Chapter 2 a simple theory of metallic SET systems is used to demonstrate the practical challenges associated with achieving room temperature operation, and the challenges associated with the integration of the device into a biosensor platform is addressed, including a final section discussing the possible applicability of the mass fabrication of metallic SETs using an ALD-based CMOS process. The experimental techniques used to fabricate, observe and characterise the MIM devices are detailed in Chapter 3, including a detailed section (Section 3.2.3) describing the deposition of  $\text{Al}_2\text{O}_3$  by ALD. In Chapter 4 the design and fabrication of the MIM devices is presented, where the geometrical parameters of the resulting junctions are described in detail. As a large number of devices were characterised, the general features of the electronic properties obtained by alternating-current (AC) and direct-current (DC) measurements are described in detail and compared extensively with the MIM literature in Chapter 5. There the scaling of the capacitance and junction resistance is described, and conclusions are presented with regards to the geometrical consistency of the devices and the implications for manufacturing tunnel barriers by ALD for SET devices. In Chapter 6, selected MIM devices are characterised in greater detail from a physics point of view, where it is shown that subtle effects of the geometry and the material properties occur as the thickness of the  $\text{Al}_2\text{O}_3$  layer is decreased. There are very few studies that treat specifically the thickness dependence of the permittivity of  $\text{Al}_2\text{O}_3$  in the thickness range used in this work, and those that have were performed long before modern techniques such as ALD were available[105], and therefore the results are difficult to interpret. Using simple models of the junction system, it is shown that the important effects observed in the MIM junctions can be accounted for. In particular, it was found

that small sharp features present on the bottom electrode cause an asymmetry in current as a function of applied voltage, described in Section 6.2. This work was published in Physical Review B[106]. Finally, concluding remarks are made in Chapter 7, where the outlook and further work is also discussed.

## Chapter 2

---

# SET biosensors, design and limitations

---

## 2.1 Introduction

The purpose of this chapter is to describe the theoretical and practical aspects of applying SETs as biosensors. This chapter begins with an outline of the orthodox theory of Coulomb blockade, followed by a description of metallic SETs with single and multiple islands, which provides sufficient background for the assessment of both the charge sensing capabilities and the engineering of such systems. The engineering challenges and practical limitations surrounding the integration of metallic SETs into biosensing technology are detailed in the following section, focusing in particular on the operating temperature and charge sensitivity of the device. With regard to these factors, the final section discusses the mass fabrication of metallic SETs and the use of ALD to form reliable tunnel barriers is discussed briefly in terms of state of the art CMOS processing trends.

## 2.2 Theory of metallic SEDs

In Chapter 1 the ground work and the important experimental and theoretical achievements in single-electronics were discussed. In this section, the electrostatic properties of small capacitively coupled conductors are described along with basic quantum mechanical considerations of isolated islands and weakly coupled junctions. These basic considerations are required for the description of the orthodox theory. In the following subsection, single electron circuits that could be used as biosensor components are described using classical electrostatics and the orthodox theory. Finally, the charge detection mechanism and the sources of noise in an SET are formally described to establish the maximum charge sensitivity of such a device.

### 2.2.1 Orthodox theory of Coulomb blockade

In order to illustrate the effect of Coulomb blockade, first consider a small isolated metallic spherical island with radius  $r_0$ , as shown in Figure 2.1 (a). In a classical description, the sphere initially has a total charge

$$Q_0 = \{Z_p - Z_e\}e = 0 \tag{2.1}$$

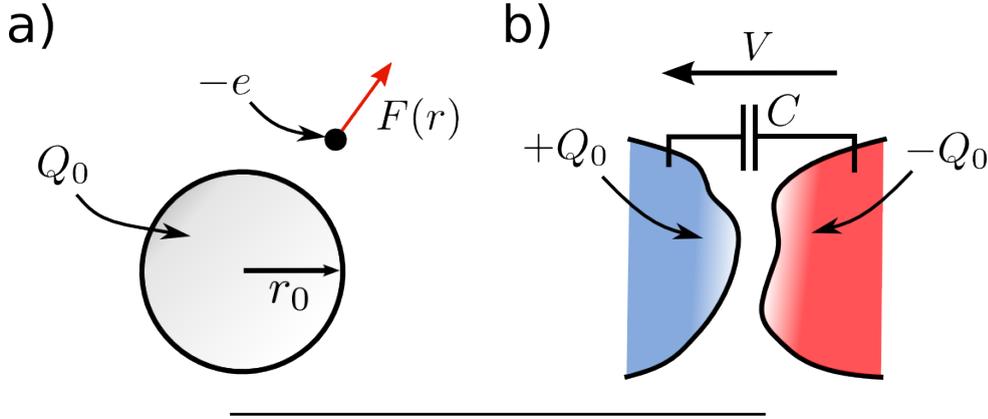


FIGURE 2.1: (a) Schematic of an isolated conductive sphere showing the force on an approaching electron when  $Q_0 < 0$ . (b) Schematic of capacitively coupled conductors showing the charges induced by the applied voltage  $V$ .

where  $e$  is the elementary charge, and  $Z_p$  and  $Z_e$  are the number of protons and electrons that form the island, respectively. In the compensated state, i.e.  $Z_p = Z_e$ , this isolated island will present no static potential outside its surface. If an electron now approaches the island at a small velocity and  $Z_p$  and  $Z_e$  are very large, a small attractive force will develop as the electron approaches the surface of the island due to the dipole effect, known as the image force[107]. If the electron is absorbed into the island, its total charge will now be

$$Q_0 = \{Z_p - (Z_e + 1)\}e = -e \quad (2.2)$$

where the island has an excess negative charge. The island now presents a potential outside its surface

$$\varphi(r) = \frac{-e}{4\pi\epsilon r} \quad \text{for } r > r_0 \quad (2.3)$$

which will present an increasing repulsive force

$$F(r) = -e \frac{d\varphi}{dr} \quad (2.4)$$

to any approaching electrons at position  $r > r_0$ . The energy required to overcome this force is termed the charging energy, which gives rise to Coulomb blockade phenomena. Defining the voltage between the surface at  $r_0$  and a point  $r$  as  $V(r) = \varphi(r) - \varphi(r_0)$ , for an electron infinitely far away, the energy required to overcome this force is

$$E_C = eV(\infty) = \frac{e^2}{C_{\text{self}}} \quad (2.5)$$

where  $C_{\text{self}}$  is the self-capacitance of the island determined by

$$C_{\text{self}} = \frac{e}{V(\infty)} = 4\pi\epsilon r_0. \quad (2.6)$$

Equation 2.5 shows that as the self capacitance of the island becomes smaller, the energy required to overcome the repulsive force increases. In an island that has large values of  $Z_p$  and  $Z_e$ , adding a single electron to such an island makes a very small difference to the total electrostatic energy. Indeed as  $C_{\text{self}}$  increases in Equation 2.5 the energy required to add electrons decreases to a negligible value compared to the thermal energy  $k_B T$ . Therefore in a very small system approaching the nanoscale,  $Z_p$  and  $Z_e$  are small, such that the addition of a single electron now makes a substantial difference to the electrostatic energy.

We now consider the two conductor system represented in Figure 2.1 (b) where a very narrow constriction is formed resulting in a small junction area. If a voltage  $V$  is applied between both conductors, a charge  $Q_0 = CV$  will be induced on each conductor surface, where  $C$  is the mutual capacitance of these two conductors. The energy stored in this capacitor is

$$U(Q_0) = \frac{1}{2}CV^2 = \frac{Q_0^2}{2C} \quad (2.7)$$

where if the case when  $Q_0 = e$  is considered, the charging energy for transfer of a single-electron in a two-conductor system can be found as

$$E_C = \frac{e^2}{2C} \quad (2.8)$$

which represents half of that for a completely isolated conductor. Therefore if electrons are permitted to flow between the two conductors, there will be no transfer of charge if the electron cannot overcome the charging energy of the system (Equation 2.8). If the capacitance  $C$  is extremely small, the ensuing energy barrier to a single electron becomes very large. It is important to note that for two separated conductors with a controlled potential, the charge is induced by the applied potential difference  $V$  and therefore  $Q_0$  can take on values arbitrarily smaller than  $e$ . Similarly, when a current is forced around the circuit, a charge will build up on each conductor surface until the charging energy threshold is reached. When  $V > \frac{e}{2C}$ , electrons from the higher potential conductor have enough energy to overcome the Coulomb blockade in the lower potential conductor. The single and dual conductor systems show that both self (or stray) and mutual capacitances play a role in determining the classical charging energy of a system of conductors.

When an isolated conductor becomes very small, quantisation effects of the electron energy levels will begin to play a role and a quantum physical description is necessary. In order to illustrate the scale at which quantisation effects occur, the isolated conductor in Figure 2.1 is approximated as an infinite square potential well in one dimension, represented in Figure 2.2 (a) where  $t = 2r_0$ . Taking the bottom of the well ( $U(x) = 0$ ) at the Fermi level, the energy eigenvalues and wavefunctions can be found by solving

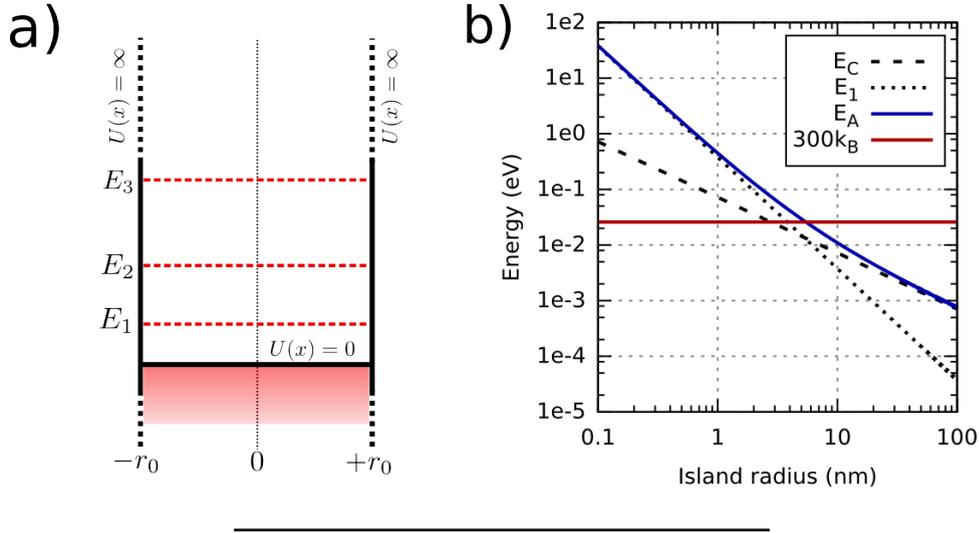


FIGURE 2.2: (a) Energy band diagram of an infinite square well showing the resulting splitting in single electron energy levels. (b) Plot of the addition energy  $E_A$  using Equation 2.13, where the first energy level  $E_n$  for  $n = 1$ , and a permittivity of  $\epsilon = 10\epsilon_0$  for the island self-capacitance are used. The thermal energy at room temperature is indicated by the dotted red line.

the one dimensional time-independent Schrödinger equation

$$\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \psi_n(x) + [E_n - U(x)] \psi_n(x) = 0 \quad (2.9)$$

where  $m$  is the free electron mass,  $E_n$  is energy eigenvalue  $n$ ,  $\psi_n(x)$  is eigenfunction  $n$  and  $U(x)$  is the potential energy of the whole system. The solution in the well is then

$$\psi_n(x) = A \cos(k_n x) + B \sin(k_n x) \quad (2.10)$$

where  $A$  and  $B$  are constants and  $E_n = \frac{\hbar^2 k_n^2}{2m}$ . Imposing the boundary conditions  $\psi_n(-r_0) = 0$  and  $\psi_n(r_0) = 0$ , only the values  $k_n = n\pi/t$  for  $n = 1, 2, 3, \dots$  provide a valid solution  $\psi_n(x)$ , so that the energy eigenvalues take the form

$$E_n = \frac{\hbar^2 \pi^2 n^2}{2mt^2} \quad (2.11)$$

where the extent of quantisation of the electron energy levels is controlled by the thickness  $t$ . Arguably this is a crude approximation of the electron behaviour in a spherical isolated conductor, however the energy eigenvalues take a similar form as long as angular momentum is ignored. Finding the solution to the wavefunctions in a spherically symmetric system requires the use of the radial Schrödinger equation, where the wavefunctions depend on  $r$ ,  $\theta$  and  $\phi$ . In this spherically symmetric potential, the wavefunctions will no longer take the form in Equation 2.10, rather they take the form of Bessel functions. Using the same procedure as before for finding the  $k$  values for which the

boundary conditions are satisfied, the energy eigenvalues take the form

$$E_n = \frac{\hbar^2 \pi^2 n^2}{2mr_0^2} \quad (2.12)$$

where the effect of the quantum numbers related to the polar ( $\theta$ ) and azimuthal ( $\phi$ ) dependence of the wavefunction is ignored. Figure 2.2 (b) illustrates the dependence of both the classical and quantum contribution (for  $n = 1$ ) to the total addition energy of an electron to the island

$$E_A = E_C + \Delta E_n \quad (2.13)$$

as a function of the radius  $r_0$  of the sphere, where Equation 2.12 was used for  $\Delta E_n$ . As can be seen from the plot in Figure 2.2 (b), the quantum confinement can be ignored for an island radius above  $t = 10$  nm.

When two conductors are brought in close proximity to each other, a small current will be allowed to flow through the finite potential barrier separating them due to the quantum tunneling effect. This effect arises due to the fact that when an electron is incident on a finite potential of higher energy, the wavefunction of the electron  $\psi(x)$  exists in the classically ‘forbidden’ potential energy region. To illustrate this effect, consider two bulk conductors separated by a thin rectangular barrier of height  $U(x) = U_0$  for  $0 \leq x \leq t$ . The bulk conductors are modelled as semi-infinite regions with potential energy  $U(x) = 0$  at  $x < 0$  and  $x > t$  so that electron kinetic energies follow a parabolic dispersion

$$E_n = E(k) = \frac{\hbar^2 k^2}{2m} \quad (2.14)$$

where  $E(k)$  represents a continuous variable, so that there are no confinement effects. Using Equation 2.9, when  $U(x) > E(k)$  the electron wavefunctions take the form:

$$\psi(x) = C \exp(\kappa x) + D \exp(-\kappa x) \quad (2.15)$$

where  $C$  and  $D$  are constants and

$$\kappa = \frac{1}{\hbar} \sqrt{2m[U_0 - E]} \quad (2.16)$$

is the decay rate of the wavefunction in the barrier region. If electrons are incident from the left, the solution in the barrier exponentially decays, and there will be a finite probability amplitude that ‘leaks’ into the region on the right. By applying continuity to the wavefunctions and their derivatives at each interface, the probability amplitudes of the incident and transmitted waves can be found. The ratio of the transmitted to incident probability amplitudes squared gives the tunneling probability for a single

incident electron as a function of the barrier parameters and the incident electron energy

$$P(E) \approx \frac{16E(U_0 - E)}{U_0^2} \exp(-2\kappa t) \quad (2.17)$$

where it is assumed that  $\kappa$  (Equation 2.16) is large, i.e. for electron energies far from the edge of the potential  $U_0$ . This function shows that for small values of  $t$ , the single electron tunneling probability becomes exponentially larger.

The classical Coulomb energy  $E_C$  and the tunneling effect provide the necessary components to describe metallic single-electron circuits. The orthodox theory of Coulomb blockade provides a simple way of reproducing the basic properties of single electron circuits relating to the classical Coulomb blockade, and is used as a starting point to describe all SET systems. In this theory, a number of important assumptions are made:

1. Quantum confinement effects are negligible. As was shown previously, these effects become negligibly small for metallic features greater than 10 nm critical dimensions. This condition can be expressed as

$$\Delta E_n \ll k_B T \ll E_C \quad (2.18)$$

which implies that a series of states are available for single electron tunneling onto and from the island whilst the charging energy ground state is within the bias window (Equation 2.36)[51]. This means an electron may occupy any one of those states lying above the charging energy level, however upon electron transfer, all of those states are shifted upwards in the band diagram to the next charging energy level, so that single-electron tunneling still occurs, just with a wider range of states available at and above the charging energy level.

2. The electron tunneling time is much smaller than other timescales of processes in the junction, including the time between consecutive tunneling events[56]. For consecutive tunneling events to occur over a long time scale, the junction time constant  $\tau = RC$  needs to be larger than the timescale of quantum fluctuations implied by the Heisenberg uncertainty with charging energy  $\Delta E \Delta \tau \geq \frac{\hbar}{2}$ , which yields the condition

$$G \leq G_Q = \frac{e^2}{\hbar} \approx 39 \mu\text{S} \quad (2.19)$$

where  $G = 1/R$  is the conductance of the junction, and  $R_Q \approx 26 \text{ k}\Omega$ . As long as  $G$  remains much smaller than the conductance quantum  $G_Q$ , random quantum fluctuations of electrons across the barriers are suppressed. To date, the traversal time of a tunneling electron is still a controversial issue. A comprehensive early discussion on the subject is provided by Landauer and Martin[108], where it is argued that the traversal time of electrons and their formal description under different conditions is still an open question

and that interpretation of experiments is sensitive to conditions for tunneling. Recent experimental and theoretical developments in the physics of high-power laser-matter interactions suggest that the tunneling time is much faster than would be expected if the electron travelled the distance of the energy barrier at a given momentum[109, 110]. These experiments show that no observable tunneling delay can be observed in attosecond ( $10^{-18}$  s) resolved ionisation experiments. The tunneling process described in these experiments occurs over atomic scales, however an increase in dimensions of the potential by an order of magnitude, such as in a MIM junction, is not expected to significantly decrease the tunneling time, therefore it is considered to be zero throughout this work. The image force effects during the tunneling of an electron are therefore ignored, since no image dipole has time to form in response to an instantaneous electron transition.

**3.** Higher order and coherent tunneling processes are not accounted for, only the sequential tunneling of single electrons is considered. Higher order tunneling processes refer to situations where multiple tunneling events occur at the same time. In multi-junction systems with at least two junctions, cotunneling of electrons may occur, also referred to as macroscopic tunneling of the quantum charge ( $q$ -MQT). The process is described as a single tunneling event crossing multiple junctions through the Coulomb blockade, where ‘virtual’ energy levels that exist within the Coulomb energy gap cause a non-vanishing conductance in the Coulomb blockade regime[111]. There are two contributions to the cotunneling conductance, an elastic contribution  $G_{\text{el}}$  where no energy is lost, and an inelastic process  $G_{\text{in}}$  where energy is lost or gained. In a two junction system with capacitances  $C_1$  and  $C_2$  (see Figure 2.5 (a)), and in the limit of small voltages, Averin *et al.*[111] estimate the elastic contribution as

$$G_{\text{el}} = \frac{\hbar G_1 G_2}{4\pi e^2 \rho} \left( \frac{C_\Sigma}{C_1} + \frac{C_\Sigma}{C_2} \right) \frac{1}{eV} \quad (2.20)$$

where  $\rho$  ( $\text{eV}^{-1}$ ) is the density of states in the volume of the island, and

$$G_{\text{in}} = \frac{\hbar G_1 G_2}{12\pi e^2} \left( \frac{C_\Sigma}{C_1} + \frac{C_\Sigma}{C_2} \right)^2 \left[ 1 + 4\pi^2 \left( \frac{k_B T}{eV} \right)^2 \right] \quad (2.21)$$

where the electron temperature plays a role. These events occur in parallel such that  $G_{\text{cot}} = G_{\text{in}} + G_{\text{el}}$ . Both contributions become large at small voltages  $V$  and when the conductances of each junction  $G_1$  and  $G_2$  are large. The inelastic contribution also increases with thermal energy.

The main result of the theory is that the flow of single electrons through a tunnel junction is mediated by the difference in the free energy of the system before and after the tunneling event[56]. The steady state current  $I_0(V)$  of the junction without any Coulomb charging effects is used to determine the tunneling rate of electrons. Then the

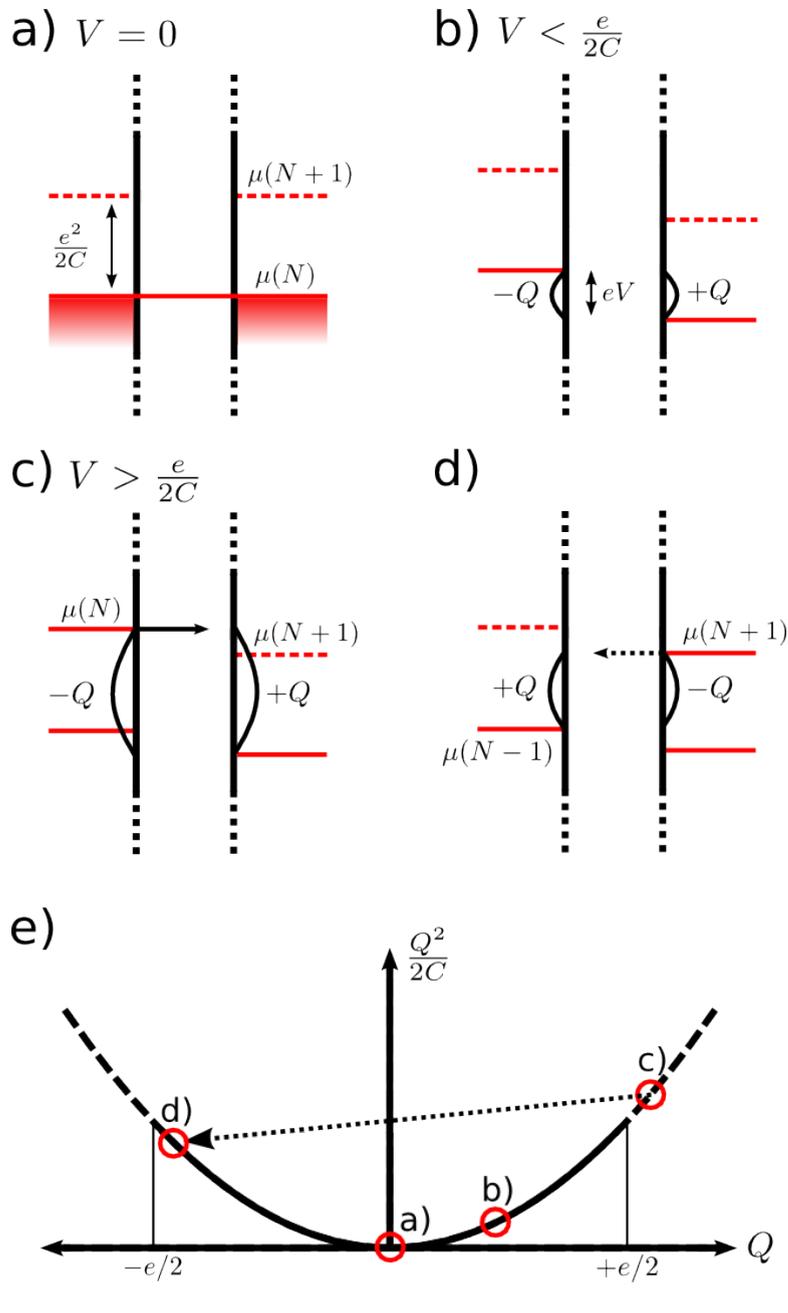


FIGURE 2.3: Band diagrams illustrating how Coulomb blockade affects the current flow in a single junction. (a) Junction at equilibrium showing the electrochemical potentials  $\mu(N)$  with  $N$  electrons. The charging energy  $E_C$  of the junction is shown. (b) Applying a small voltage  $V < \frac{e}{2C}$  induces a charge  $CV = Q$ , splitting the electrochemical potentials of each electrode by  $eV$ . (c) When  $V \geq \frac{e}{2C}$ , electrons may now tunnel into the empty energy level in the right electrode, the induced charge is now greater than  $e/2$  before the tunneling event. (d) Immediately after the tunneling event, the electrochemical potentials of each electrode change by  $E_C$ , giving a total change in the free energy  $2E_C = \frac{e^2}{C}$ . In this state, the single electron may not tunnel back into the left electrode due to the Coulomb blockade, thus the circuit discharges instead and returns to the state in (c). (e) Plot of the electrostatic energy against the induced charge  $Q$ , showing the points that correspond to diagrams (a)-(d).

measured I–V characteristic of a specific junction is modified by including the change in free energy  $\Delta U^\pm$  as an activation energy:

$$\Gamma^\pm(\Delta U^\pm/e) = \frac{I_0(\Delta U^\pm/e)}{e} \frac{1}{1 - \exp(-\frac{\Delta U^\pm}{k_B T})} \quad (2.22)$$

where  $k_B T$  is the thermal energy,  $\Gamma^\pm$  is the tunneling rate into a state that increases (+) or decreases (–) the total charge  $Q$  of the junction, in units of probability per unit time. In this junction the electrostatic energy is governed by Equation 2.8 so that the change in free electrostatic energy before and after the tunneling event can be expressed as

$$\Delta U^\pm = U(Q) - U(Q \pm e) = \mp \frac{e}{C} \left( Q \pm \frac{e}{2} \right) \quad (2.23)$$

where  $U(Q)$  is the energy before an electron tunnels, and  $U(Q \pm e)$  is the energy after the tunneling event. The charge transfer process is illustrated schematically using band-diagrams in Figure 2.3 where (a) shows a junction with no applied bias and the charging energy is shown, (b) shows a junction with a bias  $V < \frac{e}{2C}$  where the induced charge  $Q$  has been annotated. Diagrams (c) and (d) show the energy level occupancy immediately before and after the tunneling event, respectively, illustrating the drastic change in voltage. As can be seen from diagram (d), the electron now in the right contact may not tunnel back into the left contact due to the Coulomb blockade. Thus, the single electron can only flow around the circuit, and the right electrode discharges, and the left electrode recharges, returning to the state shown in (c). Diagram (e) shows the charging energy in each of the states (a)-(d).

Calculation of the I–V characteristics involves considering all the possible electronic transitions that increase and decrease the charge in the junction. In the orthodox theory the dynamics of the charge are summarised by the ‘master’ Equation[54, 112]

$$\frac{\partial p}{\partial t}(Q_0) = \sum_{\pm} [\Gamma^\mp(Q_0 \pm e)p(Q_0 \pm e) - \Gamma^\mp(Q_0)p(Q_0)] \quad (2.24)$$

where  $p(Q, t)$  is a probability distribution associated with the charge state  $Q$  of the junction. Equation 2.24 is a stochastic partial differential equation derived from the Fokker-Planck equation with a drift property represented by the transition rates, and no diffusion property. The magnitude of each transition rate  $\Gamma^\mp(Q)$  is determined by the change in electrostatic energy, where  $\Gamma^\mp$  tend to negligibly small values if  $\Delta U^\pm < 0$ . Inspecting Equation 2.23 shows that when  $Q_0 > \frac{e}{2}$ ,  $\Delta U^-(Q_0)$  becomes positive, which leads to  $\Gamma^-(Q_0)$  becoming significant, which therefore allows a finite tunnel current  $I(t)$  to flow. Figure 2.4 (a) shows the states considered in Equation 2.24 in the case where the Coulomb blockade can be overcome by a single electron. Figure 2.4 (b) shows the dependence of the transition rate against the change in free energy at certain values of

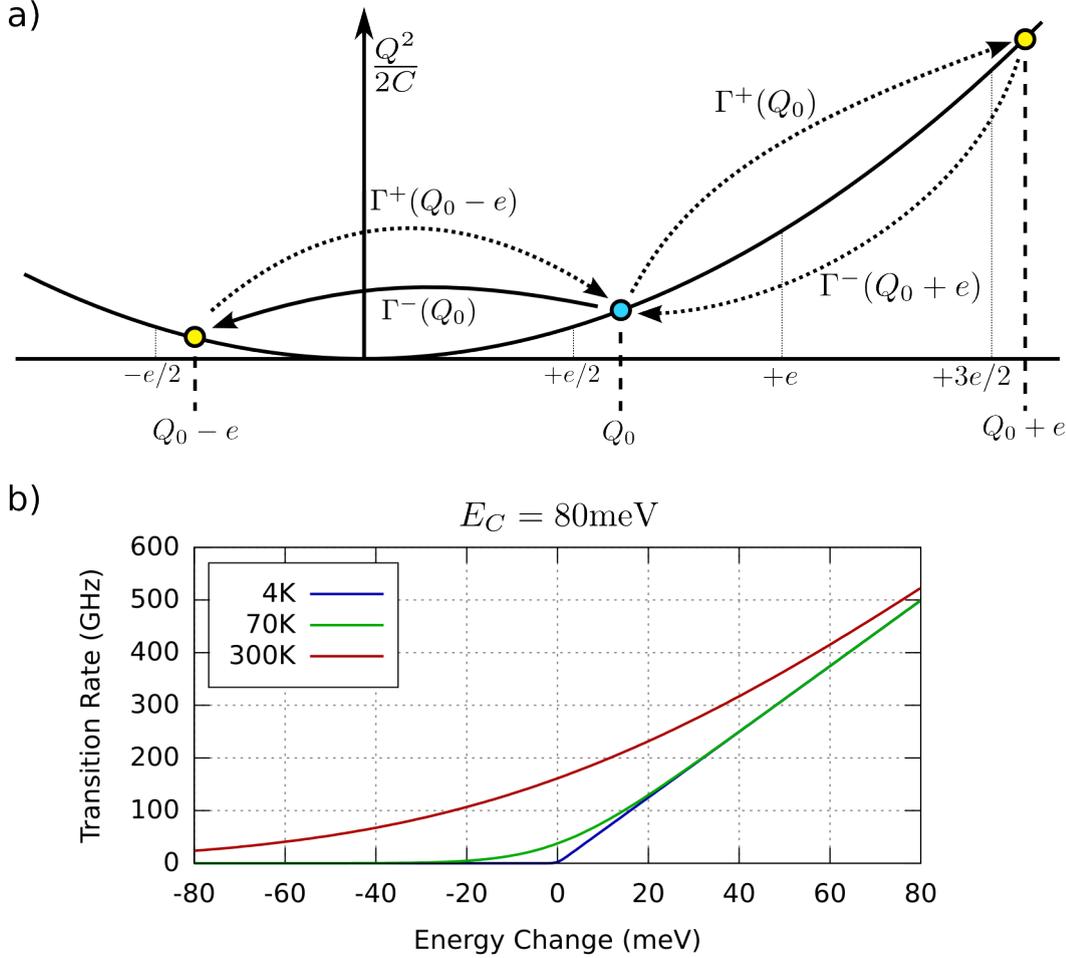


FIGURE 2.4: (a) Diagram of the possible tunneling events near the induced charge  $Q_0$ , which in this case is greater than the threshold  $e/2$ . The unfavourable transitions are denoted by the arrows with broken lines. (b) Plot of the transition rate  $\Gamma(\Delta U)$  for a junction with a charging energy  $E_C = 80\text{meV}$ .

temperature. As can be seen from this plot, at very low temperature the transition rate vanishes for negative changes in the electrostatic free energy. In this regime only the transition shown by the solid arrow in Figure 2.4 (a) is possible. If only this transition is considered, when  $Q_0 > \frac{e}{2}$ ,  $p(Q_0)$  describes a rapidly decaying state with a time constant  $\tau = (\Gamma^-(Q_0))^{-1}$ . After the tunneling event, the junction is in the blockade state until the charge  $Q_0 = CV$  is restored by the voltage source (instantaneously for an ideal voltage source). The current in this state is then only proportional to the decay frequency  $1/\tau$  of the state  $p(Q_0)$ , which gives

$$I_{\text{avg}} \propto e\Gamma^-(Q_0) \quad (2.25)$$

where the steady state current  $I_{\text{avg}}$  represents the junction I-V curve without the Blockade effect offset by a voltage  $V_{\pm} = \pm \frac{e}{2C}$  for forward (+) and reverse (-) bias configurations. As the applied voltage exceeds higher multiples of  $V_{\pm}$ , more transitions become

favourable at the same time, increasing the number of states that must be described using Equation 2.24, and the current approaches the I–V curve of the junction without blockade effects[113].

Thermal fluctuations cause electrons to exist at higher energies according to the Fermi-Dirac statistics of electrons in a bulk conductor

$$f(E_i) = \frac{1}{1 + e^{(E_i - E_F)/k_B T}} \quad (2.26)$$

where  $f(E_i)$  is the occupation probability of state  $E_i$  relative to the Fermi energy  $E_F$  of the conductor. The effect of the thermal energy is therefore to broaden the probability density  $p(Q)$  of each charge state  $Q$  such that a finite current arises due to electrons overcoming the Coulomb blockade. One of the biggest obstacles to building practical single-electron devices is the difficulty with raising the charging energy  $E_C$  such that Coulomb blockade effects occur at room temperature. As can be seen from figure 2.4 (b), the tunnel rates are highly sensitive to the thermal energy  $k_B T$ . Even for values of  $E_C$  on the order of four times the thermal energy at room temperature, the tunnel rates into various states take non-negligible values giving rise to a significant suppression of the Coulomb blockade. In order to suppress the effects of thermal fluctuations, the charging energy must be large enough such that  $f(\mu(N + 1))$  (Equation 2.26) becomes negligible which at least requires that

$$E_C > 10k_B T \quad \text{and} \quad E_C > 260 \text{ meV} \quad \text{at} \quad T = 300 \text{ K}. \quad (2.27)$$

This condition shows that charging energies approaching the eV scale are required to obtain clear Coulomb blockade effects at room temperature. For a single junction, this would require a capacitance on the order of  $10^{-19}$  F. For a square parallel plate capacitor of thickness 1.5 nm and permittivity of vacuum, this value gives a capacitor plate edge size of approximately 7.22 nm, which is beyond the current lithography resolution limits. One approach to achieving the required energy range of  $E_C$  for room temperature operation is through the formation of tunnel junction arrays in a series configuration. However, this approach comes with a tradeoff in the achievable sensitivity to external charges due to the stray and gate capacitances, presenting a challenge for sensor applications and room temperature operation. This is discussed in the next section.

### 2.2.2 Single and multi-island SET devices

The Coulomb blockade effect on electron tunneling was described previously in a single junction, where the basic quantities that determine the properties of the junction were identified as the capacitance of the junction  $C$ , which determines the magnitude of the change in free energy, and the tunneling I–V characteristic of the junction, which

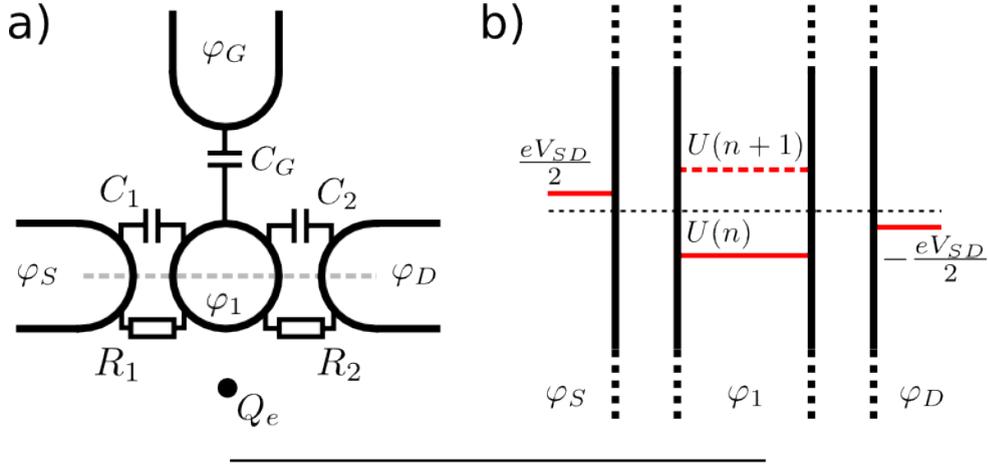


FIGURE 2.5: (a) Diagram of an SET system with equivalent circuit elements. (b) Energy band diagram of the system, taken along the dotted line in (a).

determines the transition rates. Together these properties can be used to describe the dynamics of the charge of a given tunnel junction under the orthodox theory. In a system of multiple junctions forming isolated islands, a system of equations like Equation 2.26 need to be solved for each junction. The flow of charge in such a single-electron circuit is therefore determined by the changes in the free energy  $\Delta U(Q_1, Q_2, \dots, Q_N)$  of the whole system of  $N$  tunnel coupled conductors where  $Q_i$  represent the charge state of the  $i$ -th junction.

The first single-electron circuit that displays the properties of a charge sensor is the SET shown schematically in Figure 2.5 (a), which consists of a middle island, the source and drain electrodes that form tunnel junctions to the island, and the gate electrode that is only capacitively coupled to the island. Using the principle of charge conservation, the individual charges on each conductor and an external charge  $Q_e$  can be related by

$$Q_1 + Q_S + Q_D + Q_G + Q_e = 0 \quad (2.28)$$

so that no uncompensated charges exist in the system as a whole. Each charge is related to the potentials of the conductors and the capacitances between them through the relationship  $Q = CV$  as before, which yields

$$Q_1 = (\varphi_1 - \varphi_S)C_1 + (\varphi_1 - \varphi_D)C_2 + (\varphi_1 - \varphi_G)C_G - Q_e \quad (2.29)$$

for the charge on the middle island, where  $C_{1,2}$  are the tunnel barrier capacitances and  $C_G$  is the gate capacitance. Since the gate is only capacitively coupled to the middle island in this description, no transfer of charge can occur and only induced charges may be formed. As the gate potential  $\varphi_G$  may take any value, the charge induced by the gate is a continuous value, the magnitude of which is controlled by the capacitance  $C_G$ . The external charge  $Q_e$  may in principle take any value, and the net effect is to induce

an image charge  $-Q_e$  on  $Q_1$ . In this regard, the external and gate induced charges only differ in that the latter is controlled by a varying potential, and the former is described by an arbitrary static potential and capacitance.  $Q_e$  may be arbitrarily small, taking values that are a fraction of  $e$ . This sub-electron sensitivity of the total charge on the isolated island is of considerable interest to making ultra-sensitive charge detectors.

The electrostatic free energy in this system is calculated with respect to the total charge of the island, using the expression

$$U(n) = \int_{-ne}^0 \varphi_1 dQ_1 = \frac{(-ne)^2}{2C_\Sigma} - ne \frac{\varphi_S C_1 + \varphi_D C_2}{C_\Sigma} - ne \frac{\varphi_G C_G - Q_e}{C_\Sigma} \quad (2.30)$$

where  $C_\Sigma = C_1 + C_2 + C_G$  is the total capacitance to the island and the substitution  $Q_1 = -ne$  is used, where  $n$  is the number of uncompensated electrons on the island. The single electron ( $n = 1$ ) charging energy of this system is then found when all the external potentials are zero:

$$E_C = \frac{e^2}{2C_\Sigma} \quad (2.31)$$

which shows that all the capacitances of the system, including any stray capacitance, determine the charging energy. The energy band diagram in Figure 2.5 (b) shows the island charging energy levels and the source and drain electrodes. From Equation 2.30, the basic properties of the SET can be determined by considering the change in free energy

$$\Delta U(n) = U(n-1) - U(n) \quad (2.32)$$

for adding a single electron to the island, which governs the tunneling rate into the island. First it is clear that the effect of the gate potential is to increase or reduce the free energy in a continuous way, which allows for the source-drain current  $I_{SD}$  to be modulated continuously, as required by a transistor component. Under a very small bias such that the second term on the right-hand-side of Equation 2.30 is negligible and when  $Q_e = 0$ , the free energy is minimised when

$$\frac{e^2}{2C_\Sigma} - e\varphi_G \frac{C_G}{C_\Sigma} = 0 \quad \Leftrightarrow \quad \varphi_G = \frac{e}{2C_G}. \quad (2.33)$$

Since the gate potential can be arbitrarily increased, more electrons can be added to the island and the cost to add the  $n$ -th electron  $\Delta U(n)$  is minimised at gate voltage values of

$$\varphi_G(n) = \frac{e}{2C_G}(2n-1). \quad (2.34)$$

The free energy is thus minimised periodically as a function of the gate voltage, with a period  $\frac{e}{C_G}$ . The effect of the external charge  $Q_e$  is then to offset the gate voltage at which the energy is minimised by an amount  $Q_e/C_G$ .

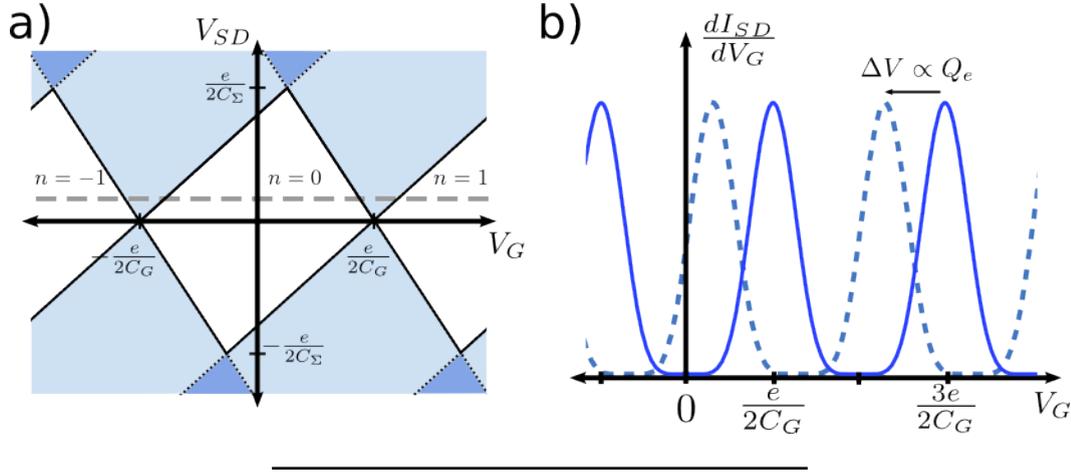


FIGURE 2.6: (a) Simplified Coulomb diamond plot of the SET oscillations. The white regions represent the Coulomb blockade regions with a very small transconductance, where a minimal current flows. The coloured regions indicate where a larger current is allowed to flow, with a darker colour indicating regions with a greater transconductance. As  $V_{SD}$  is increased other diamonds appear and multiple electrons contribute to the tunnel current lifting the Coulomb blockade regions.  $n$  is the number of uncompensated electrons trapped on the island. (b) Coulomb oscillations of the transconductance against the gate voltage  $V_G$  for a small constant value of  $V_{SD}$ . The effect of the external charge is illustrated as the shifted dashed line curve, where in this case the charge is positive.

At near-zero temperatures and including a bias voltage  $V_{SD} = \varphi_S - \varphi_D$  such that  $\varphi_S = V_{SD}/2$  and  $\varphi_D = -V_{SD}/2$  and using Equation 2.32, the change in free energy can be written as

$$\Delta U(n) = \frac{e^2}{C_\Sigma} \left( n - \frac{1}{2} \right) - e\alpha_G V_G - \frac{e}{2} V_{SD} (\alpha_2 - \alpha_1) + \frac{eQ_e}{C_\Sigma} \quad (2.35)$$

where  $\alpha_i = C_i/C_\Sigma$  for  $i \in \{1, 2, G\}$  represent coupling constants, and where  $V_G = \varphi_G$ . This equation determines the values of  $V_{SD}$  and  $V_G$  for which tunneling becomes energetically favourable. The factor  $(\alpha_2 - \alpha_1)$  represents the asymmetry of the capacitive coupling of each electrode to the dot, where it is equal to zero if  $C_1 = C_2$ . According to Figure 2.5 (b), tunneling becomes favourable when

$$-e\frac{V_{SD}}{2} < \Delta U(n) < e\frac{V_{SD}}{2} \quad (2.36)$$

which is referred to as the bias window. Using this equation with Equation 2.35 leads to a set of boundary conditions that form a series of lines in the  $V_{SD}$ - $V_G$  plane. These lines form the so called ‘Coulomb diamonds’ shown in Figure 2.6 (a), where the white regions are Coulomb blockade regions with  $n$  uncompensated electrons localised on the island. The periodic fluctuations of the current against the gate voltage are termed the Coulomb oscillations, controlled by Equation 2.33 for low bias values. A slight asymmetry is shown in the diamonds to show the effect of a slight difference between  $C_1$

and  $C_2$ . These oscillations are represented schematically in Figure 2.6 (b), where the bias condition is represented by the dashed dark line in Figure 2.6 (a). Under small values of  $V_{SD}$  and at very low temperatures in an experiment ( $eV_{SD}, k_B T \ll E_C$ ), measurements of the Coulomb oscillations serve as a spectroscopic probe of the electron energy levels of the island. The shift of these oscillations caused by the external charge  $Q_e$  represents a change in  $I_{SD}$  which can be measured experimentally.

The change in current caused by the externally induced offset charge is the simplest example of an SET acting as a charge sensor, where the detected charge is directly coupled to the central island. The sensitivity of the SET as a charge sensor in this configuration is determined by the both the coupling strength of the external charge to the island, determined by the surrounding environment; and the linewidth of the Coulomb oscillations, which are intricately tied to the coupling between the states in each electrode and the available states in the island. A detailed account of the conductance in the SET circuit under the orthodox theory is provided by Beenakker[51], from which the qualitative dependence of the conductance on the electrostatic energy state of the island is described. As long as the conditions of the orthodox theory are fulfilled, the dynamics of the electrons are determined by a probability distribution  $p$  as in Equation 2.24, only extended to include the transitions across the second barrier. The solution of  $p$  describes a chain of rapidly decaying states in and out of the island. Under a small voltage bias  $eV_{SD} \ll E_C$ , the conductance through the island is due to single electrons tunneling in and out of the island as independent, sequential events,

$$\frac{1}{\Gamma_{\text{eff}}} = \frac{1}{\Gamma_{\text{in}}} + \frac{1}{\Gamma_{\text{out}}} \quad (2.37)$$

where the in and out superscripts denote tunneling into and out of the island respectively, and the direction is determined by the sign of the applied voltage. As each tunneling event occurs independently and at random, the current can be expressed as

$$I_{\text{avg}} = \beta e \Gamma_{\text{eff}} = \beta e \frac{\Gamma_1^{\text{in}} \Gamma_2^{\text{out}}}{\Gamma_1^{\text{in}} + \Gamma_2^{\text{out}}} \quad (2.38)$$

where  $\beta$  is a constant that depends on the density of electron states in the dot. The slope of the oscillations are only determined by the tunneling characteristics (measured I-V curves) of each junction under the approximations considered. With the conditions imposed by the bias window, Equation 2.36, and the periodic minimisation of the energy by the gate electrode, Equation 2.38 oscillates as tunneling rates are decreased and increased.

The single island transistor circuit described previously presents an important drawback towards the mass fabrication of practical devices for widespread commercial use. Using conventional metal evaporation techniques and nanolithography techniques such

as electron beam lithography, and high resolution photolithography, it is currently not possible to make junctions and islands small enough such that the charging energy of the circuit (Equation 2.31) far exceeds the room temperature thermal energy, fulfilling Equation 2.27. As discussed briefly in Chapter 1, there are now many new avenues for achieving room temperature devices, and new fabrication techniques that show potential for mass-fabricating devices. However, a method of increasing the operating temperature of the SET circuit that is compatible with metal evaporation and ALD thin-film deposition, is to use a chain of islands to take advantage of the reduction of the effective capacitance along the chain, which leads to an increase in the charging energy.

The multi-island SET circuit is illustrated in Figure 2.7, showing an SET with  $N - 1$  islands each with potentials  $\varphi_i(n_i)$  where  $n_i$  is the number of uncompensated electrons on island  $i$ , and a single gate with a potential  $\varphi_G$ . As with the single island case, the conservation of charges requires that

$$Q_S + Q_D + \sum_{i=1}^{N-1} (Q_i + Q_{Gi}) = 0 \quad (2.39)$$

where each  $Q_i = -n_i e$  represents the charge state of island  $i$ , and  $Q_{Gi}$  represents the combined arbitrary gate and external charges of island  $i$ . All the charges of the system are interdependent through the potentials and capacitive coupling of all the conductors. The system can be solved by considering the first and last islands separately as boundary conditions[114], and the islands in between as a recursive relation

$$Q'_1 = Q_1 + \varphi_S C_1 + \varphi_G C_{G1} = \varphi_1 C_{\Sigma 1} - \varphi_2 C_2 \quad (2.40)$$

$$Q'_i = Q_i + \varphi_G C_{Gi} = \varphi_i C_{\Sigma i} - \varphi_{i-1} C_i - \varphi_{i+1} C_{i+1} \quad (2.41)$$

$$Q'_{N-1} = Q_{N-1} + \varphi_D C_N + \varphi_G C_{G(N-1)} = \varphi_{N-1} C_{\Sigma(N-1)} - \varphi_{N-2} C_{N-1} \quad (2.42)$$

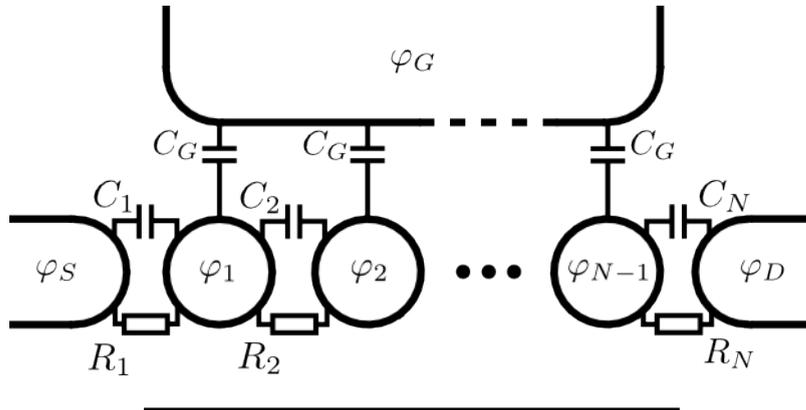


FIGURE 2.7: Diagram of a multi-island SET. The potentials and capacitances between each conductor are indicated.

which represents a system of  $N - 1$  coupled equations which can be written as a matrix equation. Each of  $Q'_i$  represent both discrete charges  $Q_i$  and induced charges  $Q_{Gi}$ ,  $Q_S$  and  $Q_D$ . The  $C_{\Sigma i}$  term represents the sum of each capacitance neighbouring island  $i$ :

$$C_{\Sigma i} = C_{Gi} + C_i + C_{i+1}. \quad (2.43)$$

Writing the matrix as a capacitance matrix, the system of equations takes the form

$$\begin{pmatrix} Q'_1 \\ Q'_2 \\ Q'_3 \\ \vdots \\ Q'_{N-1} \end{pmatrix} = \begin{pmatrix} C_{\Sigma 1} & -C_2 & 0 & \dots & 0 \\ -C_2 & C_{\Sigma 2} & -C_3 & \dots & 0 \\ 0 & -C_3 & C_{\Sigma 2} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & C_{\Sigma(N-1)} \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \varphi_3 \\ \vdots \\ \varphi_{N-1} \end{pmatrix} \quad (2.44)$$

which can be generalised as

$$\mathbf{Q} = \bar{\mathbf{C}} \cdot \mathbf{V}. \quad (2.45)$$

The system of potentials can then be solved by finding the inverse of the capacitance matrix

$$\mathbf{V} = \bar{\mathbf{C}}^{-1} \cdot \mathbf{Q} \Rightarrow \mathbf{V} = \bar{\mathbf{R}} \cdot \mathbf{Q}, \quad (2.46)$$

where  $\bar{\mathbf{R}} = \bar{\mathbf{C}}^{-1}$  is the reciprocal capacitance matrix. Therefore starting from a known charge configuration  $\mathbf{Q}$  specified by the potentials  $\varphi_G$ ,  $\varphi_S$  and  $\varphi_D$ , and each of  $n_i$ , all the potentials may be calculated which in turn allows the electrostatic energy to be calculated for a given charge configuration. The charging energy of each island can again be found by setting all the external potentials to zero, and finding the total capacitance seen by each island. In the case of a chain of islands where only an inter-island capacitance exists, the total capacitance of each island can be written as

$$C_{\Sigma}(i) = \left( \sum_{j=1}^{j=i} \frac{1}{C_j} \right)^{-1} + \left( \sum_{j=i+1}^{j=N} \frac{1}{C_j} \right)^{-1} \quad (2.47)$$

which shows that for islands near the middle of the chain, the capacitance seen by each island becomes much smaller, leading to a larger charging energy. When capacitances arising from the gate electrode and external charges are taken into account, the effect of series capacitances is compensated for by these parallel contributions. The total capacitance is more easily found by considering the change in potential caused by the addition of a single electron on the island  $i$  in consideration. When all  $j$  charges  $n_{j \neq i}$  are equal to zero and  $n_i = 1$ , then

$$\varphi_i = R_{ii} Q_i \Rightarrow C_{\Sigma}(i) = \frac{1}{R_{ii}} \quad (2.48)$$

where  $R_{ii}$  is the  $i$ -th diagonal element of the reciprocal capacitance matrix  $\bar{\mathbf{R}}$ .

After solving the island potentials through Equation 2.46, the voltages across each junction can be computed as  $V_1 = \varphi_S - \varphi_1$ ,  $V_i = \varphi_{i-1} - \varphi_i$  and  $V_N = \varphi_{N-1} - \varphi_D$ , and the island-gate voltages as  $V_{Gi} = \varphi_i - \varphi_G$ . The total electrostatic energy of this system is then written as

$$U(n_1, n_2, \dots, n_{N-1}) = \frac{1}{2} \sum_{i=1}^N C_i V_i^2 + \frac{1}{2} \sum_{i=1}^{N-1} C_{Gi} V_{Gi}^2 - \varphi_S(n_+ e + C_1 V_1) - \varphi_D(n_- e - C_N V_N) \quad (2.49)$$

where each potential implicitly depends on the charge state defined by the set of values  $n_i$  and where  $n_+$  and  $n_-$  represent the number of electrons the source and drain have supplied into the chain. This equation is suitable to numerically compute the change in free energy caused by any particular tunneling event. The free energy can also be expressed in terms of the reciprocal capacitance matrix[89]

$$U(n_1, n_2, \dots, n_{N-1}) = \frac{1}{2} \mathbf{V}^T \cdot \bar{\mathbf{C}} \cdot \mathbf{V} = \frac{1}{2} \mathbf{Q}^T \cdot \bar{\mathbf{R}} \cdot \mathbf{Q} \quad (2.50)$$

which in the last form explicitly contains the set of charge states  $n_i$ , allowing direct calculation of the free energy. When analysing a particular transition however, it is more helpful to consider the change in voltage before and after a single tunneling event

$$\Delta U = -\frac{1}{2} e (V_i + V'_i) \quad (2.51)$$

where  $V_i$  is the voltage across junction  $i$  before the tunneling event, and  $V'_i$  is the voltage after. The tunneling rates across individual junctions are then easily calculated through Equation 2.22.

Figure 2.8 (a) depicts the potential profile of a chain of 10 islands, where a potential is applied symmetrically at each electrode with  $\varphi_S = 0.5$  V and  $\varphi_D = -0.5$  V, and the potentials are solved using Equation 2.46. The charging energy of each island is plotted in Figure 2.8 (b), where the curves correspond to the legend in (a). In this instance, the junction capacitance used was  $C = 3 \times 10^{-18}$  F, the width of each junction was 2 nm and the width of each island was 4 nm. The solid black line shows the case where there is no external gate capacitance coupling to each island, resulting in the voltages across each junction being equal. In this case, the increase in charging energy is most effective, since there are no parallel capacitances to each island. The solid blue line shows the case where the gate capacitance is a tenth the value of the junction capacitance, which results in a decrease of the voltages across junctions deep in the chain. This effect is due to the sharing of the electric field coming from the island, between the neighbouring islands and the externally coupled entity, be it a gate electrode, external charge or stray capacitance. The solid green line represents a more pronounced case where the

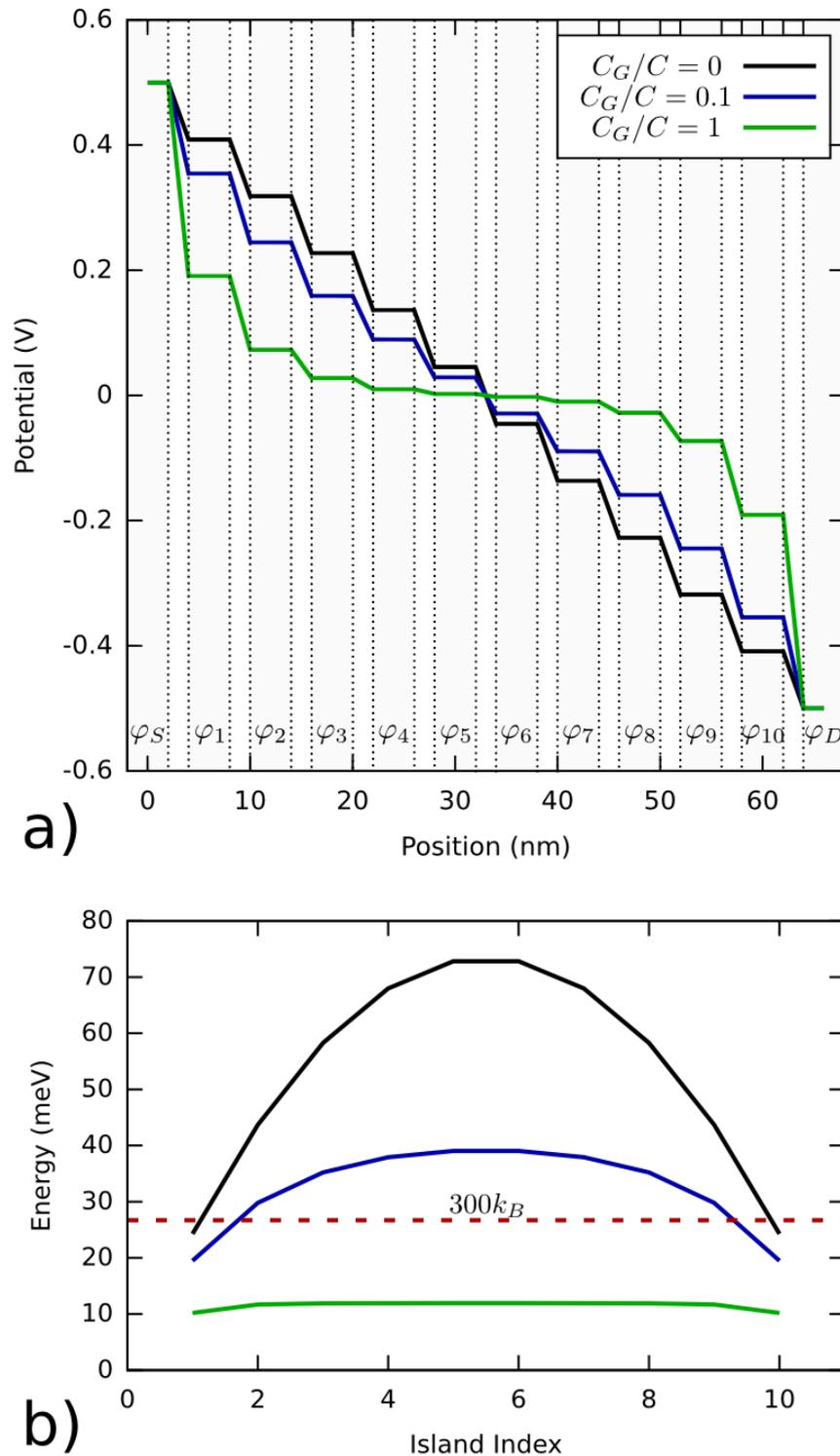


FIGURE 2.8: (a) Plot of the potentials of each island, and the resulting potential differences across each junction in a system of 10 islands. The effect of an increasing gate capacitance is shown. (b) Plot of the charging energy of each island, where the curves correspond with the legend of (a). The thermal energy at 300 K is represented by the red dashed line.

gate capacitance is of the same order of magnitude of the junction capacitance. In this instance a significant amount of the coupling between islands is lost to the gate, resulting in a severe suppression of the increase in the charging energy along the chain. The voltages across the boundary junction of the chain become larger as a result. The distribution of the voltages across each junction is characterised by the ratio of the gate capacitance to the junction capacitance

$$\alpha_J = \frac{C_G}{C}, \quad (2.52)$$

and the number of junctions in the chain. For a non-zero ratio, there always exists an upper limit to the number of islands that can be used to increase the charging energy within the chain. The result in long chains is the existence of single electron solitons in the regions far from the electrodes, where the potentials between neighbouring islands are initially near-zero as in the middle of the green curve of Figure 2.8 (a).

The effect of the gate voltage in this system is determined by analysing the second form of Equation 2.50. As before, when the source and drain voltages are negligible, the charge of each island  $i$  reduces to  $\varphi_G C_{Gi} - n_i e$ . Initially all  $n_i$  are zero. If all the gate capacitances are equal to one another, the free energy is minimised in periods of  $e/C_G$  when all islands have gained an extra electron, implying that a transconductance peak occurs at  $V_G = \pm e/2C_G$ . The gate voltage threshold for a single electron entering island  $i$  directly from the source or drain can be calculated by minimising the change in free energy through Equation 2.50, which after some rearrangement gives

$$\phi_G = \frac{e}{C_G} R_{ii} \left[ \sum_{j=1}^{N-1} (R_{ji} + R_{ij}) \right]^{-1} = \gamma_i \frac{e}{C_G} \quad (2.53)$$

where all the gate capacitances are again assumed equal. From this equation it is clear that  $\gamma_i < 1$ , indicating that a single electron may enter the chain at a fraction of the threshold voltage of a single island system with the same gate capacitance. Importantly, the term  $\gamma_i$  decreases along the chain and has its smallest value in the middle. This indicates that once a single electron is allowed to enter the chain, the most stable configuration is for the single electron to migrate towards the middle of the chain. Therefore the values of  $\gamma_1$  and  $\gamma_{N-1}$  determine the gate voltage threshold for admitting in the first uncompensated electron into the chain. For a specific gate voltage, the change in free energy for a specific transition from island  $i$  to  $j$  can be written as

$$\Delta U(i \rightarrow j, \varphi_G) = e^2(R_{ii} - R_{jj}) + e\varphi_G C_G \sum_{k=1}^{N-1} (R_{kj} - R_{ki} + R_{jk} - R_{ik}) \quad (2.54)$$

where the first term on the right-hand-side represents the difference in charging energy

for island  $i$  and  $j$ , in accordance with Equation 2.48. Beginning with  $\varphi_G = 0$  and assuming a perfectly homogeneous chain, Equation 2.54 shows that only transitions migrating out of the chain from the middle are favourable, as the charging energy of each island decreases in these directions. Now as  $\varphi_G$  is increased, transitions towards the middle of the chain become more favourable. When  $\varphi_G = e\gamma_1/C_G$ , an electron is able to enter the chain, and the transitions towards the centre of the chain then become favourable as the difference between the charging energies become smaller. At low temperatures and when the applied bias can be ignored, the electron becomes trapped in the middle of the chain and no conduction occurs. In contrast, at higher temperatures the thermal energy of electrons may easily exceed the small energy required to leave the chain, therefore the chain may become conductive under a small applied bias near the gate voltage threshold determined by  $\gamma_1$  or  $\gamma_{N-1}$ .

Figure 2.9 illustrates the behaviour of the potentials of each island under an applied bias (a) and when a gate voltage is applied (b). In this case a 5 island system is depicted with  $\alpha_J = 0.1$  and a symmetrical bias is applied to the source and drain. The transitions of single electrons between each island are highlighted by the arrows, where Equation 2.51 can be used to assess the favourability of the transition. It can be seen in (a) that even with a large bias, the transitions from the source towards the middle island are unfavourable at low temperatures, so that a small gate voltage must be applied to overcome this barrier. In (b), all the transitions from the source towards the drain are now favourable with a finite gate voltage applied. As discussed previously, at higher temperatures the small energy barrier to the initial transitions in (a) may

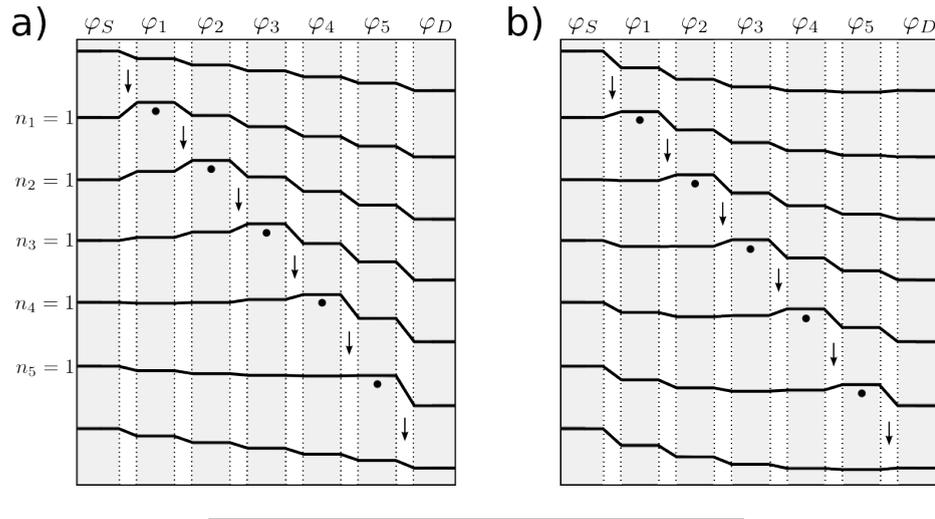


FIGURE 2.9: (a) Potential profile of a 5 island SET, where a source and drain voltage are applied, and the gate voltage is zero. (b) Same system as in (a) but with a finite gate voltage applied, making the initial transitions from the left energetically favourable. The dots indicate single-electron occupancy of the island, and the arrows highlight the voltages before (tail) and after (tip) the tunneling event.

easily be overcome, such that the channel becomes conductive. The behaviour of the conductance and the transconductance of this system is difficult to assess without a statistical simulation of the probable events for a given configuration. A single state of the system is defined through the capacitance matrix, the set of island occupations  $n_i$ , the gate potential  $\varphi_G$ , the source and drain potentials  $\varphi_S$  and  $\varphi_D$  and the temperature  $T$ . The dynamics of each junction can then be described through Equation 2.22 as long as the assumptions of the orthodox theory are satisfied. The change in electrostatic energy due to any tunneling event between neighbouring islands must be considered, which then allows the transition rates between each state to be calculated. For an ideal finite voltage source connected to the source and drain, the current is completely determined by the decay rate of each state.

Up to this point the presence of external charges in a multi-island SET system has not been considered. As seen for the single island case, external charges cause a charge offset in the island, the magnitude of which depends on the coupling between the external charge and the island. Similarly, in this system external charges will cause offsets in the charges of each island. However if these external charges are scattered around the chain, each island  $i$  may have a different offset charge  $Q_{ei}$ . This presents a potentially significant drawback to the use of long chains for a number of reasons. If the offsets are significantly different, the alignment between the charging energy levels in each island may be significantly affected, and the chain may not be easily turned on as a result. If there is a high density of trapped charge in the region surrounding the chain, the offset may be homogeneous across each island, however the external charges may be effectively screened. The screening of external charges poses a significant challenge to achieving a high charge detection sensitivity and will be discussed in the last section. The charge detection sensitivity is itself fundamental to the application of this device as a biosensor, and thus the limitations caused by noise are described in the next subsection, and values of the charge sensitivity are estimated in terms of theoretical and experimental results from the literature.

### 2.2.3 Charge sensitivity

The external charge sensitivity of the single and multi-island systems under the orthodox theory is determined by the change in current  $\delta I$  caused by a change in the external offset charge  $\delta Q_e$ , resulting from the modulation of the change in free energy for the relevant tunneling events. If  $\delta Q_e$  is much smaller than  $e$ , the change in current can be expressed as

$$\delta I = \frac{\partial I}{\partial Q_e} \delta Q_e. \quad (2.55)$$

A similar expression can be obtained for the sensitivity of the current to the gate voltage by replacing  $\delta Q_e$  with  $C_G \delta V_G$ . In the single island case, the sensitivity is thus

determined by the modulation of the rates in and out of the island as a function of the magnitude of the external charge, the applied bias, and the temperature of the system. In Figure 2.6 (b), the width and slope of the conductance peaks are controlled by the magnitude of the charging energy relative to the temperature, and the applied bias. In Figure 2.4 (b), the tunneling rate through a particular junction becomes very sharply defined by the threshold voltage at very low temperatures, which under the orthodox theory leads to very sharp oscillations of the conductance controlled only by the magnitude of the bias window (Equation 2.36). In this idealised regime, the SET may act as a switch, controlled by the gate. At higher temperatures, the threshold voltage is no longer well defined, and the conductance peaks thermally broaden as transitions between islands become more favourable, decreasing the slope of the conductance leading to the peak. As the thermal energy exceeds the charging energy, the oscillations are significantly suppressed and the SET cannot be turned off. By comparison in the multi-island system, since it is possible for each island to independently couple to external charges, a change in current could be determined by a fraction of the islands in the chain. As the charging energy of islands in the middle of the chain is largest, they have the greatest potential to modulate the total current through the chain.

In practice, the sensitivity of these SET systems is limited by a number of noise sources. Each source causes current and voltage fluctuations with a certain frequency distribution, that may drown out conductance oscillations and cause significant uncertainty in fast measurements when averaging is not used. The sensitivity is determined by the smallest detectable fluctuations of the current due to the external charge, as described by Equation 2.55. In general, noise processes are characterised by a power spectral density that specifies the average power of the noise at a given frequency. The total noise present in the measurements is then determined by the bandwidth of the system  $\Delta\omega = \omega_2 - \omega_1$  and the power spectral densities  $S_s$  of the noise sources  $s$

$$P_{\text{noise}} = \sum_s \int_{\omega_1}^{\omega_2} S_s(\omega) d\omega. \quad (2.56)$$

If the power spectral density is ‘white’ or approximately frequency independent in the measurement bandwidth, the noise power is then proportional to the bandwidth. The bandwidth is determined by the frequency response of the measurement system connected to the SET including external filters. The r.m.s. noise current  $i_{\text{noise}}$  can then be found from

$$P_{\text{noise}} = i_{\text{noise}}^2. \quad (2.57)$$

The minimum detectable external charge is then found by comparing Equations 2.55 and 2.57 resulting in the general expression

$$\delta Q_e = \sqrt{P_{\text{noise}}} \left( \frac{\partial I}{\partial Q_e} \right)^{-1}. \quad (2.58)$$

Equation 2.58 gives the absolute minimum detectable charge in the given bandwidth[115]. Since measurement systems and experimental conditions differ, it is more convenient to consider the spectral minimum detectable charge  $\delta Q_e/\sqrt{\Delta\omega}$ , in units of  $C/\sqrt{\text{Hz}}$ .

The intrinsic noise of an SET system is treated in detail under the orthodox theory by A. N. Korotkov[50], and sheds light on the origin of shot and thermal noise in tunnel junctions. Equation 2.22 can be solved numerically in matrix form for all states of the system, and a spectral density  $S_{II}(\omega)$  of the total current  $I$  along the chain and spectral density  $S_{\varphi\varphi}$  of the average island potential can be derived, which describes the frequency distribution of the tunneling events in the system of junctions. The current and potential spectral densities are a function of the applied voltage, the temperature of the system, the capacitance matrix and all the tunnel junction impedances of the system. When the thermal energy is much greater than the charging energy and the applied bias, many electrons may tunnel back and forth between each electrode. This noise can be considered to be thermal in nature and therefore the power spectrum takes the form of Johnson-Nyquist noise. Single electron effects are completely suppressed in this regime as discussed, and therefore there is no correlation between tunneling events, and the junctions behave as normal tunnel junctions. When the thermal energy is much smaller than the charging energy and falls within the Coulomb blockade range  $V \ll e/2C_\Sigma$ , the intrinsic noise is completely suppressed ( $S_{II} = S_{\varphi\varphi} = 0$ ). In this range, quantum noise becomes much more important, such as random quantum fluctuations and cotunneling as described by Equations 2.20 and 2.21. Ignoring cotunneling effects, the sensitivity to small changes in the external charge in this range is 0, since no current flows at all. The maximum sensitivity to the external charge can be obtained at the threshold of Coulomb blockade by the applied bias  $V \approx e/2C_\Sigma$  or by the gate voltage  $V_G \approx e/2C_G$ , where only a very small charge can turn the channel on. The minimum detectable charge  $\delta Q_e$  per unit bandwidth ignoring all noise sources other than the intrinsic noise is therefore determined by the noise floor at the threshold of Coulomb blockade. In this state,  $\delta Q_e$  is approximately [116]

$$\delta Q_e \approx 1.9C_\Sigma \sqrt{R_\Sigma k_B T \Delta\omega} \quad (2.59)$$

derived from Equation 2.58, where for a two junction SET  $R_\Sigma = R_1 + R_2$  and  $C_\Sigma = C_1 + C_2 + C_G$ . This expression is only valid near the Coulomb blockade threshold, in a small bandwidth, and at low temperatures according to Equation 2.27. With values

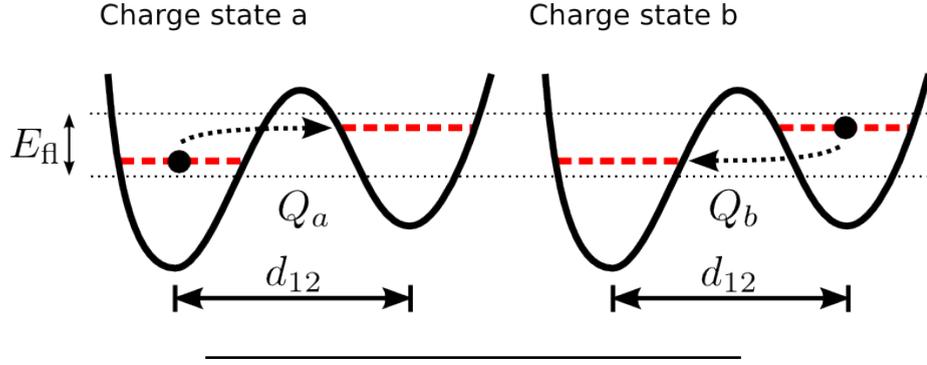


FIGURE 2.10: The two charge states of a two-level system. The particle can tunnel between the two levels shown in red, across the barrier which determines the decay rate of each state. The excitation energy  $E_H$  is usually much smaller than  $k_B T$ .

$C_\Sigma = 1$  fF and  $R_\Sigma = 200$  k $\Omega$ , and at  $T = 100$  mK, the maximum external charge sensitivity is  $\delta Q_e \approx 6.2 \times 10^{-6} e \sqrt{\text{Hz}^{-1}}$ . In a small bandwidth this represents a very small fraction of the elementary charge  $e$ . Therefore in a system with very small  $C_\Sigma \leq 1$  aF, and at higher temperatures, a similar ultimate intrinsic charge sensitivity is possible. For a multi-island system, by treating the islands either side of the middle island as electrodes with an effective resistance and capacitance, Equation 2.59 can be used to a first approximation. The effective capacitance seen from the middle dot  $C_\Sigma(N/2)$  is considerably reduced. The effective resistance can be treated as due to the series contributions from each junction  $R_\Sigma = \sum_i^N R_i$ . Using  $C_\Sigma = 1$  aF,  $T = 300$  K, and  $R_\Sigma = 2$  M $\Omega$  for example, an ultimate sensitivity  $\delta Q_e \approx 1.1 \times 10^{-6} e \sqrt{\text{Hz}^{-1}}$  is obtained, which is on the same order of magnitude as in the single island case. Typically in charge sensitivity measurements the smallest bandwidth achievable is desired. Lock-in amplifiers are typically used, which is described in Section 3.4.2. Very low frequency measurements can be performed using this system, usually less than 1 Hz, and the bandwidth can consequently be lower than 1 Hz. However in this frequency range, low frequency noise processes limit this intrinsic charge sensitivity in practice.

An important source of noise that is frequently encountered in experiments with SETs is  $1/f$  noise[60, 64, 117–121]. This extrinsic source of noise is seen in most electronic devices operating at low frequency and may have different origins depending on the device. In conventional metallic SETs,  $1/f$  noise usually dominates over the intrinsic noise described above. The noise spectral density is characterised by an inverse law

$$S_f(\omega) = \frac{1}{a\omega^b} \quad (2.60)$$

where  $a$  and  $b$  are empirical fitting parameters characterising the amplitude and slope of the spectral density, respectively. The most plausible explanation for the origin of this type of noise is the dynamic behaviour of external impurity charges trapped in dielectric materials surrounding the device. Impurity charges can contribute a static charge, which

is part of the external offset charge  $Q_e$ , and a dynamic charge that fluctuates. The simplest model of the origin of  $1/f$  noise is to consider two particle traps that are close enough together for tunneling to occur, i.e. when  $d_{12}$  is small, allowing electrons or atoms to tunnel between the two states[120]. The potential energy of this situation is illustrated in Figure 2.10 where the excitation energy of the fluctuator  $E_{\text{fl}} = \sqrt{U^2 + \Lambda^2}$  is highlighted, and the two resulting different charge states  $Q_a$  and  $Q_b$ . Here  $U$  is the difference between the two energy levels, and  $\Lambda = \hbar\omega_{\text{ph}}e^{-\lambda}$  is the broadening due to the vibrational mode of the system  $\omega_{\text{ph}}$ , and  $\lambda$  characterises the opacity of the potential barrier between the energy levels[122], analogously to the exponent  $2\kappa t$  in Equation 2.17, and therefore depends on  $d_{12}$ . The charge spectral density of this system can be expressed as[120]

$$S_Q(\omega) = \frac{(Q_a - Q_b)^2}{\cosh^2(E_{\text{fl}}/2k_B T)} \frac{\tau}{\omega^2 \tau^2 + 1} \quad (2.61)$$

where  $\tau$  is the relaxation time. The r.m.s. charge fluctuations can then be found from the spectral density. These external charge fluctuations give rise to a noise current found using Equation 2.55. Under normal conditions, the difference  $Q_a - Q_b$  is expected to be much smaller than the electronic charge  $e$ , since  $Q_a$  and  $Q_b$  correspond to the effective charges of slightly different configurations of the same particle. The excitation energy  $E_{\text{fl}}$  is also usually expected to be much smaller than the thermal energy such that the denominator of the first factor in Equation 2.61 is unity. Experimental and theoretical studies show that a group of fluctuators with a distribution of relaxation times and energy gaps are a plausible origin for  $1/f$  noise in SETs[117, 119, 121]. The spectral distribution of the r.m.s. external charge fluctuations due to each contributing fluctuator  $i$  can then be estimated as[118]

$$Q_{\text{noise}}(\omega) = \left( \sum_i \frac{\Delta Q_i^2 \tau_i}{\omega^2 \tau_i^2 + 1} \right)^{\frac{1}{2}} \quad (2.62)$$

where  $\Delta Q_i$  is the difference in charge between the two states of fluctuator  $i$ , and  $\tau_i$  is the relaxation time of fluctuator  $i$ .

From these equations it is clear the external charge noise increases with increasing fluctuator concentration and increasing relaxation time. When  $\omega\tau_i \ll 1$ , the charge noise is large, and for a single fluctuator,  $Q_{\text{noise}}(0) = \Delta Q_i \sqrt{\tau_i}$ . To characterise  $1/f$  noise, the value of  $\tau_i$  is expected to be on the order of 10 ms, corresponding to a cutoff frequency of 100 Hz. The charge sensitivity when  $1/f$  noise is present is often between 10 and 100 times worse at low frequencies[60, 118], giving values on the order of  $\delta Q_e \approx 10^{-4} e \sqrt{\text{Hz}}^{-1}$ . The  $1/f$  noise due to fluctuating impurity charges can effectively be reduced by suspending the SET islands to separate it from materials by air gaps, as has been shown experimentally[64]. Nonetheless, such a high charge sensitivity is likely to be

adequate for the detection of a single molecule, provided room temperature operation can be achieved. As will be discussed in Section 2.3.3, the charge of single biomolecules depends on their composition, their environment, and their size. Saving the details for that discussion, the total biomolecule charge can be substantially greater than  $e$  under the right conditions, and in such cases and using a low bandwidth, the SET charge sensitivity is more than adequate. However, only a certain fraction of the total charge of the biomolecule, depending on the geometry of the coupling and screening effects, will play a role in shifting the island charge, and therefore the charge sensitivity is discussed separately.

## 2.3 SET biosensor design

This section will explore how an SET biosensing device would be formed in practice, given the theoretical considerations discussed in the previous sections. This section begins with a description of the components required for an SET biosensor system that fulfils the challenges laid out in Chapter 1. The broad practical requirements of an SET system applied as a biosensor are then described in terms of the most important challenges. In terms of the SET technology, the first most pressing challenge is increasing the charging or addition energy significantly beyond the thermal energy at room temperature, 26.7 meV. This challenge is addressed by estimating the required properties of the system through the orthodox theory described previously. The second most pressing challenge is the coupling between the islands and the target charge, which must occur across a bioelectronic interface. The basic properties of such an interface and the implications for the system are discussed.

### 2.3.1 Proposed system architecture

Figure 2.11 shows a schematic that summarises the important features of a SET biosensor system that are required for multiplexed diagnostic of a panel of biomarkers, employing an array of  $N$  devices. The envisaged end-user device is a point-of-care device such as shown in Figure 1.2. The biological input to the system would be, for example, a sample of the user's blood. Once instructed through the user interface, the device would begin a diagnostic sequence for a particular biomarker panel, chosen by the user. In a device designed for diagnosing a specific panel, a simple microfluidic system would be used to deliver the sample to the array of floating gate surfaces. For a given biomarker panel, each biomarker antigen requires a specific antibody tethered to a floating gate surface. The antibody itself must be tethered to a surface that is chemically functionalised accordingly, the whole structure can then be called the capture molecule for a specific biomarker. The deployment of the capture molecules in the sensor are discussed in more detail in further paragraphs. To make the control measurements

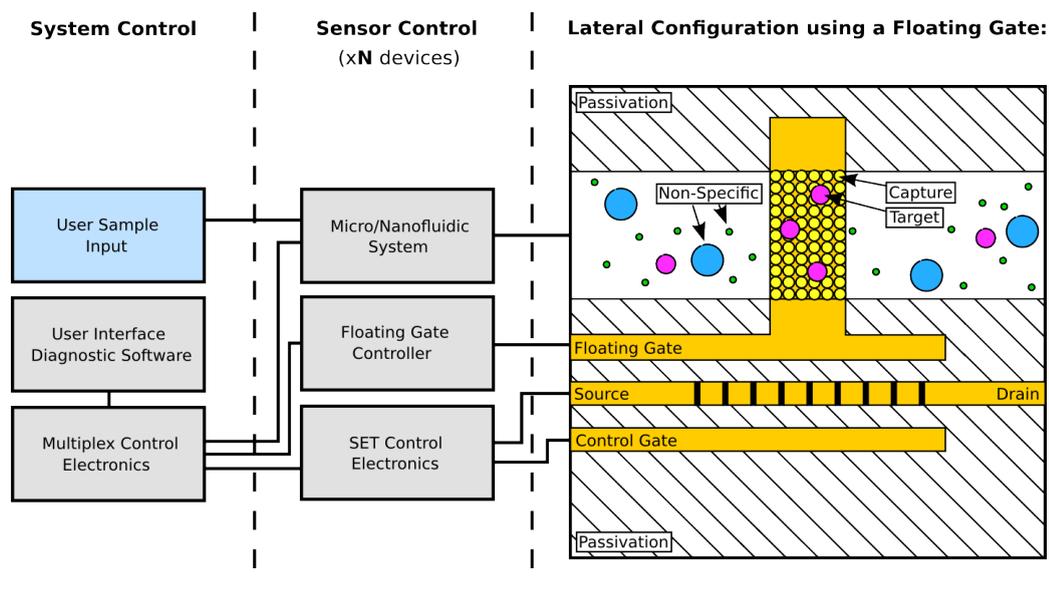


FIGURE 2.11: Schematic of the proposed SET biosensor system, showing the important components that allow full control of the system.

(no binding), the SET must be biased to its most sensitive point at the threshold of Coulomb blockade, as discussed in the previous section. This is achieved with the gate, which serves to set the bias point of the SET. The current or voltage of the channel is then measured using suitable peripheral electronics. After the initial measurements, the floating gates can be exposed to the blood sample. The biomarkers, once in the sensing area, must diffuse through the medium to the desired floating gate so that the binding reaction can take place. The geometry and material properties of the microfluidic channel will influence the average diffusion time. After sufficient time was allowed for the biomarkers to bind, another measurement is made to find the change in external charge of the floating gate, caused by the binding event. If this is performed on a biomarker panel, the results of the measurements can be analysed by the software and a diagnosis can be reported with a certain confidence. Admittedly, this technology is highly ambitious for a number of reasons, and not only the challenges associated with SETs operating at room temperature. Before looking at the optimisation of the SET system, some of the other challenges surrounding the SET biosensor are briefly discussed for completeness.

For dilute concentrations of a target species, the volume of the chamber becomes an important factor to consider. For example, given a molar concentration  $c_i = 1$  fM of a species  $i$ , there is on average one molecule of that species in a  $1/c_i N_A \approx 1.66$  mm<sup>3</sup> (166 nL) volume. If the sensing area chamber volume is smaller than the volume containing on average one target particle, the fluid must be pumped whilst ensuring enough time is allowed for diffusion and reaction to occur.

Typically in analytical experiments where the blood plasma is required, the blood

must be filtered to remove large cells, such as red and white blood cells, to leave only the plasma containing the proteins of interest. Normally this is achieved by centrifugation, so that the constituents of the blood are phase separated by mass, leaving only the blood plasma at the top of the vial used in the centrifugation. Experiments integrating microfluidics and blood paper have shown however that it is possible to separate the plasma from the blood in a contained on-chip device[123, 124]. If a suitable plasma sample can be provided, the main challenge associated with the microfluidic system is then the delivery of the sample, limited by the properties of the channels that constitute the fluidic network, and the fluid properties of the blood plasma sample itself[125].

The capture molecules can be realised using self-assembled-monolayers (SAMs)[126], which have been widely used in biosensing platforms[127–129], and which may allow for the multiplexed detection of different biomarkers, improved speed and throughput of the sensor system, and reduced costs for a given diagnosis[130]. SAMs are a type of self-forming and self-packing molecular structure on a substrate surface, where self-assembly refers to the high reactivity and reproducibility of the chemistry, and a monolayer implies the chemistry is self-terminating. SAM chemistry is based on the fact that metal and metal-oxide surfaces can readily adsorb certain organic chemical groups into a stable dense structure, modifying the chemical properties of the surface. In the simplest case, a suitable solvent (or vapour), such as ethanol or water for example, carrying the desired organic molecules is

placed in contact with the substrate surface, and enough time is allowed for the molecules to diffuse and react with the substrate, until SAM coverage reaches almost 100% and optimal packing density is achieved. The part of the SAM molecule that reacts with the substrate surface is named the head group, and the rest of the molecule consists of a hydrocarbon chain, and an end group that under dense packing conditions determines the chemical properties of the surface after SAM formation[126]. The most commonly used head group is the thiol group  $R-S-H$ , where  $R$  is the chain and end group. The thiol group reacts with a clean Au surface, and forms  $S-Au$  bonds up to  $50 \text{ kcal.mol}^{-1}$  in strength[131], which represents approximately 2.1 eV per bond, significantly larger than the room temperature thermal energy 27 meV. Figure 2.12 shows the chemical structure of an alkylthiol SAM bonded to a gold surface, with 4 methylene groups. The length of the chain and the end group can be synthesised so that a wide variety of SAM

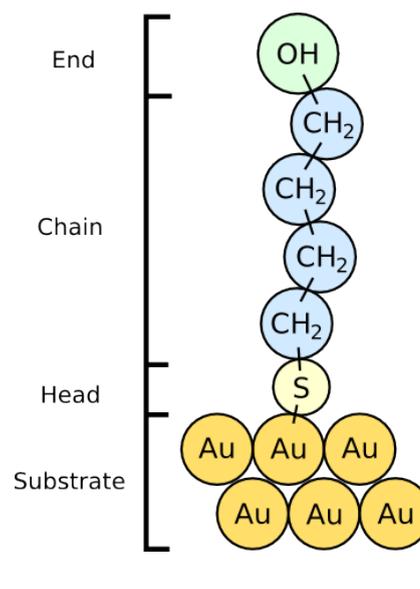


FIGURE 2.12: Diagram of an 4-alkylthiol molecule with a hydroxyl end group, tethered to an Au surface by a sulfide bond.

molecules can be created with specific properties before self-assembly. It is also possible to further modify the chemistry of a preformed SAM by reaction or adsorption of new surface functional groups. For example, poly ethylene glycol (PEG) modified alkylthiolates have been extensively used to protect an Au surface from non-specific binding of biomolecules[132], since the end groups of these molecules are highly inert. This method has been successfully applied to the multiplexed detection of different proteins tethered to peptide aptamer functionalised Au electrodes, using electrochemical impedance spectroscopy (EIS) to measure the change in impedance of the biological interface after adsorption[133].

The use of SAM tailored surfaces provides a means of designing the surface chemistry of the floating gates such that the affinity is high only to the target biomolecules, reducing non-specific binding. The floating gate is capacitively coupled to the SET sensor so that changes in the surrounding charge polarise the floating gate, inducing a charge in the SET islands. The floating gate surface is very difficult to clean in a microfluidic system if it cannot be electrically addressed. Through the use of a highly insulating and isolating switch, or an additional control gate[134], the potential of the floating gate could be controlled, allowing the use of electrochemical methods within the microfluidic system. With a switch, the floating gate would be connected to a true voltage source, allowing a current to flow. The control gate method doesn't provide a true voltage source, rather the floating gate potential is matched to that of the control gate. The coupling between the SET, the floating gate and the adsorbates is discussed in a later subsection. It has been shown that a combination of SAM chemistry and electrochemistry can be used to functionalise individual electrodes[132, 135]. Using an Ag/AgCl reference electrode, thiol based SAMs formed on a particular electrode were removed by applying a 30 s 1.4 V pulse to the electrode in a phosphate-buffered saline (PBS) solution. This allows different thiol based SAMs to be used with the cleaned electrodes, providing a route to the selective functionalisation of different electrodes. In addition to this it is possible to use such electrochemical methods to enhance the SAM formation rate down to a matter of minutes[136], showing that such a system could be a viable means of selectively functionalising large sensor arrays[137]. It has also been demonstrated that these processes can be used at the nanoscale, where electrodes separated by even a nanoscale gap can be functionalised differently[138], allowing DNA strands to be anchored with DNA modified alkylthiol SAMs across the nanogap[139]. Figure 2.13 shows an example circuit diagram and schematics of a lab-on-a-chip biosensor array, combining microfluidic, SAM and electronic technology. A pump or force establishes a flow rate of a carrier solution, valves control which fluids are allowed to flow through the sensing chamber, so that a recipe can be sequenced for the formation of SAMs before admitting the sample. When the floating gates are prepared, control measurements can be performed to measure the initial charge. The sample is then admitted into the sensing chamber and

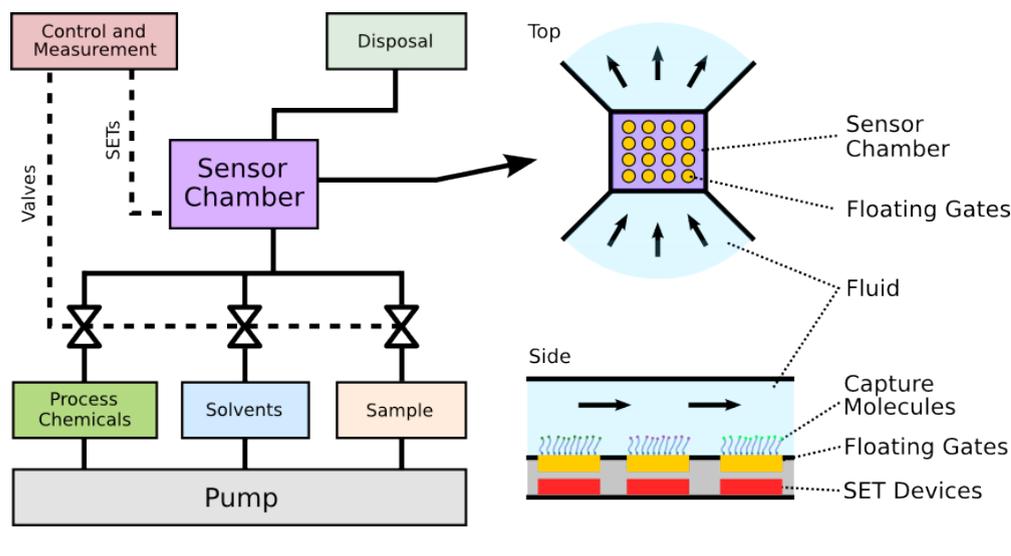


FIGURE 2.13: Fluidic (solid lines) and control (dashed lines) circuit diagram for an example self-contained SET biosensor, with capabilities for functionalising the floating gates on chip. Multiplexing is achieved if the floating gates can be functionalised with different capture molecules, so that multiple antibodies can be detected separately at the same time.

measurements of the change in charge can be recorded. Such a device is not necessarily important for POC applications however, since the floating gates could in principle be prepared before distribution to detect a specific panel of biomarkers only. However if more sophisticated devices were required, such as in a research setting, a more versatile but more complicated system, such as shown in Figure 2.13, could in principle be used.

The SET is coupled to the floating gate, but should be shielded from the fluid and unwanted external sources of charge. The microfluidic channels are usually either made from PDMS and/or  $\text{SiO}_2$ , materials with excellent resistance to organic chemicals used in biomedical assays[140]. Passivation of the SET against the biological environment can be achieved by the floating gate material and the insulator that separates the SET islands and floating gate. ALD  $\text{Al}_2\text{O}_3$  has been shown to be an excellent passivation material for organic transistors[141], however its amorphous structure may lead to a significant charge offset or fluctuation noise in the SET islands[142], due to the random atomic potential causing electron localised states. Alternatively to using metallic floating gates, the capture molecules could be formed directly on the passivation surface to directly couple the target biomolecule to the channel. For example, *n* and *p*-type Si-nanowires coated in aldehyde propyltrimethoxysilane (APTMS) to form aldehyde groups, can be subsequently functionalised with specific antibodies for label-free biomolecule detection[15]. Some SAM chemistries for  $\text{Al}_2\text{O}_3$  surfaces also exist nonetheless[143], which could lead to solutions becoming available for biomarker binding with this material also.

### 2.3.2 SET optimisation for operating temperature

Two configurations may be discerned in the multi-island SET system, by comparing the gate-junction capacitance ratio  $\alpha_J$  and the number of junctions  $N$ . An in depth analysis of the properties of long chains of tunnel junctions is given by Bakhalov *et al.* [144]. Considering an infinite system of series junctions with equal parameters, the decay constant of a soliton solution of the potential deep inside the chain is found to be

$$\lambda = \operatorname{arccosh} \left( 1 + \frac{\alpha_J}{2} \right) \quad (2.63)$$

which has units of inverse number of junctions. The width of the soliton is then

$$W = \frac{2}{\lambda} \quad (2.64)$$

where  $W$  is considered in both directions along the chain. It is clear from Equation 2.63 that when  $\alpha_J = 0$  the width is infinite, as expected, since all the voltages of the junctions in the system are affected by the addition of an electron on an island in the middle of the chain. For finite values of  $\alpha_J$ , and in an infinite chain, there may in principle be an infinite number of solitons within the chain. This leads to interactions between solitons, consisting of repulsion between like-charge solitons, and attraction and subsequent annihilation between opposing charge solitons. In a finite system with a specific number of junctions, there are then two configurations that lead to different behaviour.

**Soliton operation** when  $W \ll N$ . In this case, voltages applied at the source and the drain decay at the same rate  $\lambda$  into the chain, and once electrons are able to tunnel into the chain, a stream of solitons of width  $W$  will flow into the system, the number limited by the strength of the repulsion between neighbouring solitons. The solitons can then be trapped by applying a negative voltage on both the drain and the source electrodes. The charging energy of each island becomes constant beyond  $W/2$  junctions away from the source and drain electrodes. In this limit the charging energy of each island takes the value

$$E_C = \frac{e^2}{2\sqrt{C_G^2 + 4CC_G}}. \quad (2.65)$$

This operation is not suitable for higher temperature operation as excessively small values of  $C$  and  $C_G$  are required for small values of  $N$ .

**Coulomb operation** when  $W \approx N$  and  $W \gg N$ . In this case, almost all the junctions in the chain are affected by the presence of a single electron in the chain. Only a single soliton is permitted to exist in the chain, since the addition of subsequent electrons is prevented by the Coulomb blockade. The charging energy in this system increases up to the middle region of the chain, which is of greatest interest to increasing

the operating temperature of the chain. This mode of operation is closer to the original single-island system, which only allows sequential tunneling of single electrons (under the orthodox theory).

For a sensor system working at higher temperatures, the case when  $W \approx N$  is more suitable since the charging energy can be increased. However, paradoxically, for a stronger coupling of the target charge, a large value of  $C_G$  is required, therefore a trade-off is established between the number of islands and the value of  $\alpha_J$ . Considering the cases described in Figure 2.8, the soliton widths are found to be infinite for the  $\alpha_J = 0$  case,  $W = 2$  for  $\alpha_J = 1$  and  $W = 6$  for  $\alpha_J = 0.1$ . For  $\alpha_J = 0$  and  $\alpha_J = 0.1$ , the circuit should operate in the Coulomb mode so that only single electrons can exist in the chain. For  $\alpha_J = 1$ , more than one electron may enter the chain, and therefore the circuit operates in the soliton mode. As both the sensitivity of the chain to external charges, requiring a large value of  $C_G$ , and a high charging energy, requiring a small value of  $\alpha_J$ , are important, the number of islands in the system must be limited such that  $W \geq N - 1$ . The condition can then be expressed as

$$N - 1 \leq \frac{2}{\operatorname{arccosh}\left(1 + \frac{\alpha_J}{2}\right)} \quad (2.66)$$

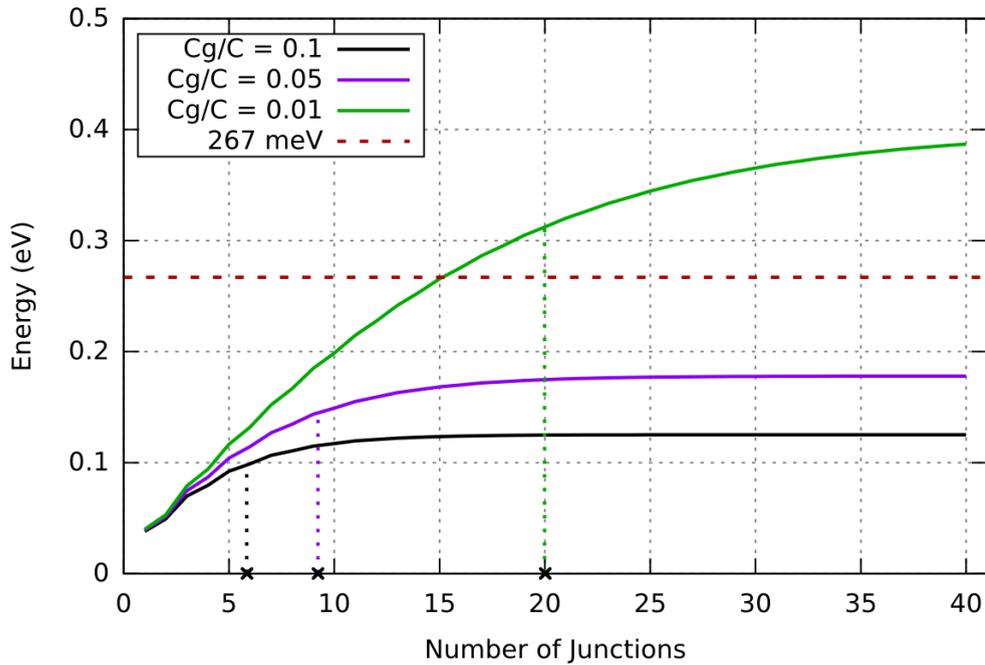


FIGURE 2.14: Plot of the charging energy of the middle island of a multi-island SET, against the number of islands. The red dashed line represents the energy condition imposed by Equation 2.27, and the crosses on the x-axis correspond to values of the soliton width  $W$  for each  $\alpha_J$ , calculated through Equation 2.64.

which restricts the range of interest of the number of islands for a given value of  $\alpha_J$ . Alternatively

$$\alpha_J \leq 2 \cosh\left(\frac{2}{N-1}\right) - 2 \quad (2.67)$$

can be used to estimate  $\alpha_J$  for a given number of junctions. These expressions do not provide a guide to the required values of  $C$  and  $C_G$  however, so they can only be used to estimate the optimal value of  $N$  or  $\alpha_J$  for a given system configuration.

The first consideration should be that the value of  $N$  needs to be large enough for the middle island charging energies to satisfy Equation 2.27, and yet retain a suitable value of  $\alpha_J$ . This problem can be solved iteratively by starting with a known value of  $C$ , and optimising for the value of  $N$  and  $\alpha_J$ , until the middle island charging energy exceeds  $10k_B T$ . Using a minimal value of  $C = 1 \text{ aF} = 10^{-18} \text{ F}$ , achievable at 10 nm length scales, Figure 2.14 shows a plot of the charging energy of the middle island against the array length, computed by finding the reciprocal capacitance matrix  $\bar{\mathbf{R}}$ , and using

$$E_C(i) = \frac{e^2}{2C_{\Sigma}(i)} \quad (2.68)$$

where  $C_{\Sigma}(i)$  is found using Equation 2.48. The condition set by Equation 2.27 is represented by the dashed red line. The soliton width  $W$  for each value of  $\alpha_J$  (approximately 6, 9 and 20) is indicated by the crosses on the x-axis. Clearly  $C$  must initially be small enough to have any chance of providing a large charging energy. The value of  $\alpha_J$  sets the limiting value the charging energy can take for large arrays, as discussed previously. With the current parameters, Equation 2.27 is then satisfied when  $N - 1 \geq 15$  and  $\alpha_J = 0.01$ , and if  $W \geq N - 1$ , a good balance between  $\alpha_J$  and  $N$  is obtained from the point of view of operating temperature optimisation. Thus to decrease  $N$  and increase  $\alpha_J$ ,  $C$  must be made even smaller.

### 2.3.3 SET optimisation for charge sensitivity

A number of configurations are available for coupling the target biomolecule charge  $Q_e$  to the SET islands. The charge can be directly coupled to the islands, such that the passivating insulating dielectric materials and the SAM biological interface form a capacitor between the charge and the SET islands, and thus a charge is directly induced into the SET islands, similarly to a gate electrode. Alternatively, the charge can be coupled to a floating gate, which itself is coupled to the SET islands, forming two capacitors in series to the island channel. Both methods impose constraints on the fabrication of the devices, and different considerations for the layout and operation of the biosensor. The origin of the effective charge of a biomolecule is the incomplete screening (by the medium and the macromolecule itself) of the charged chemical groups that form its macroscopic structure. To first approximation, the net charge of a biomolecule can

be assessed by summing the contributions from each ionisable chemical group  $i$ [145],

$$Q_{\text{net}} = \sum_i \frac{z_i e}{1 + 10^{z_i(\text{pH} - \text{p}K_i)}} \quad (2.69)$$

where  $z_i = +1$  for a basic group,  $z_i = -1$  for an acidic group, pH is the pH of the solution, and  $\text{p}K_i = -\log_{10}(K_i)$ , where  $K_i$  is the dissociation constant for group  $i$ .  $\text{p}K_i$  describes how many protons  $\text{H}^+$  are dissociated (associated) from (to) an acid (base) in solution at equilibrium, and hence describes the strength of an acid (base) in a solution. The pH of the solution determines the magnitude of the net charge of a single group in the solution. For example, for a single weakly acidic species with a value of  $\text{p}K_i = 8$ , if the pH of the solution is much less than this value, describing a high concentration of dissociated protons in the solution, then the weakly acidic species cannot readily donate a proton, and therefore the net charge becomes very small. Increasing the pH to match the  $\text{p}K_i$  value is termed the isoelectric point, at which the charges in the solution and those of species  $i$  are perfectly compensated. When the pH exceeds the  $\text{p}K_i$  value, the solution is more basic, and therefore protons are more readily donated, leaving a net negative charge on the acidic group. For a basic group, a net positive charge is contributed when the pH is lower than the  $\text{p}K_i$  value of the group. Therefore in a complex molecule such as a protein, formed of amino-acids with different individual charge properties, the net charge is a function of the pH, where if the pH is lower than the effective isoelectric point, the protein is net positively charged, and net negatively charged otherwise. Gel-electrophoresis is a widely used experimental technique that makes use of this to measure to separate proteins by mass and charge. According to Equation 2.69, the net charge of a large biomolecule can be significant under the right conditions, with values on the order of  $-100e$ [145].

In a solution of given pH, a protein can be described as a system of point charges distributed in space according to the molecular structure, surrounded by the solution ions. When the protein is adsorped onto a floating gate, the charge system changes slightly due to the reaction, and is brought in close proximity to the sensing surface. The effective charge seen by the sensing surface is therefore dependent on the electrostatic properties of the medium between the protein charge system and the sensing surface, governed by Poisson's equation

$$\nabla^2 \varphi(\mathbf{r}) = -\frac{\rho(\mathbf{r}, \varphi)}{\epsilon_r \epsilon_0}, \quad (2.70)$$

where  $\varphi(\mathbf{r})$  is the spatially dependent electric potential of the system,  $\rho$  is the charge density which may depend on position and the potential of the system itself, and  $\epsilon_r$  is the relative permittivity of the medium, assumed constant everywhere. The solution

carrying the proteins is usually a buffer solution, a weak acid-base mixture in an equilibrium state with a pH that is independent of small additions of strong acids and bases to the solution. The buffer solution therefore contains ions that can effectively screen the protein charges. The ionic strength characterises the concentration of ions  $j$  in the buffer solution,

$$I = \frac{1}{2} \sum_j N_A c_j z_j^2 \quad (2.71)$$

where  $c_j$  is the concentration of ions  $j$  in  $\text{mol.m}^{-3}$ , and  $z_j$  is the number of charges per ion. A buffer with a high ionic strength will have a large number of charged species or a smaller number of highly charged species, and as a result will have a high conductivity[146].

The classical Debye-Hückel theory can be used to estimate the effect of charge screening on the potential[147]. The concentration  $c_j$  is assumed to follow a Boltzmann distribution, where the activation energy is dependent on all the ions of the solution through the potential

$$c_j = c_{0j} \exp\left(-\frac{z_j e \varphi}{k_B T}\right), \quad (2.72)$$

where  $c_{0j}$  is the concentration with no external potential  $\varphi$ . The charge density  $\rho$  is then due to the sum of each charged species in the solution. Due to the complex resulting expression for  $\rho$  and Equation 2.70, a Taylor expansion is taken and only the first two terms are retained, resulting in

$$\rho = \frac{N_A \varphi}{k_B T} \sum_j z_j^2 e^2 c_{0j} = \frac{2N_A I e^2 \varphi}{k_B T}, \quad (2.73)$$

where the first term of the expansion is equal to zero due to charge conservation. The ions in the solution are approximated to be spherically symmetrical with minimum contact distance  $r_j$ , the potential is then solved by considering Poisson's equation (Equation 2.70) in radial form, which gives a solution for the potential local to ion  $j$

$$\varphi(r) = \frac{z_j e}{\epsilon_r \epsilon_0 r (1 + \kappa_D r_j)} e^{\kappa_D (r_i - r)} \quad (2.74)$$

where  $\kappa_D$  is the inverse Debye screening length, which describes the rate at which the potential due to the ion decays into the surrounding medium. The Debye screening length  $\lambda_D$  is independent of the geometry of the ions, and is defined as

$$\lambda_D = \sqrt{\frac{\epsilon_r \epsilon_0 k_B T}{2Ie}}. \quad (2.75)$$

The potential due to an ion is therefore more effectively screened in a buffer with high ionic strength and smaller permittivity. Therefore, the potential due to the charges in

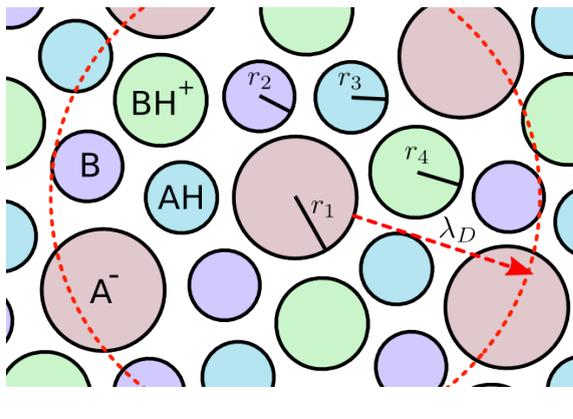


FIGURE 2.15: Schematic illustrating the Debye length in a buffer solution with an acid AH, a base B, and the conjugate base  $A^-$  and acid  $BH^+$  in equilibrium.

the biomolecule will decay into the solution. Figure 2.15 illustrates the Debye length in a buffer solution formed of an acid AH, a base B, and the conjugate base  $A^-$  and acid  $BH^+$ . Using Equation 2.74, the effective charge due to an ion  $j$  can be expressed as

$$Q_j(r) = z_j e \exp\left(\frac{r_j - r}{\lambda_D}\right) \quad (2.76)$$

for  $r \geq r_j$ , which shows that when  $|r_j - r| > \lambda_D$  the effective charge appears less than 37% of the original value at  $r = r_j$ . Optimisation of the buffer solution is therefore a critical consideration when performing a charge measurement in a buffer solution, and the effect of the solution on the activity of the proteins of interest must be carefully considered[148]. The total effective charge seen by the sensing surface is due to both the buffer solution and the target biomolecule, therefore as long as the pH of the buffer solution doesn't change when biomolecules bind onto a surface, the change in charge detected should be due only to the target biomolecules.

Recent measurements of the effective charge of biomolecules using a novel escape-time electrometry technique[145] have shown that the effective charge is a significant fraction (between 20 and 80%) of the net charge  $Q_{\text{net}}$  (Equation 2.69). A dependence of the effective charge on the size and structure of the biomolecules used was observed, which was found to be due to the internal screening properties of each molecule. Large spherical biomolecules contain many charged groups within the structure, leading to a screening of the internal charges by the exterior charges. In contrast, elongated molecules with less structural volume are less affected by the screening effect, and therefore the effective charge is a greater portion of the net charge in those cases. The effective charge seen by a planar sensing surface will be again a fraction of the effective charge of the biomolecule, due to the spatial distribution of the biomolecule effective charges, and additional screening from the medium between the biomolecule and the sensing surface. The Debye-Hückel theory is only valid for small concentrations  $c_j$  due to the

linear approximation used, however recent experiments show that in a high concentration solution, the Debye length increases with concentration[149]. The Bjerrum constant

$$l_B = \frac{z_j^2 e^2}{\epsilon_r \epsilon_0 k_B T}, \quad (2.77)$$

is the length scale at which ion interactions are comparable to the thermal energy. The Debye-Hückel theory breaks down when  $l_B^3 c_j \ll 1$ , and in this range it has been shown that the Debye length varies as[150]

$$\lambda_D \propto l_B c_j r_j^3. \quad (2.78)$$

At these high concentrations however, the proteins are at risk of losing their function or decomposing, so increasing the concentration to increase the screening length may not be a viable option.

Within a molecular layer, such as a SAM, trapped ions and defects are different to ions in solution. In a solution, an ion responds to the presence of all other ions and moves according to the forces in the liquid state, however when trapped in a potential well in a solid structure, a significant force may be required to displace the ions. If the activation energy of the traps and defects are much higher than the thermal energy and the concentration very low, the state is stable over time and the charge is static. Within the SAM layer, the distribution of charges will give rise to static potentials and each

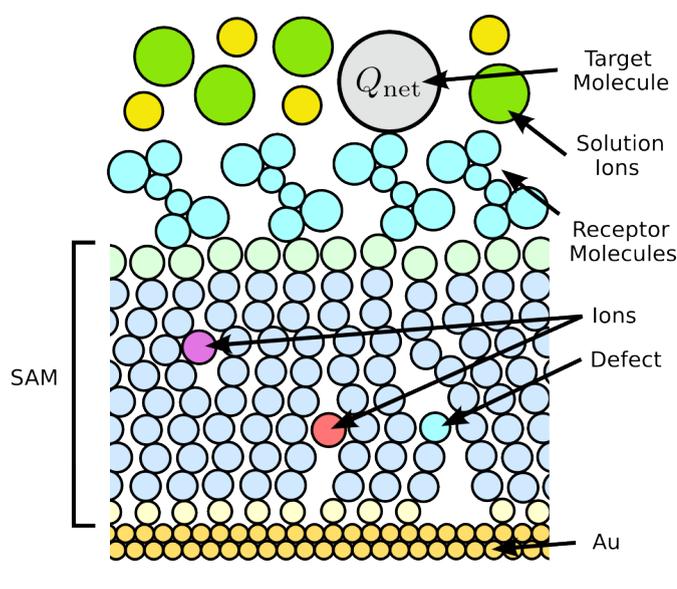


FIGURE 2.16: Illustration of trapped ions and defects in molecular layers, such as SAMs, that cause static and possibly fluctuating sources of charge. The solution ions are also illustrated, which contribute to the total charge seen by the SET and form a double layer at the interface with the SAM.

pair of ions  $i$  and  $j$  will experience a force

$$\mathbf{F}_{ij} = \frac{z_i z_j e^2}{4\pi\epsilon_0\epsilon_r |\mathbf{r}_i - \mathbf{r}_j|^2} \cdot \hat{\mathbf{r}}_{ij} \quad (2.79)$$

where  $\mathbf{r}_{ij}$  are the position vectors of each ion, and  $\hat{\mathbf{r}}_{ij}$  is the unit vector  $(\mathbf{r}_i - \mathbf{r}_j)/|\mathbf{r}_i - \mathbf{r}_j|$ . For ions close to the floating gate, the image force will be so strong that migration towards the floating gate is possible, at which point the ion is passivated. The stable charges will contribute a significant offset charge to the nearest SET islands in possibly an inhomogeneous way. In addition, a small spatial separation between traps or defects of similar energy give rise to charge fluctuations that impede the SET sensitivity, as discussed previously. If the molecular layer has insulating and uniform dielectric properties, which implies a low density of defects and trapped ions, the screening of the external charge should be minimal. If ions and defects can move within the molecular layer, then screening could be as effective as in a electrolyte solution.

The nature of the effective charge is therefore determined by the environment of the surface-bound biomolecule and the dielectric properties of the capture molecular layer. When the external charge is near a metallic surface, a charge is induced that depends on the capacitance between the external charge and the metallic surface. If the external charge is dimensionally very small, the self-capacitance (Equation 2.6) becomes very small, and therefore it is usually dropped from the capacitance matrix and the charge is expressed directly as an offset. When the point charge is coupled to a surface, the mutual capacitance will be very insensitive to the distance from the surface. For instance, the capacitance between concentric spheres is given by

$$C_{pc} = 4\pi\epsilon \left( \frac{1}{r_i} - \frac{1}{r_o} \right)^{-1}, \quad (2.80)$$

where  $r_i$  and  $r_o$  are the inner and outer diameters respectively. The value of  $r_o$  must be on the same order of magnitude as  $r_i$  to make a significant difference to the capacitance, so for a point charge coupled to a plane at a large distance, the capacitance will be at least on the order of the self-capacitance. With dimensions of  $r_i = 1 \text{ \AA}$ , an outer diameter of  $r_o = 1 \text{ nm}$  leads only to an 11% increase in the capacitance. With larger conductors such as patterned electrodes, the capacitance becomes considerable and therefore should be included in the capacitance matrix. Another important contribution that should be included is due to the coupling between the ions in solution and the conducting surfaces. The potentials of the conducting surfaces cause an electrical double layer to exist at material interfaces, the scale of which can be determined using the Debye-Hückel theory. The electrical double layer has an associated capacitance that is in parallel with capacitance associated with the external charge. The external charge can be coupled directly and indirectly to the SET islands:

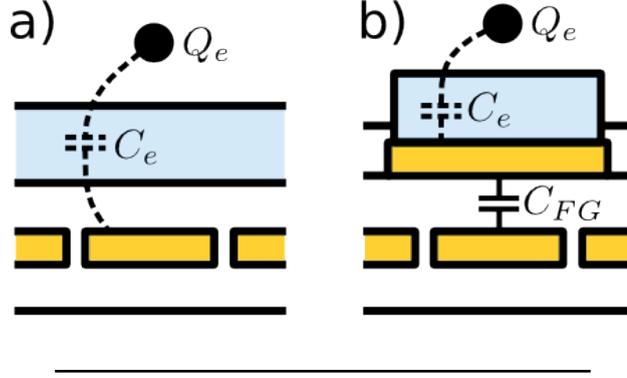


FIGURE 2.17: (a) Direct coupling of the external charge  $Q_e$  to an SET island, through a very small capacitance  $C_e$ . (b) Coupling of  $Q_e$  to the SET island through the floating gate. The potential of the floating gate is changed by the external charge, which can be detected through  $C_{FG}$ . The SAM layer is shown in blue, the SET components are shown in orange, and the uncoloured regions represent insulation or passivation layers.

**Direct coupling** is illustrated in Figure 2.17 (a) for an SET with a single island. In this case, the external charge  $Q_e$  shifts the electrostatic energy of the island by an amount  $eQ_e/C_\Sigma$  from Equations 2.30 and 2.32. Using the gate to set the electrostatic energy of the system at the threshold of Coulomb blockade allows a baseline measurement  $I_1$  to be taken, and when the charge  $Q_e$  is coupled to the island, the current is changed by an amount  $\Delta I(Q_e) = I(Q_e) - I_1$ . For sub-electron charge ( $Q_e < e$ ) measurements, the SET can be biased with a small enough voltage across the source and the drain such that the charging energy lies within the bias window (Equation 2.36) when the second measurement  $I_2$  is made. If  $Q_e$  is very large and the SET has periodic oscillations, the change in charge can not be measured quantitatively in a metallic system due to the indistinguishability of the oscillations associated with each charge state  $n$ . In a multi-island system,  $Q_e$  will be shared between each island  $i$  such that  $Q_e + \sum_i Q_{ie} = 0$ , so that each induced charge depends on the coupling between the external charge and the islands. If the source of the charge is very small compared to the dimensions of the island chain, the response of the system will depend on the location of the charge.

**Indirect coupling** is illustrated in Figure 2.17 (b), where the SET island is coupled to a floating gate, which is isolated to allow the potential to change according to the surrounding electrostatic environment. Starting from charge conservation the charge in an arbitrary multi-island system, the floating gate charge can be expressed as

$$Q_{FG} = - \left( Q_e + Q_S + Q_D + \sum_i Q_i \right), \quad (2.81)$$

where  $Q_i$  is the charge of island  $i$ . Proceeding in the same way as in Section 2.2.2, the potential of the floating gate can be expressed as

$$\varphi_{FG} = \frac{Q_{FG} + \varphi_S C_{FGS} + \varphi_D C_{FGD} + Q_e}{C_{F\Sigma}} + \frac{1}{C_{F\Sigma}} \sum_i \varphi_i C_{FGi}, \quad (2.82)$$

where  $C_{F\Sigma}$  is the total capacitance neighbouring the floating gate,  $C_{FGS}$  and  $C_{FGD}$  are parasitic capacitances between the floating gate and the source and drain electrodes, and  $C_{FGi}$  are the capacitances between the floating gate and island  $i$ . The desired situation is for the floating gate potential to change as much as possible with respect to the external charge, so that a larger charge  $Q_i$  can be induced in the SET islands. From Equation 2.82, it follows that

$$\frac{\partial \varphi_{FG}}{\partial Q_e} = \frac{1}{C_{F\Sigma}}, \quad (2.83)$$

so that it is clear that as  $C_{F\Sigma}$  increases, the floating gate potential is less sensitive to the external charge. Expressing the external charge in terms of the potential  $Q_e = (\varphi_e - \varphi_{FG})C_e$ , Equation 2.81 leads to

$$\frac{\partial \varphi_{FG}}{\partial \varphi_e} = \frac{C_e}{C_{F\Sigma} + C_e}, \quad (2.84)$$

which shows that when  $C_{F\Sigma} \ll C_e$ , the potentials of the floating gate and the external charge are matched. This effect has been used to control the potential of a floating gate via a control gate[134], shown in Figure 2.18, where if the control gate capacitance is large compared to the others, the potential of the floating gate follows that of the control gate, and can therefore be used to bias the transistor as a normal gate. Optimal sensitivity of the floating gate potential to the external charge is determined by the coupling to the islands and the control gate. In a multi-island system, the size of the floating gate must increase if it is to couple to multiple islands, which also increases  $C_{F\Sigma}$ , reducing its efficacy. Considering that the magnitude of the effective charge of a biomolecule can be quite large under the right conditions (up to  $100e$  depending on the size and structure of the molecule[145]), higher values of  $C_{FGi}$  at the cost of a reduced  $\Delta\varphi_{FG}$  are likely tolerable.

If the capacitance values of the external charge, the floating gate, the control gate and the double layer capacitance are large, they should be included in the capacitance matrix of the system. The capacitance seen from an island  $i$  in the chain gains a contribution

$$\left( \frac{1}{C_{FGi}} + \frac{1}{C_e + C_{CG} + C_{DL}} \right)^{-1} \quad (2.85)$$

where  $C_{CG}$  is the control gate capacitance and  $C_{DL}$  is the double layer capacitance. Therefore the use of a large floating gate closely coupled to the islands will reduce their charging energy along the chain in a similar way as a gate electrode, the magnitude

of the reduction being mitigated by the values of  $C_e$ ,  $C_{CG}$  and  $C_{DL}$ . If the value is considerable, this contribution should therefore be added to the gate capacitance value in the estimation of the number of junctions  $N$  and the value of  $\alpha_J$ .

## 2.4 Mass fabrication of metallic SETs with ALD

The systems discussed in the previous section describe the envisaged routes for the implementation of the final system. The most formidable fabrication challenge is with mass producing room temperature operable SETs in a suitable configuration for use as a biosensor. The SET device can be laid out in two general ways if the fabrication processes permit it. In a vertical configuration, the islands are stacked on top of each other. In this configuration it is much easier to form a very thin island so that quantum confinement effects can be utilised. This method is more useful for studying the physics and electronic properties of a multi-island SET, such as spectroscopically probing the confined states with the gate electrode[74]. The difficulty is with the fabrication steps required to form vertical columns, and then the need to include a contact to the top of the column. Although the confinement effect can be used to increase the charging energy, a vertical configuration for a biosensing SET makes the fabrication process very challenging. Metallic SETs are more commonly fabricated using shadow-angle evaporation[60–62, 65, 67, 151], a technique that involves evaporating a metal through a suspended mask connected to the substrate. The first metal layer, usually Al, is deposited at a given angle, after which controlled *in-situ* oxidation of the metal film is performed to form the tunnel barriers. The second metal layer is then deposited at a different angle to create overlapping structures separated by the tunnel barrier thin film. Although this technique is very effective in an experimental environment, the mass fabrication of such devices for commercial use is not feasible using this method. The complexity of forming a suspended mask with densely packed features, the difficulty

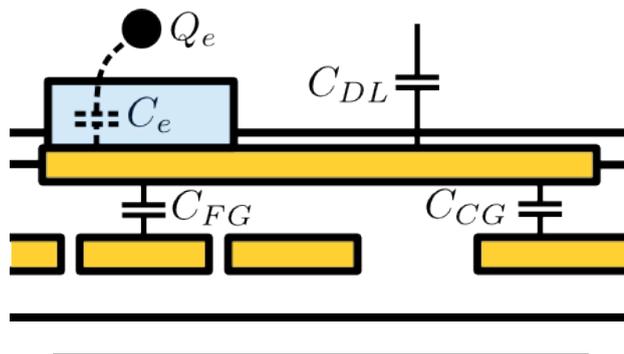


FIGURE 2.18: Use of a control gate to modify the floating gate potential like a normal gate, while still serving as a sensing surface. The capacitance caused by the electrical double layer at the interfaces of the materials is included.

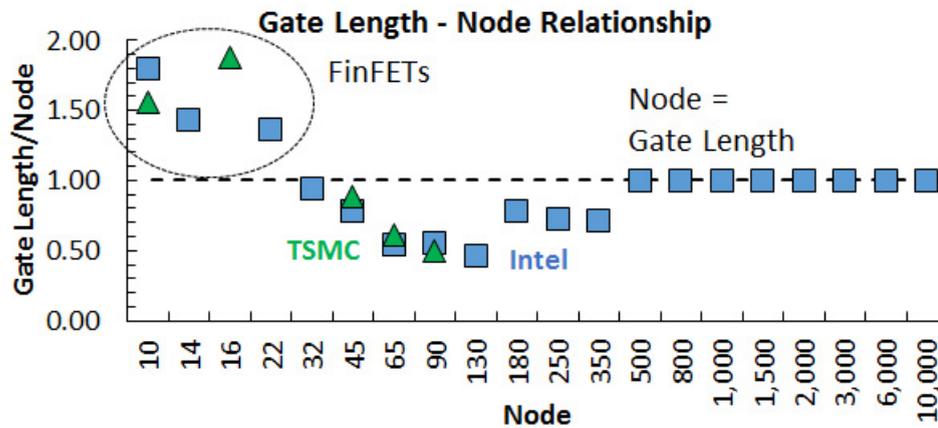


FIGURE 2.19: Variation of the gate length for each technology node, from manufacturers Intel and TSMC. The 10 nm node gate length is a projection. Image found here[155].

with achieving reproducible uniform thin films using oxidation, and the need for complex metal-evaporation equipment are all factors that limit the ease of fabrication of reproducible biosensor devices.

A potential route to the mass fabrication of metallic SETs is through the use of metal evaporation and ALD deposition of  $\text{Al}_2\text{O}_3$ . The experimental work presented in this thesis will show that ALD, a technique described in detail in Section 3.2.3, shows potential for producing reproducible thin films of  $\text{Al}_2\text{O}_3$  over a large surface area with desirable properties for a SET operating at a high temperature. The highly uniform thickness and uniform insulating properties are promising for creating a high density of small capacitors and tunnel barriers with minimal spread in the tunneling and dielectric properties. ALD was integrated into the CMOS manufacturing process flow for the 45 nm Intel<sup>®</sup> technology node[152], to reduce the gate leakage current in Si FETs by a few orders of magnitude by increasing the thickness of the gate capacitor and using high permittivity of  $\text{Al}_2\text{O}_3$ . It has also been used more recently as a gate oxide in  $\text{MoS}_2$  based FETs at sub-nanometer thicknesses, to considerably reduce gate leakage[153]. As discussed in Chapter 1, finFET CMOS technology has shown some potential for room temperature SET operation, but the Si quantum dot properties are irregular and highly non-reproducible. Metallic interconnect technology, which connects the FET networks to make logic circuits, is driven by similar scaling requirements to the FET dimensions, and therefore much research is focused on achieving the smallest metallic linewidth. It is expected that 10 nm linewidths will become possible within a decade according to the 2015 International Technology Roadmap for Semiconductors (ITRS) for interconnects[154], so the patterning of ultra-small metallic electrodes using photolithographic methods is an important part of industrial-scale manufacturing.

The fabrication of the islands can be approached in two ways, using top-down or bottom-up methods. Bottom-up methods involve patterning the metals directly into the required structure, such as using the lift-off method (discussed in Section 3.2.1). In a typical CMOS process, top-down methods are usually used, where in this case the metal and a hardmask is deposited on the entire substrate first, after which the photoresist is deposited and patterned, allowing the metal patterns to be formed by etching techniques. There are two general steps in the CMOS processing line in which metals can be used, the front-end-of-line (FEOL), which is the transistor fabrication step including the gates, and the back-end-of-line (BEOL), which involves fabrication of contacts to the FET source and drain, and subsequently the interconnects. Metal gates (proprietary materials) were used in Intel 45 nm node technology for FETs, and very small gate lengths on the order of 35 nm with a pitch of roughly 160 nm in SRAM cells could be achieved[152]. The 10 nm and lower technology nodes are currently still in development, however various semiconductor manufacturers have revealed their estimations of the gate lengths, some of which are summarised in Figure 2.19, including the past technology nodes. As can be seen, the gate lengths increase relative to the node size as of the most recent trends.

Figure 2.20 (a) and (b) illustrate an example lateral SET system that makes use of the small half-pitch achievable in the CMOS process. Making use of the minimum half-pitch, the gate length is on the order of half its value, so that overlaid metal lines can produce small capacitors defined mostly by the contact area of the overlap. Use of ALD provides the junction capacitance and tunnel dielectric. The size of the floating gate is determined by the dimensions of the channel. By densely packing the lines that form a channel, the floating gate can be scaled down, reducing the capacitance with the multi-island SET. The packing density of the devices determines how much of the exposed surface area consists of sensing surfaces. Since each island has a planar structure, and an exposed region above and below, the capacitance between electrodes placed in these regions could be quite large. The floating gate capacitance can be tuned by changing the thickness of the dielectric between the islands and the floating gate.

The use of ALD for the precise control of the tunnel barrier thickness has some potential drawbacks depending on the substrate materials used. The standard ALD- $\text{Al}_2\text{O}_3$  recipe used to make high quality tunnel barriers makes use of either water or oxygen plasma as one of the gas precursors. Therefore most materials will inevitably oxidise to some extent during the ALD process, which can cause significant uncertainty in the thickness of the tunnel barrier. Multi-layered insulator tunnel barriers show properties that are generally undesirable for SETs, such as an enhanced asymmetry in the I-V characteristics, a diffuse interface, and a change in the effective permittivity of the whole film. To completely avoid such oxidation in the ALD process, highly inert materials such as Au or Pt should be used. This considerably restricts the range of materials that can be used to form SETs using this particular process. The implications

of using the ALD- $\text{Al}_2\text{O}_3$  process for precise tunnel barrier fabrication is discussed in further detail in Chapter 5.

According to the 2015 ITRS for lithography[156], a number of technologies are in various stages of development for the 10 nm node technology and beyond. One of the most promising technologies is extreme ultra-violet (EUV) lithography which uses a 13.5 nm wavelength electromagnetic radiation source, compared to the current state of the art deep ultra-violet (DUV) lithography that uses a 193 nm source. Despite the numerous challenges facing the use of EUV for mass fabrication, the performance of resists that are sensitive to 13.5 nm radiation were tested using a synchrotron source, and it was shown that half pitch value of less than 20 nm could be achieved for line-space patterns, which represents a linewidth of less than 10 nm[157] (shown in Figure 2.21). There is therefore some potential for mass producing nanoscale metallic electrodes, that may be used for forming SET devices.

## 2.5 Conclusion

It was shown in the first section that multi-island metallic SET systems may be used to significantly enhance the Coulomb blockade, at the cost of a tradeoff with coupling to gate electrodes. Using the orthodox theory of Coulomb blockade, the important parameters of the system were identified as the capacitances and electron tunneling properties of the junctions, which need to be determined experimentally. The important sources of noise in an SET system were discussed, with particular attention to the effect on the charge sensitivity of the device. Some of the most important challenges

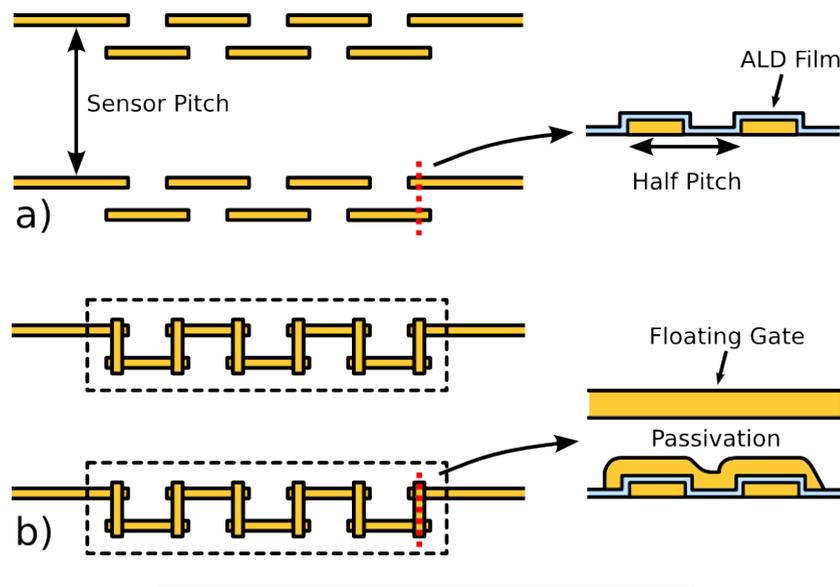


FIGURE 2.20: (a) First layers of the SET fabrication, and thin ALD coating for tunnel barriers. (b) Final device layers, showing the position of the floating gate on top of the channel.

surrounding the integration of an SET as a biosensor were discussed, including the integration of microfluidics and the use of SAMs. The geometrical requirements and the arrangement of the system was discussed in terms of the operating temperature and the charge sensitivity. Finally, the use of mass fabrication techniques for fabrication of metallic SETs was briefly discussed. It was shown that at the scale of lithography limits

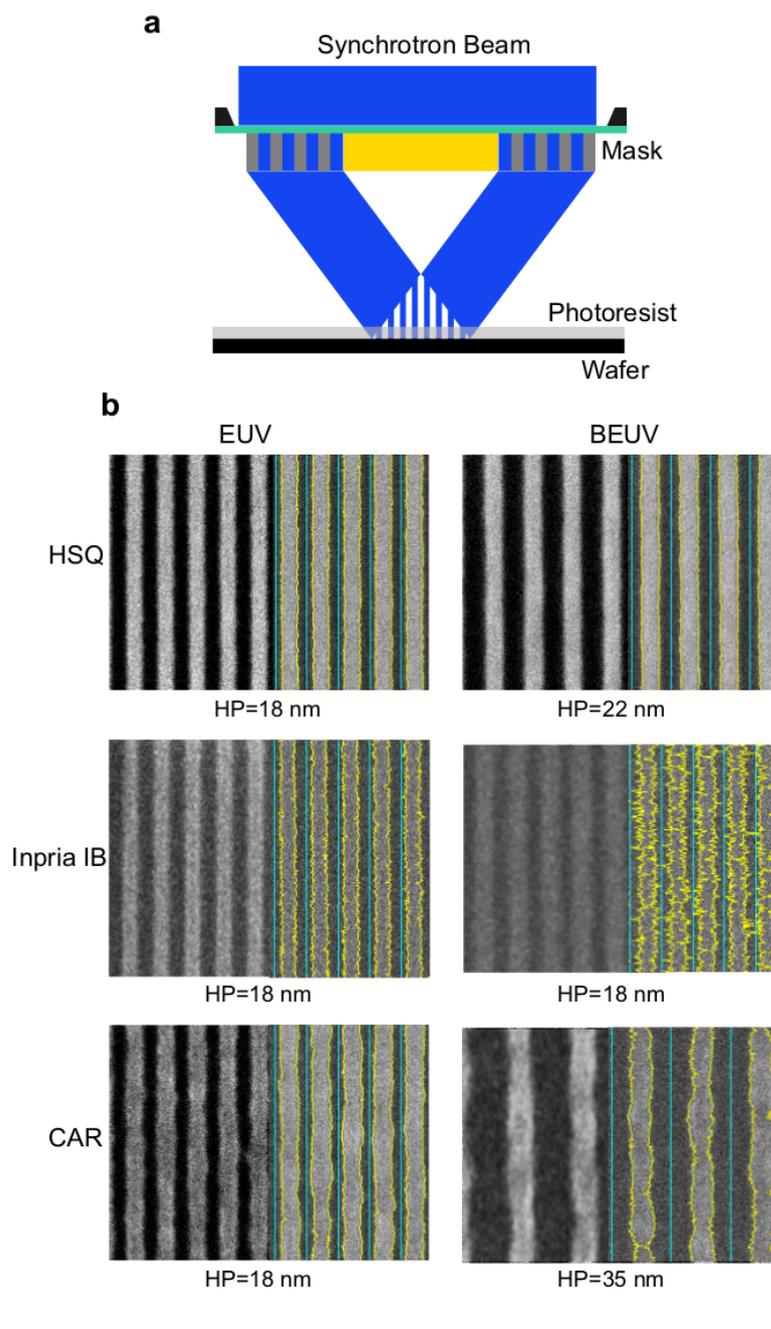


FIGURE 2.21: (a) Basic schematic of the EUV process used to pattern the resist. In this case the EUV source is a synchrotron beam, which allows precise control the electromagnetic radiation. (b) Comparison of linewidths achieved using different resists (HSQ, Inpria IB and CAR), and different EUV techniques (BEUV result is not important here). Adapted from Mojarad *et al.* [157].

( $\approx 10$  nm), the thermal energy at room temperature could be significantly exceeded with long chains, on the order of 10 islands, and that the charge sensitivity should still be sub-electronic in this range. The use of ALD and metals to mass fabricate SETs was discussed in terms of state of the art, and anticipated CMOS fabrication processes.



# Experimental Details

---

## 3.1 Introduction

This chapter details the experimental techniques used during the project. The fabrication equipment used to fabricate metal-insulator-metal (MIM) capacitors of different surface area and insulator thickness is detailed first. Typical process parameters are given here if they were varied little or not at all throughout the fabrication work, otherwise they are detailed in the relevant chapters. The tools used to characterise the fabrication steps and materials of the structure are described next. The measurement equipment used to electronically characterise the devices is described last.

## 3.2 Fabrication techniques and apparatus

### 3.2.1 UV Photolithography

To create micrometer-scale patterned structures, photolithographic techniques can be used to create a patterned sacrificial layer, which acts as a stencil for a fabrication process, such as metal evaporation, or wet/dry etching, allowing the desired patterns to be projected onto the underlying substrate. The sacrificial layer is formed from a photoresist, a curable resin that is designed to be sensitive to light in a specific wavelength range, such that a controlled dose of such radiation causes a chemical reaction that determines the photoresist's sensitivity to a developer solution, used to form the stencil pattern. A positive resist becomes soluble in its associated developer after exposure to a sufficient radiation dose, the reverse type of resist is called a negative resist, where exposure leads to insolubility in the developer. To create patterns, a lithography mask is required to protect the desired regions from the radiation. The mask is formed from a high quality plate of glass which is optically transparent to the light source, on which the design patterns are created using electron-beam lithography and a special set of metal coatings designed to reduce optical aberrations. The mask and the substrate are then loaded into a mask-aligner, which provides the radiation source and the controls to set a precise exposure time. When the mask is properly aligned with the substrate, the mask is brought into contact with the resist on the substrate surface, and the controlled dose can then be delivered. This method is called contact printing, since the mask is in direct contact with the resist surface. After development of the patterns in the developer solution and any additional resist processing post-development (such as

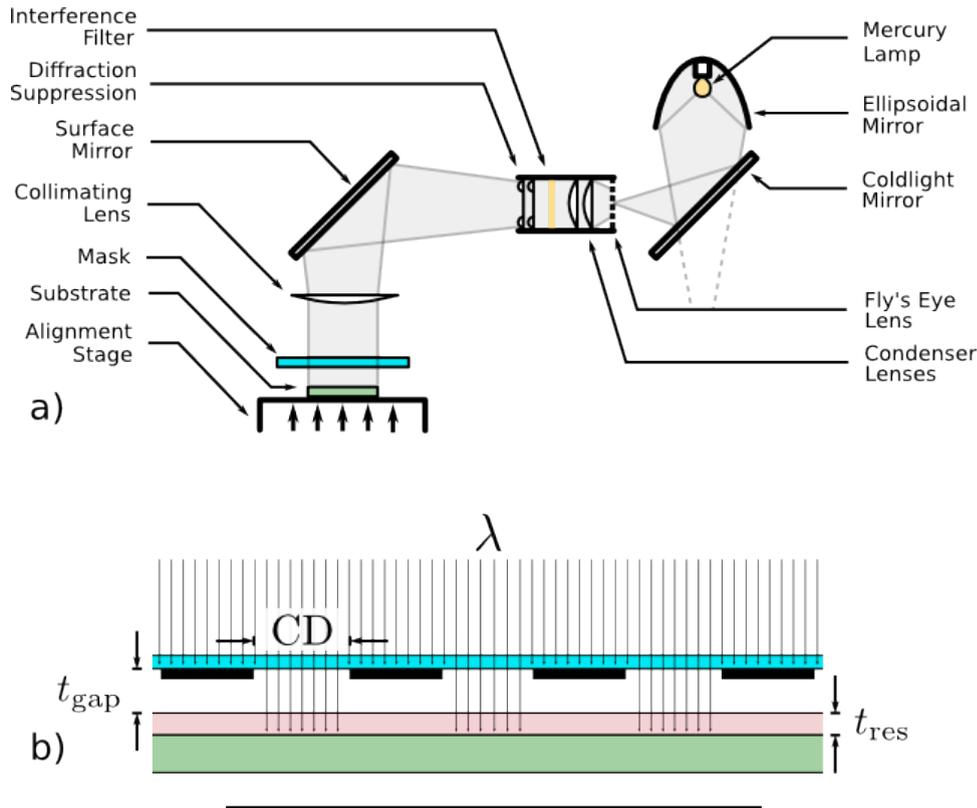


FIGURE 3.1: (a) Schematic of the Karl Suss MJB3 system used for aligning a mask with a substrate and exposing the substrate to UV radiation. (b) Geometry of the mask-substrate system. The smallest critical dimensions CD achievable are related to the thickness of the gap between the mask and the resist  $t_{\text{gap}}$ , the thickness of the resist  $t_{\text{res}}$  and the incident radiation wavelength.

extra baking steps), the substrate is ready for the next processing step to be used as a sacrificial layer. After the deposition or etch step, the remaining resist is stripped in a suitable solution, exposing the protected regions of the substrate.

Figure 3.1 (a) shows a schematic of the Karl Suss MJB3 system used in this work[158]. The UV source is a mercury lamp that emits a broad range of wavelengths, those of interest between 310 and 436 nm. The ellipsoidal mirror focuses the light emitted from all angles of the lamp to maximise the intensity. The coldlight mirror is specially designed to transmit infra-red (IR) radiation and reflect UV radiation. The IR radiation is absorbed into a heat sink, and the UV radiation is focused into a system of lenses for conditioning. The fly's eye lens disperses the light evenly into a series of condenser lenses that collimate the light into a uniform beam, which are adjusted spatially for optimal intensity and uniformity. An optional interference filter can be placed to filter unwanted spectral lines, however none were used in this work, so that the maximal intensity from the mercury lamp is used. The beam is then fed through a series of 12 lenses designed to reduce diffraction effects. The surface mirror then reflects all light coming out of the lens system into a vertical beam, and finally through a collimating lens. The mask is

fixed patterns-down to a metallic plate using a vacuum seal around an aperture 5 cm in diameter. The substrate is placed on a chuck connected to the alignment stage, so that the substrate can be aligned with the patterns on the mask, using a microscope aligned with the aperture. When the substrate is aligned, the chuck is pushed upwards into contact with the mask to minimise the gap between the substrate and the mask patterns. The exposure time is set using a dial and initiated using a button, after which the optical system is moved into alignment with the aperture and a shutter is opened for the specified time.

The resolution and precision of the photolithography process is determined by many factors. Fundamentally, as the wavelength of light approaches the scale of the clear patterns on the mask, diffraction begins to occur, leading to a diffuse image of the pattern projected onto the resist. Figure 3.1 (b) shows a diagram of the patterned mask and the resist-coated substrate. In contact printing, the gap  $t_{\text{gap}}$  is reduced as much as possible to make use of the maximum resolution available by the wavelength of the incident light, and therefore the main limitation on the resolution is the diffraction due to opaque edges on the mask. The maximum resolution can be expressed as[159]

$$2\text{CD} = 3\sqrt{\lambda \left( t_{\text{gap}} + \frac{t_{\text{res}}}{2} \right)} \quad (3.1)$$

if the patterns are considered to be a grating with pitch  $2\text{CD}$ . Therefore with a wavelength  $\lambda = 365$  nm and a resist thickness  $t_{\text{res}} = 1$   $\mu\text{m}$ , the best achievable resolution is  $\text{CD} \approx 640$  nm. If a poor contact is made with the resist, a finite gap  $t_{\text{gap}}$  is introduced which decreases the resolution. For instance a gap of 10  $\mu\text{m}$  worsens the resolution to approximately 3  $\mu\text{m}$ . Therefore a good contact is essential for small features. The drawback of contact printing is the contamination risk and deformation of the resist under contact, which causes reliability issues. The mask must therefore be thoroughly cleaned before use. Depending on the severity, an acid clean may be required using piranha (1:H<sub>2</sub>O<sub>2</sub> (30%) to 3:H<sub>2</sub>SO<sub>4</sub> (98%)) to remove tough resist residues, or a simple rinse in acetone followed by isopropyl alcohol (IPA), and plenty of N<sub>2</sub> blow-drying, is sufficient directly before and after use. The mask used in this work was designed for positive resists.

Positive resists are generally formed of a photoactive compound (PAC), a novolak resin, and a solvent. The resin is diluted in the solvent such that when it is cured, the resin remains and provides a strong and stable structure. In this work, the substrates used to make devices were placed in a resist spinner. The resist spinner is essentially a chuck connected to an axle with a vacuum line that holds the substrate in place on the chuck. The axle is connected to a motor which spins the substrate on a precise vertical axis, with the required velocity profile. The resist in its dilute form is administered to

the substrate surface using a small pipette, slowly enough to avoid the formation of small bubbles, after which the programmed velocity profile can be initiated. The viscosity of the resist, determined by the solvent concentration, limits the amount of resist ejected from the substrate surface at a given velocity, and therefore highly uniform thin layers of resist can be obtained. The spin speed curves provided in the resist datasheets are used to determine the speed required for the desired resist thickness. The resists used here are then baked for a specific time at a given temperature, after which the substrate is ready for patterning. The baking or curing step evaporates the solvent, and the PAC is trapped in the resin structure as a result. The most important resist used in this work was Shipley S1813[160], a positive resist that uses tetramethylammonium hydroxide (TMAH) (MF-319, approximately 2% TMAH in water) as a developer. S1813 is optimised for sensitivity to the g-line (435 nm) of a mercury lamp, achieved using a proprietary diazo based compound, such as diazonaphthoquinone (DNQ). DNQ is a compound that is insoluble in water and TMAH before exposure to UV radiation, and upon sufficient UV absorption, becomes soluble[161]. The exposure dose delivered to the resist is a function of the UV power, time, and the geometry of the mask patterns. The UV power seen after the collimating lens stage in the mask aligner can be measured using a tool which yields a power on the order of  $20 \text{ mW.cm}^{-2}$ , depending on the age of the mercury lamp. The recommended exposure dose for S1813 is  $150 \text{ mJ.cm}^{-2}$ , such that approximately 7 s exposure times are sufficient to a first approximation. The mask and the substrate influence the exposure time however. If the mask is highly opaque, i.e. with small ( $< 10 \mu\text{m}$ ) and disparate clear features, a higher exposure time is generally needed due to the low optical transmittance of the mask. As a result, a mask pattern that contains both large and small features will show an inhomogeneity in the image tolerance printed to the resist after development. If the substrates are UV transparent, such as those used in this work, the mask aligner chuck can scatter some of the radiation back through the substrate into the resist, increasing the UV absorption, and thus the exposure time must be limited to avoid over exposure. The trade-off between these two factors results in an optimal exposure time, derived experimentally using the available mask (described in Chapter 4), between 4 and 10 seconds, depending on the fabrication step. Modern masks use anti-reflective coatings to minimise reflections between the mask and the substrate/chuck.

For lift-off lithography, where a material is deposited on the patterned resist and subsequently removed to leave the deposited pattern, a single resist layer is generally insufficient to provide reliable image transfer. The reason is the deposited material usually ends up on the resist walls due to the non-ideality of the source of the material being deposited, or simply due to the nature of the deposition (such as ALD). The solution is to use another resist deposited before the S1813 to act as a buffer layer. If the buffer layer is more soluble in the developer than S1813, it will dissolve at a

higher rate, and therefore produce an undercut that leaves the edges of the S1813 layer hanging. This way the deposition source is shadowed by the overhanging S1813, which leaves clean edges on the patterns as long as the buffer layer is thick enough. The usual guideline is that the buffer layer should be at least 3 times thicker than the deposited material for a reliable undercut, and the method is therefore called the bilayer lift-off method. Two different resist materials were used as buffers in this work, depending on the fabrication step. LOR-3A is a resist widely used with S1813 to produce a reliable undercut[162]. It is baked at a higher temperature than S1813 and is developed in the same developer, so that generally only one development step is needed. The thickness is determined by the spin speed, and the dissolution rate by the bake temperature and duration, as well as the TMAH concentration of the developer. The second resist used as a buffer layer was PMMA[163], which is a commonly used electron-beam lithography (EBL) resist. PMMA was required mainly to protect the underlying aluminium oxide film from dissolving in the TMAH developer, as a different developer is required. The PMMA-S1813 bilayer process is complicated for a number of reasons. PMMA is sensitive to deep-UV (DUV) radiation at 248 nm, and for lack of a dedicated DUV source for lithography, the DUV source of an ozone cleaning unit was used, which resulted in the need for long exposure times on the order of 20 minutes. Additionally, the interface between PMMA and S1813 becomes significantly diffuse, which reduces the sensitivity of the diffuse layer to both the UV and DUV sources. The process is not reliable for producing a good undercut for thick deposition layers, however the PMMA is readily removed from the underlying surface in a suitable developer, leaving a clean surface free of residue. PMMA comes in a series of different molecular weights and solvents. Higher molecular weight PMMA leads to thicker deposited layers at a given spin speed. In this work thick layers were required so the less diluted and higher molecular weight variants were used (495K and 950K series), and the anisole solvent series was used, rather than chlorobenzene which poses greater safety hazards. A range of developers were tested, including varying concentrations of MIBK in IPA, and the 7:3 IPA:H<sub>2</sub>O developer.

Table 3.1 summarises the equipment used for photolithography processes. The wet-bench is supplied with all the required solvents for cleaning and resist development, and

<b>Equipment</b>	<b>Purpose</b>
Wetbench (inc. Solvents)	Cleaning/Processing
Hotplates	Resist baking/Surface treatment
MJB3	Resist patterning
UVO Cleaner	PMMA exposure/Cleaning
Ultrasonic Agitator	Cleaning
O <sub>2</sub> Asher	Cleaning
Spinner 1	Primer/S1813
Spinner 2	Primer/LOR-3A
Spinner 3	Primer/PMMA

TABLE 3.1: List of equipment used for photolithography processes.

a range of glassware. Resists are stored in a separate fridge to avoid contact with light and to prolong shelf-life. Separate spinners are used for each resist, to reduce risk of cross contamination. The ultrasonic agitator, UVO cleaner, and O<sub>2</sub>-plasma asher can be used to clean substrates at various stages of fabrication. The substrates were handled with carbon fibre tipped tweezers at all times during photolithography processing. A N<sub>2</sub> gun is used to dry substrates after cleaning in solvents. The substrates used in this work required a surface-priming step before coating with S1813 and LOR-3A. The primer solution used was hexamethyldisilazane (HMDS), which is applied after heating the substrate to remove unwanted solvent residue and water, which makes the SiO<sub>2</sub> surface hydrophobic. The adhesion of the resist to the substrate surface is then improved, which is important for contact lithography and clear substrates.

### 3.2.2 Metallic thin film deposition

Thin metallic films were deposited using electron-beam evaporation (EBE), to create metallic electrodes. The system used is a Leybold Univex 350, which uses a curved-beam electron source. Figure 3.2 shows a schematic of this system. Substrates are placed on a large chuck, which is placed inside a large chamber that is brought under vacuum by a roughing pump and then a turbo pump for a higher vacuum. An electron beam is created using a filament, usually composed of tungsten, that is heated to boil electrons off the surface, and an anode plate held at a high voltage of approximately 7.5 kV accelerates electrons. A transverse magnetic (TM) field is applied that deflects the beam emitted from the anode towards the target. Deflectors control the beam  $x$  and  $y$  position by controlling the current through inductive coils that induce the desired deflecting magnetic field. The metal target is placed in a liner, designed to thermally isolate the target from the crucible as much as possible. The crucible is cooled by a cold water or liquid nitrogen source. The shutter when closed shades the chuck away from the metal vapour released from the metal target. When the shutter is opened the metal vapour is exposed to the substrate and material is deposited on the substrates. The crystal monitor is used to measure the thickness of the film during the deposition, by measure of the change in resonant frequency of a quartz crystal microbalance due to the change in mass, and knowledge of the mass density and acoustic impedance of the target material. The crystal monitor must be at the same radial distance from the source as the substrates, to ensure an accurate reading. The deposition rate depends on the distance from the source, since the vapour is emitted at broad angles.

The system used has a rotary crucible that supports multiple metal targets. A system module provides an interface to create recipes that allow sequential deposition of metallic films without breaking vacuum. The created recipe sets up the crystal monitor for the desired film, and sets up an optional automatic electron-beam control and shutter

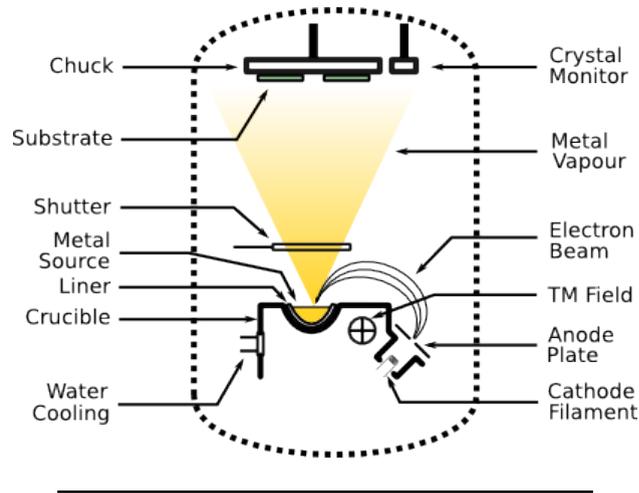


FIGURE 3.2: Schematic of the electron beam evaporation system used in this work.

operation. The electron beam, when focused and incident with the target, will effectively heat the target. Depending on the target material, a certain minimum amount of power is required to bring the material to its melting, or sublimation point. The power is increased by increasing the current emitted from the filament, which leads to a higher electron flux to the target. The characteristics of the deposition are determined by the phase transition properties, the path between the target and the substrates, and the characteristics of the condensation of the vapour onto the substrate. The vapour pressure, the pressure exerted by the vapour on the condensed phase (liquid or solid) in equilibrium, as well as the properties of the target, determine the evaporation rate. The vapour pressure is increased by increasing the temperature, which leads to higher evaporation rates. The air pressure in the vacuum chamber must be minimised so that only the target vapour pressure limits the rate at which atoms transition to a vapour state. A high vacuum then allows a long mean-free-path for both the electron beam and the evaporated beam, and reduces the organic impurity content of the deposited material. Vapour pressures are usually determined experimentally, and they may differ strongly between materials[164]. The materials used during this work were Au (gold) for the electrode layers and Cr (chromium) for the adhesion of the electrode layers to the substrate. Au requires a significant electron beam power to bring the vapour pressure to a high value, partly due to the high heat capacity, which indicates a high energy is required to heat Au. The beam current required to observe a practical evaporation rate ( $> 0.1 \text{ \AA}\cdot\text{s}^{-1}$ ) for Au is on the order of 0.25 A at 7.5 kV accelerating voltage and vacuum pressure of approximately  $10^{-6}$  bar, or 750  $\mu\text{Torr}$ . For Cr, the vapour pressure exceeds the equilibrium pressure before melting, and thus sublimates rather than evaporates. The equilibrium pressure refers to the vapour pressure at which the condensed and vapour phase are in equilibrium, i.e. when the transfer of substance between the two phases are equal across a boundary (the surface of the target). A significantly smaller power is

required for Cr compared to Au. A beam current on the order of 20 mA is sufficient to produce practical deposition rates on the order of  $0.5 \text{ \AA}\cdot\text{s}^{-1}$  in the available system.

The morphology and properties of the deposited films depends on the growth characteristics during condensation of the vapour onto the substrate. The growth can be characterised in four steps as determined by electron microscopy observations of the substrate at various stages of growth[165]. The first impinging nuclei diffuse on the surface and nucleate to form small grains (on the order of 2.5 nm), such that growth occurs both in parallel and perpendicular to the surface. As the deposition continues the nucleation centres coalesce into larger patches on the surface and the original substrate becomes more covered. As the substrate becomes covered, small channels and holes remain. Beyond a certain thickness the film then becomes continuous, although metallic thin films show a grain size that is dependent on film thickness. The grain structure depends on the metals, where for Au, columnar grains are formed that impinge on each other during growth. In order to create good quality electrodes, the metal film thickness must be continuous, and as homogeneous as possible. Films at least 20 nm thick in total were used, which was found to be sufficient to achieve a dense and continuous film. Such thin films were used firstly because the lift off process is more reliable, and secondly because the roughness of such evaporated thin films increases with thickness and deposition rate[166, 167], amongst other factors. After the deposition and lift-off process, the thickness and surface roughness of the films can be measured using either an alpha-step or an atomic-force microscope (AFM). The AFM system is described in a separate section, however the alpha-step is a similar system that is much less accurate for fine measurements such as roughness or very thin layers ( $< 20 \text{ nm}$ ), but is much simpler and quicker to use. The alpha-step drags a diamond tip connected to a cantilever across the substrate surface, and deflections of the cantilever are measured to produce a trace of the surface profile. Film thickness measurements showed that at most a 10% error in thickness was present between uses and maintenance of the system. As the metal film thickness was not so critical in this work, no tooling factor was needed to achieve precise film thicknesses with less than 10% error.

### 3.2.3 Atomic layer deposition

ALD can be defined as a fabrication tool that makes use of surface-limited, irreversible solid-gas reactions to provide uniform and conformal thin-films of many materials onto many substrates. The material deposited is a result of the surface chemistry between the substrate and usually two precursor gases that are supplied sequentially into a reactor via high speed switching valves[98, 168]. The primary driving force of the development of ALD is the need for higher-k dielectrics in the CMOS technology nodes

as conventional SiO<sub>2</sub> gate oxide scaling leads to excessive gate leakage current in MOS-FETs, leading to higher power dissipation and unreliable operation. In the sidelines of this research goal in CMOS technology, many groups are using this technology in areas that require reliable and well defined tunnel barriers, such as single electron memories based on floating gates[63, 169], MIM tunnel diodes[170–174] and graphene spintronic devices[175]. Only the use of the trimethylaluminium (TMA) precursor for depositing Al<sub>2</sub>O<sub>3</sub> will be discussed, which is by far the most studied and understood ALD process, as well as the most ‘ideal’ in terms of the physics and chemistry of the precursors. The growth process in ALD is intended entirely as a chemical process as opposed to a physical process such as EBE and sputtering. It is the properties of the specific reactions between precursors that distinguish it from other chemical-vapour deposition (CVD) process such as Molecular-Beam Epitaxy (MBE) where the growth is limited by the MBE system itself through the use of shutters. To create a successful ALD chemistry, it is necessary that the surface reaction be irreversible so that the deposited precursor layer cannot react back into volatile compounds, and it must be self-terminating or saturating, so that only one precursor layer is deposited. Figure 3.3 illustrates the kinetics of adsorption where inset (a) depicts the ideal conditions for an ALD process. In the TMA/H<sub>2</sub>O or O<sub>2</sub>-plasma process, the adsorption is of the chemisorption type where the interaction between the reactants is strong, meaning that a given surface of reaction sites cannot rearrange itself to acquire more packing efficiency. This leads to the distinction between a physisorption monolayer and a chemisorption monolayer, where the latter considers the density and structure of reaction sites on the surface as well as the structure of the binding molecule, as is discussed by Puurunen *et al.*[98].

An overview of the ALD system (Fiji model by Cambridge Nanotech) used for the experimental work is shown in Figure 3.4. The substrate is delivered to the reaction chamber through a load-lock system, which permits the reaction chamber to remain under vacuum. Heaters in the reaction chamber control the temperature of the chuck, which the substrate is placed on. Further heaters control the temperature of the precursor delivery lines (denoted in green). The delivery of the precursor gases is controlled by high-speed switching valves that ensure a precise precursor dose is delivered to the reaction chamber. A carrier gas (light blue arrows) maintained at a constant flow rate is used to continually purge the reaction chamber, and is used to clear the headspace of remaining precursor gas between the manually operated precursor inlet valve and the switching valve. Depending on the adsorption kinetics, the high speed switching valve is opened for long enough for saturation of the surface to occur, after which the system is purged for a certain amount of time to ensure no precursor remains. This is an ALD half-cycle for a two-precursor deposition process, such as for Al<sub>2</sub>O<sub>3</sub> and many other materials[176]. The second half-cycle then depends on the kinetics of the second precursor, in this case water vapour or O<sub>2</sub>-plasma. Therefore, every cycle provides a

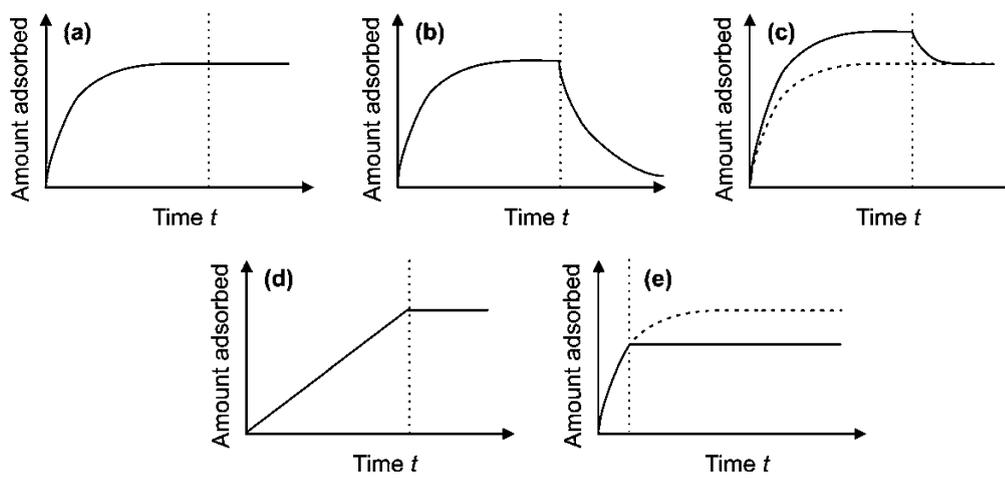


FIGURE 3.3: Adsorption kinetics, (a) irreversible saturating (self-terminating), (b) reversible saturating, (c) combined irreversible and reversible saturating, (d) irreversible nonsaturating (deposition), and (e) irreversible saturating adsorption prevented from saturating. The vertical dashed line marks the end of an injection of precursor into the reaction chamber, which is cleared of any remaining precursor by the purge gas.

Source: [98]

certain amount of growth of the target material, which is described by the growth-per-cycle (GPC), usually quoted in units of  $\text{\AA}\cdot\text{cycle}^{-1}$ . The user interface to the system is a LabView program that provides both manual and programmatic control over all aspects of the system. Control over the execution of the cycle loop and the individual parameters are controlled in an editable list with a simple programmatic structure providing loops and conditional statements.

The chemistry of the TMA/ $\text{H}_2\text{O}$  process for the deposition of  $\text{Al}_2\text{O}_3$  is considered chemically ideal for a number of reasons. The TMA molecule is thermally stable up to approximately  $300\text{ }^\circ\text{C}$ , which allows a wide range of temperatures, and it is highly reactive to water vapour and oxygen containing compounds, allowing short exposure times. Additionally, the products of the reaction in general do not interfere with the growth process. The stoichiometric equation of the reaction between the two precursors is



where the underlying mechanism is the oxidation of the aluminium atom through the reaction of the hydrogen and methyl groups leading to the release of methane. The reaction of TMA with a surface occurs through a number of chemisorption mechanisms depending on the surface chemical composition of the substrate. On a substrate that presents a high surface density of OH groups, TMA reacts predominantly through ligand exchange and dissociation of the  $\text{CH}_3$  groups. TMA molecules may also react associatively to surface groups other than OH, such as the native substrate terminations, in

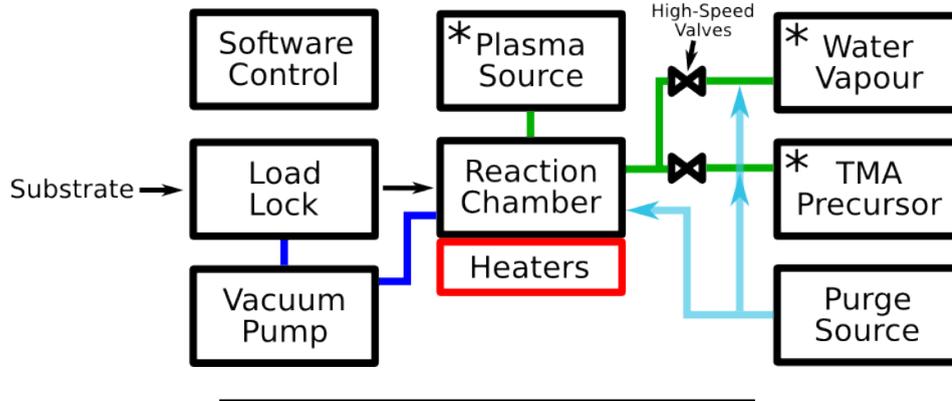
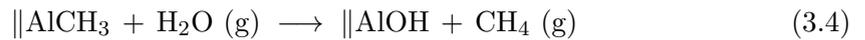


FIGURE 3.4: Overview of the ALD system used for the fabrication of the devices presented in this thesis. The precursors are denoted by the asterisks.

which case no reactants are formed. In general, the TMA half-cycle is intended to terminate the surface with  $\text{CH}_3$  groups, so that in the  $\text{H}_2\text{O}$  half-cycle, the  $\text{CH}_3$  surface reacts and is turned back into an OH terminated surface. In ligand exchange, the TMA half-reaction equation for single ligand exchange is



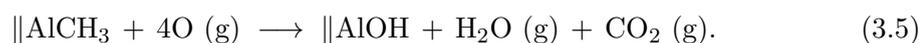
where  $\|$  represents any supports for the binding site. This reaction can continue through the same process until all the methyl ligands are turned into methane, such that an aluminium atom is fully adsorbed and forms part of the solid oxide, at which point it can be reacted with  $\text{H}_2\text{O}$  to form new surface OH groups. The  $\text{H}_2\text{O}$  half-reaction equation is



where the same ligand exchange reaction has occurred to switch the terminating groups back to OH, ready for the next TMA exposure. These equations describe the process in general and are useful as a starting point for understanding the growth mechanisms. In fact many types of surface species have been identified by *in-situ* infrared spectroscopy where in the TMA half-reaction, the TMA monomer can react with two surface OH groups simultaneously and it can react dissociatively with coordinatively unsaturated Al-O pairs. Similarly for the  $\text{H}_2\text{O}$  half-reaction, ligand-exchange and dissociation reactions occur to produce OH groups. The oxidising half-reaction involving water vapour is usually always made more effective with the application of heat, however the GPC decreases at low temperatures as dehydroxylation of the surface occurs during the long purge times required for water vapour, reducing the surface density of reaction sites[98, 168].

In the TMA/ $\text{O}_2$ -plasma process, often called plasma-enhanced ALD (PEALD), oxygen free radicals are used in the second half-cycle rather than water molecules for their

much higher reactivity. In the Fiji system, the plasma type is a remote plasma, where the plasma is kept some distance away from the plasma boundary, providing some protection from high energy electrons, ions and photons that can cause physical damage to the substrate, or alter the growth. The plasma source is a cylindrical inductively-coupled plasma (ICP) that feeds directly into the reaction chamber from above. In a research setting the remote plasma method is the most versatile, as it allows the substrate and plasma parameters to be controlled independently. Since the plasma is far away from the substrate, the substrate plays no important part in species generation, allowing the composition of the plasma to be controlled, as well as the temperature of the substrate surface[177]. The proposed mechanism for the surface chemistry of the O<sub>2</sub> plasma is a combustion-like process where in the plasma half-cycle, H<sub>2</sub>O, CO<sub>2</sub> and CO are released as products[178]. The dominant surface groups created in the remote plasma process are confirmed to be CH<sub>3</sub> groups for the TMA half-cycle, and OH groups for the plasma half-cycle in the temperature range 25°C to 150°C[179]. The suggested O<sub>2</sub>-plasma half-reaction equation in the simplest case can then be written as



The main motivations for the use of PEALD over ALD is that it offers more flexibility over the choice of substrate, as the deposition can occur at much lower temperatures. PEALD has been used for example to coat sensitive organic materials, such as organic light emitting diodes (OLEDs)[180]. It has also been shown that there are a number of advantages in the material and electronic properties, such as a higher breakdown field, higher permittivity and more reproducible electronic properties[181–183], compared to the water vapour process. Furthermore, the process is faster overall since the plasma is purged more easily than water vapour and thus cycle times are optimised with short purge times. This in turn leads to a higher growth rate, due to the limited dehydroxylation within the purge time. As the surface OH concentration is temperature dependent and the O<sub>2</sub>-plasma highly reactive independently of temperature, higher growth rates are achieved at lower temperatures[178]. However due to the nature of the plasma, and the geometry of the reactor, conformality and uniformity are reportedly not as consistent over large areas covering a 300 mm Si substrate[181, 184, 185]. This can be expected as the reaction with radicals is no longer saturating as with water vapour, rather, since the radicals come from a source that is contained and not spread equally over the reactor volume, the reactants must follow a directional path to reach the surface. The fact that the reactions with radicals are not thermally induced provides better control over the surface chemistry and the species supplied to the substrate. Overall the use of plasma is a much more versatile methodology since different precursors can be used, additional gases can be introduced to dope the film and the plasma can be used to clean

Parameter	Value
Process temperature	200 °C
Process pressure	50 mTorr
Carrier gas and flow	Ar at 100 sccm
ICP power	300 W
O <sub>2</sub> flow	50 sccm
TMA pulse	60 ms
TMA purge	10 s
Plasma exposure	20 s
O <sub>2</sub> purge	5 s
GPC	1.14 Å.cycle <sup>-1</sup>

TABLE 3.2: Fiji F200 ALD system parameters used for the fabrication of MIM devices.

a surface before starting deposition without breaking vacuum. Further benefits of the PEALD-Al<sub>2</sub>O<sub>3</sub> process over the conventional process include a higher film density and lower impurity content[177], and the impurity content is reported to reduce with a longer plasma exposure time[178, 179].

Table 3.2 details the PEALD system parameters used during the fabrication of the devices presented in Chapter 4. The parameters were not changed throughout the experimental work carried out, as the dielectric properties of interest were found to be within the narrow range of widely reported values. The highest permitted temperature of 200 °C was selected as it is reported to feature a smaller impurity content, higher density and overall better electronic characteristics[181–185]. The GPC was determined from all the film thickness measurements presented in Chapter 4. All other values are those recommended by the system manufacturer. In order to accurately reproduce the results obtained here, and those presented in the rest of this thesis, the ALD parameters need to be carefully controlled. The most important parameters to maintain with regards to the GPC are the process temperature and duration of the gas purging steps. The temperature of the substrate influences the reactivity of the precursors, where for PEALD-Al<sub>2</sub>O<sub>3</sub>, the growth rate increases significantly at lower temperatures due to the high reactivity of the oxygen radicals and the higher density of OH groups on the substrate surface, as discussed previously. To estimate the growth rate for a given process temperature, see Figure 8 of Profijt *et al.*'s work[177], where the GPC is found to vary between 1.75 and 1 Å.cycle<sup>-1</sup> from room temperature to 400 °C, respectively. The purge duration is important for ensuring the reaction chamber has been cleared of any remaining precursor, and therefore a minimum duration is required to avoid overgrowth and CVD deposition of the film. The values used here were fine-tuned by the manufacturer, however extensive literature suggests that times of at least 5 s are required to ensure the chamber is cleared of oxygen plasma, and significantly longer times are required for the water vapour precursor, particularly at low temperature[177, 178]. The other ALD parameters already discussed typically affect the stoichiometry and

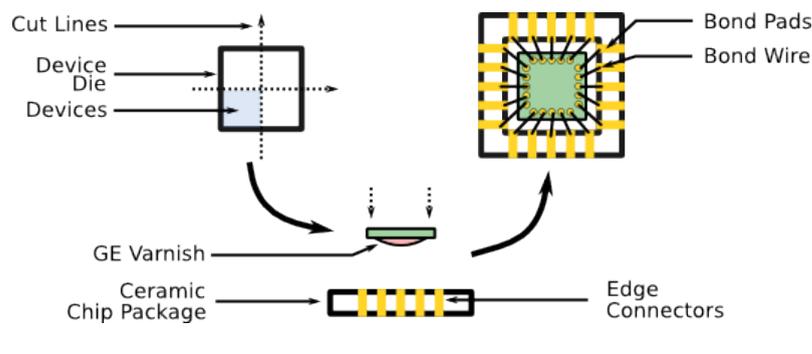


FIGURE 3.5: Diagram of the packaging steps. A finished device die is cut using the wafer saw into four equal pieces, each containing a set of devices. The interior of the ceramic chip package is then coated with a small amount of GE varnish, which produces a good thermal contact. Finally the pads are connected using Au bonding wires.

impurity content of the films, and thus they may not be so important for achieving a desired GPC, but more relevant to detailed comparisons of the material properties of the deposited films.

### 3.2.4 Packaging techniques

A number of tools are required to process the wafers used for device fabrication, and the final device packaging steps. A wafer saw (Loadpoint Microace 66) is used to both dice a quartz wafer into square substrates suitably large for the devices, and to further dice the square substrates once all the materials have been deposited. The wafer saw uses a special blade selected according to the material which is being cut. The blade is connected to a computer-numerical-control (CNC) system that is programmed to guide the blade at the desired velocity and depth to produce a precise cut. The blade material and speed must be selected according to the material, in this case resin blades were used at 20 krpm spin speed and  $2 \text{ mm}\cdot\text{s}^{-1}$  maximum velocity, as recommended for quartz substrates. To reduce the risk of damage to the substrate, the cuts are made in two runs or ‘pecks’. The substrates are placed on a chuck with XY translational control, and is aligned using a camera-optical system. The blade height is calibrated through an automated procedure that measures the blade diameter. The depth of cut can then be determined, considering the thickness of the substrate and the tape placed on the chuck to hold the substrates in place. The substrates used were  $500 \mu\text{m}$  thick, and the tape approximately  $70 \mu\text{m}$  thick, so that the total thickness seen by the blade is  $570 \mu\text{m}$ . A cut  $350 \mu\text{m}$  deep, leaving approximately  $150 \mu\text{m}$  thickness under the cut is sufficient to hold the substrates in place during sawing, and to gently separate the devices after sawing. While cutting, the blade is sprayed with plenty of water to avoid local heating of the substrate and subsequent damage. To protect the devices from damage during this process, the substrates are coated in S1813 and cured at  $115 \text{ }^\circ\text{C}$  for at least 3 minutes.

Figure 3.5 shows a simple diagram of the process. After dicing the processed square substrates, the devices must be placed in a ceramic chip package that provides a means of connecting the device to measurement equipment with a suitable carrier. The square ceramic chip packages used have a side length of 0.9 cm, and consist of 20 Au electrodes approximately 1 mm wide, leaving plenty of space for a connection to be made. The devices are sealed to the chip package using GE varnish, a cryogenic sealant which provides a strong bond to the ceramic chip package and provides good thermal conductivity for cryogenic measurements. The GE varnish is applied to the interior of the chip package, and the device substrate is pressed onto it to spread the varnish equally between the substrate and the chip package. The varnish is then cured at 100 °C for at least 5 minutes which solidifies the structure. Finally, the bond pads on the substrate must be connected to those on the chip package. This is done using a ball bonder, which is used to connect very fine gold wire to both pads. The ball bonder creates a bond in two steps, the first bond is usually made to the device and the second to the chip package. The gold wire is fed through a thin capillary tube. The wire exposed at the opening of the capillary is melted to form a ball larger than the capillary diameter, using a high voltage needle that is brought in close contact with the wire end. The ball then holds the wire in place, such that it can be pulled on when the capillary is lowered towards the substrate. The Z position of the capillary is adjusted for the first and second bond through a calibration step, which consists of setting the maximum displacement of the capillary, such that the ball makes contact with each surface. The first bond is formed by welding the ball to the bond pad through applied force and ultrasonic vibration of the capillary. The wire is then directed to the chip package bond pad, and similarly, the wire is brought into contact with the pad and welded. The second bond step breaks the wire and forms a new ball for the next contact.

### **3.3 Imaging techniques and apparatus**

#### **3.3.1 Optical microscopy**

A series of optical microscopes were used during fabrication, to monitor the lithography steps and to make measurements of the patterns. The microscopes are placed near the wetbenches and the mask aligner used for photolithography processing. A Carl Zeiss Axio-Imager A1m fitted with a camera,  $\times 5$ ,  $\times 10$ ,  $\times 20$ ,  $\times 50$  and  $\times 100$  objectives, and a yellow filter was used to take images of substrates at various stages of processing. The camera is connected to a PC, running image processing software calibrated such that rough measurements can be made of the dimensions of features on the substrate. A Nikon Eclipse LV100ND with bright and dark-fields, and  $\times 5$ ,  $\times 10$ ,  $\times 20$ ,  $\times 50$  and  $\times 150$  objectives was used for similar purposes. The dark-field mode was used to inspect the

surfaces of features, and the areas where resist has been developed to check for small contaminating particles. In the dark field mode, the light path is modified such that light is incident on the substrate only from a steep angle. The light is then diffracted by small edges and particle features on the substrate surface. The contrast between large uniform structures and fine details and defects is then increased significantly.

### 3.3.2 Electron microscopy techniques

A Leo 1530 field-emission gun (FEG) scanning electron microscope (SEM) was occasionally used to observe the edges of deposited structures, however these were readily observable by AFM, which was more versatile for this purpose. Figure 3.6 shows a simplified diagram of the Leo 1530, which consists of a GEMINI electron-optic column, and with an Everhart-Thornley secondary electron (SE) detector. Most of the innovative features of the Gemini column were not required, such as the in-lens SE detector, which requires a very short working distance (distance between the sample and the final objective lens) and low electron energies of the primary beam. A FEG is a fine conducting tip that may be formed of multiple layers, that is biased as a cathode to a very high voltage. In contrast to thermal emitters, in a FEG the potential near the tip is so steep due to the high voltage and the sharp geometry, that the potential energy seen by electrons appear as a triangular barrier. As the voltage is increased the triangular barrier decreases in thickness, and the tunneling probability of electrons increases. FEG sources are preferred over thermal sources in high resolution imaging, partly because the spread of energies is drastically reduced, leading to less dispersion of the electron beam formed in the column, improving the resolution. The interior of the SEM is divided into multiple vacuum regions, such that the FEG is always held under a high vacuum during sample loading. The suppressor plate is slightly negatively biased compared to the FEG, and placed slightly behind the tip apex such that emitted electrons are accelerated away from the tip. The anodes are at a significantly lower potential than the FEG, which provides significant acceleration of electrons into the magnetic condenser lens stage. Magnetic electron-lenses are formed from a high precision coil wrapped in a highly permeable material, where under application of a large current through the coil, a strong magnetic field is induced within the outer material. The air gap in the outer material then presents a magnetic field in the middle of the column that points down the column. The magnetic field causes the electron path to become curled down the column, and therefore can act as a lens if the magnetic field is precisely controlled and maintained. The beam is conditioned and positioned by the scanning lenses, after which a final electrostatic lens fine-tunes the beam focus for optimal resolution. Images are formed by raster scanning the beam while performing measurements of the signals created by the interaction of the electron beam and the substrate.

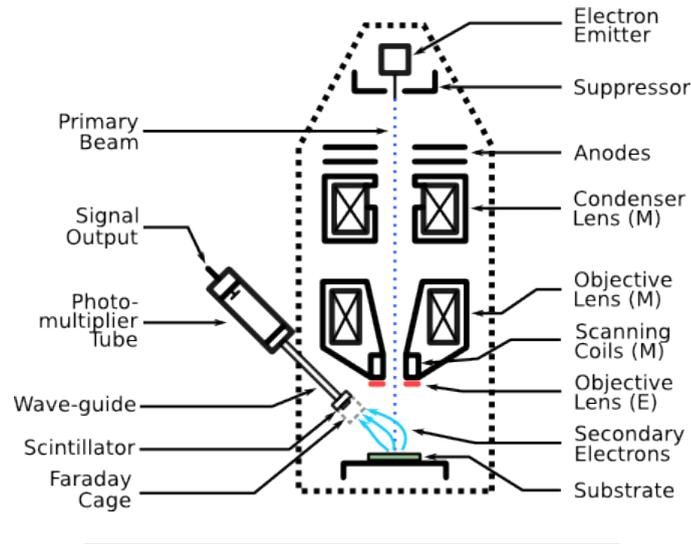


FIGURE 3.6: Diagram of the SEM system used in this work. Many additional components are omitted as they are not important in this work.

When a high energy electron beam is incident on a sample, an interaction volume with various layers corresponding to different interactions occurs. A great number of electron and photon signals are created over a broad range of energies, depending on the interaction. Image creation is performed through the detection of secondary electrons (SEs), which are electrons that have lost most of their energy through inelastic processes. SEs are emitted from within a small volume under the substrate surface at the beam incidence location, and therefore provide topographical information. The quantity of SEs emitted is determined by the topography of the surface, where outward sharp features yield a higher signal due to the reduced recapture of SEs by surrounding material. Therefore the quantity, and not the energy of the SEs, is important for topographical imaging. The Everhart-Thornley SE detector provides a method of quantifying the SE signal. A Faraday cage composed of a fine mesh is positively biased at a high voltage, which attracts the low energy emitted electrons towards a scintillator, such as a phosphor screen. The incident electrons cause emission of electromagnetic radiation from the scintillator, which is coupled to an optical waveguide or fibre. The waveguide directs the light signal to a photomultiplier tube. The photomultiplier tube converts the electromagnetic radiation back to electrons, which are accelerated under vacuum into metallic plates that emit more electrons. The electrons are collected at an anode, creating a measurable current signal that is proportional to the SE intensity. An SE image is formed from the current measurements taken at each incremental X or Y step of the raster scan. The system is operated through a PC running dedicated software and with control hardware extensions, such as controls for zoom, focus, and beam conditioning. The sample is fixed to a standard pin-mount sample holder with a double sided tape suitable for use in high vacuum. The substrates were not processed in any way prior

to imaging in the SEM due to the process flow. Charging effects due to the trapping of electrons in materials may cause significant artefacts in the rendered image, usually anomalously bright regions that build up over time. However for the large metallic features imaged, charging was not a major concern for image fidelity. The primary beam voltage used for imaging was usually between 2 and 5 kV, with an aperture of 30  $\mu\text{m}$ , and a working distance on the order of 4 mm. Resolutions better than 10 nm could be achieved with a slow scan speed.

Transmission electron microscopy (TEM) and scanning-TEM (STEM) was used to produce atomic resolution ( $\leq 0.1$  nm) images of a fabricated MIM structure. TEM imaging is a much more involved process than that described previously, and thus only the important details will be covered. With this technique, a very high energy electron beam is incident on a cross section of the sample. The wavelength of the electrons can be estimated using the De Broglie wavelength,

$$\lambda_{\text{DB}} = \mu(V_a) \frac{h}{\sqrt{2meV_a}}, \quad (3.6)$$

where  $V_a$  is the acceleration potential, and  $\mu(V_a)$  is a relativistic correction factor, as free electrons in such a potential approach a significant fraction of the speed of light  $c$ . The relativistic correction factor for the wavelength is given as

$$\mu(V_a) = \left(1 + \frac{eV_a}{2mc^2}\right)^{-\frac{1}{2}}, \quad (3.7)$$

which decreases the wavelength. For values of  $V_a \geq 100$  kV, the correction factor is  $\mu(V_a) \approx 0.95$ , and the electron wavelength becomes  $\lambda_{\text{DB}} < 4$  pm, much smaller than the inter-atomic spacing, which is on the order of a few  $\text{\AA}$ . If the sample is thin enough, it becomes transparent to electrons, due to the significantly reduced interaction volume. As electrons are incident on the structure, parts of the beam are diffracted to different angles past the sample. As electrons interact strongly with matter, the exit beams are strongly correlated to atomic spacing, which acts similarly to an optical grating of given pitch. A series of objective and projection lenses manipulate the transmitted beams, which are captured by different detectors. Due to the interference of the electron paths, the captured diffraction pattern depends strongly on the geometry of the beam and the sample, which induce phase shifts of the transmitted electron waves. The diffraction pattern is recorded, and an image can be formed from the two-dimensional Fourier transform of the circular diffraction pattern. As a diffraction spot represents a vector in reciprocal-space (or  $k$ -space), there is an associated periodicity along a plane in the real-space image. In this work, TEM was not used to determine the precise atomic positions of composite materials, therefore no detailed modelling of the beam and quantum mechanical behaviour of electrons in the sample was required to determine the exact

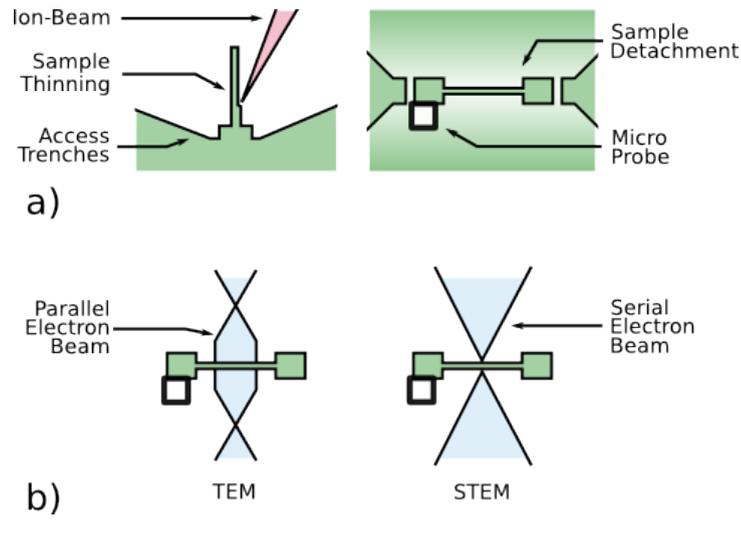


FIGURE 3.7: (a) Schematic showing part of the sample preparation process for creating electron transparent cross-sections. (b) Diagram of the beam shape (qualitative) for the TEM and STEM imaging modes.

conditions for the diffraction pattern obtained. Boundaries between materials are relatively clear in TEM images, due to the changes in the periodicity, composition and density of a structure in real-space, leading to significant changes in the phase of the transmitted electron waves. This allows the basic determination of the thin film properties such as the diameter of metallic grains and thin film thicknesses. There is a good resolution of the boundaries between materials due to the significant  $Z$ -number (atomic number) contrast, which arises as lighter elements deflect primary beam electrons less, yielding a brighter signal (in the bright-field imaging mode) in these regions. Using the dark field mode, i.e. focusing the post-specimen lenses to capture scattered electrons, then reverses the signal.

A significant part of TEM imaging involves the preparation of the sample cross-section, which has to be milled out of the substrate and thinned to a thickness less than 100 nm for good transparency. Figure 3.7 (a) illustrates part of this process, where first a Ga focused ion beam (FIB) system with an *in-situ* SEM is used to carve out trenches and thick support structures at the sides and bottom of the cross-section, and then the finely milled region in the middle that is left quite thick (on the order of  $10\ \mu\text{m}$ ). A microprobe can then be directed and attached to the edge of one of the supports, by depositing a thick layer of tungsten. The region at the bottom of the side trenches is then milled out to separate the cross section from the substrate. The cross section can then be mounted onto a specimen holder or retained on the microprobe. High precision milling of the middle region of the cross section is then performed to thin the sample down to the required thickness for electron transparency. The milling system used was a FEI Helios G4 CX DualBeam, which consists of two columns aimed at the sample

space, used for milling and imaging in one sample space. Similarly to an SEM, the FIB uses a  $\text{Ga}^+$  ion source to produce an ion beam, and is accelerated through a series of lens stages to produce a focused high energy beam. The TEM/STEM microscope used was a FEI Titan Themis 300 kV system, which supports many additional electron microscopy and spectroscopy modes, some of which are discussed in a later subsection. The technical aspects of the sample preparation and operation of the TEM were handled by a trained operator.

Figure 3.7 (b) shows the difference between the TEM and STEM operating modes. In the TEM mode, the sample is exposed to a wide collimated beam, whereas in the STEM mode, a probe is formed with the beam and thus it must be scanned across the surface to produce an image of the cross section. The STEM mode allows new information about the sample to be collected that is not attainable in the TEM mode. In TEM mode, electrons scattered by the sample at a high angle result in a defocusing effect of the diffraction spot projected onto the detector, due to the objective lens chromatic aberration. In STEM, there is no objective lens below the specimen, such that the electrons emitted at high angles are captured directly using large annular detectors, and inelastically scattered electrons can be detected without defocusing effects due to chromatic aberrations. Additionally, emitted SEs yield topographical information in STEM mode due to the small probe size. The Titan Themis has the three modern STEM detectors, the bright-field detector that detects primary beam electrons (non-scattered) and low-angle scattered electrons, an annular dark field (ADF) detector, and a high-angle ADF (HAADF) detector. The ADF detector can be used to detect electrons deflected at a moderate angle, such as electrons elastically scattered by the Coulomb repulsion of nuclear cores (Rutherford scattering), or certain inelastic collision that lie with the ADF detector angle. The HAADF detector is intended to detect only inelastically scattered electrons, usually scattered at the highest angles, so that a very high Z-contrast at a high spatial resolution can be achieved[186].

### 3.3.3 Scanning probe microscopy techniques

Atomic force microscopy (AFM) was used to measure the thickness of deposited materials at the nanoscale, through step height measurement, and to estimate the roughness of surfaces. Figure 3.8 shows a schematic of the essential components of an AFM system. The substrate is placed on a chuck with a vacuum line to hold it in place. The substrate is aligned with the AFM probe, the cantilever and the tip, using the computer controlled stage connected to the chuck. A suitable height of the probe is adjusted by finding the focal point of the camera system with the substrate surface. AFM measures the force of interaction between a very sharp tip, and the surface near the tip. A range of forces are experienced depending on the distance to the sample. These forces cause

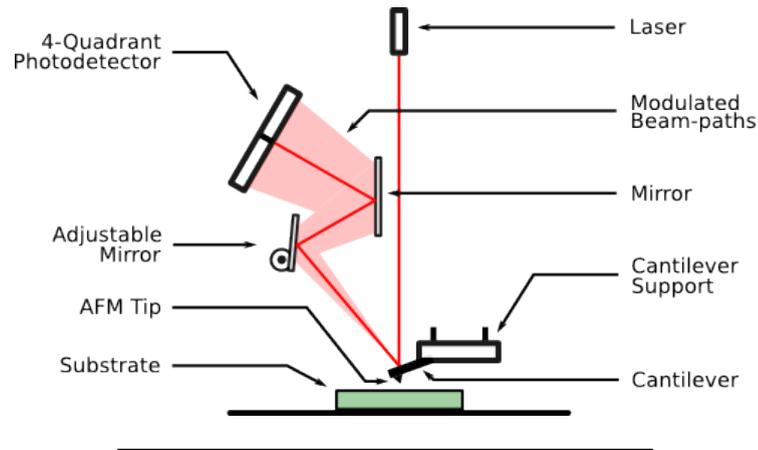


FIGURE 3.8: Schematic of the basic components of the AFM system used.

the cantilever connected to the tip to bend. A red laser is then aligned onto the back of the cantilever, such that the bending cause a change in the beam trajectory. The changes in the beam are captured by an optical system and shone onto a four-quadrant photodiode. An adjustable mirror allows the beam spot to be centred in the middle of the photodiode when not engaged with a surface. When the cantilever is deflected the photodiode detects lateral and transverse motion through the deflection of the laser onto different quadrants, yielding four signals. If the tip is dragged along a surface in the direction parallel to the cantilever arm, the differential signal between the top and bottom quadrant provides a measure of height. The differential signal between the left and right quadrants provides a measure of the torsion experienced by the cantilever, caused by deflections perpendicular to the cantilever arm. Generally the AFM system can be operated in two modes in terms of contact with the surface. A direct mode consists of keeping the tip at a constant height, scanning the surface and measuring tip deflections. Alternatively, in feedback mode the deflection amplitude can be fed back into the height control system, to maintain a given height of the tip above the surface at the given position. These two modes are considered variants of contact mode.

The interaction of the tip with a sample is characterised by the force-distance curve, which is simply a plot of the force determined from the cantilever deflection, against the distance between the surface and the tip (in relative terms). As the tip approaches the surface, intra-molecular forces become important, and can repel or attract the tip, depending on the type of interaction. The  $z$ -scale of the height measurements can be calibrated when the tip first makes contact with the surface, by measuring the deflection of the laser beam as a function of cantilever displacement towards the surface, which for a solid surface yields a linear response. The force experienced by the tip can be determined using Hooke's law

$$F_d = k_{cl}\Delta d \quad (3.8)$$

where  $\Delta d$  is the deflection amplitude relative to the equilibrium position, and  $k_{cl}$  is the spring constant of the cantilever. The spring constant is determined by the geometry and the materials of the tip and the cantilever structure, it is usually provided by the manufacturer if commercial tips are used. However the spring constant is very sensitive to the length and thickness of the cantilever[187], so that quantitative force measurements require accurate knowledge of the tip geometry. If the tip is soft relative to the surface, attractive and repulsive forces will deflect the cantilever at a greater distance than with stiffer cantilevers. The true force-distance curve of the sample surface is found by accounting for the tip displacement, which was previously used as a measure of height.

The tip may be operated in a number of different ways to measure the surface topography of a sample, that provide more information and some significant advantages over the normal contact mode. The system used was a Bruker Icon AFM operated in tapping mode, and the tips used were standard silicon nitride OTESPA tips with approximately 300 kHz resonant frequency. In tapping mode, the tip height is oscillated at or near the resonant frequency of the cantilever, such that when the oscillating tip is lowered towards the surface and impinges on it, the amplitude of the oscillation (the drive amplitude) decreases as a function of distance from the surface. The drive amplitude is then recorded and fed back to the Z-motion controller to maintain a constant tapping force (the tracking force) on the sample. In this mode, weak surface-tip interactions such as Van der Waals forces and liquid surface tension effects are screened by the substantial force of the oscillating cantilever. An image is rendered by raster scanning the tip across a specified surface area up to  $100 \mu\text{m} \times 100 \mu\text{m}$  and measuring changes in amplitude at each grid point of the surface. As the measurement at each position is differential, nanoscale measurements shows constant offsets due to external effects, such as a slight inclination of the substrate on the chuck. The internal image processing software attempts to correct the offset if there is a clear reference plane. With images of a surface showing only one step, the correction usually has to be done manually with image processing software.

### 3.3.4 Compositional imaging and mapping

To determine the elemental composition of the fabricated MIM structures, electron and X-ray spectroscopy techniques were used. Electron energy loss spectroscopy (EELS) and energy dispersive X-ray (EDX) capabilities are built into the Titan Themis system used for S/TEM imaging. In the STEM mode, EDX data can be collected at the same time as forming an image using the HAADF detector, which allows high spatial resolution mappings to be created. When high energy electrons are incident on a material, a series of interactions occur that generate new electrons and X-rays characteristic of the



beam electron occasionally displace core electrons, leaving significant positively charged holes. Electrons from higher energy shells may or may not relax into these holes, depending on the selection rules imposed by the particular transition. The selection rules are represented by the change in quantum numbers of an electron during a transition[189], which results in certain direct transitions being forbidden due to the violation of conservation laws. The characteristic X-rays are named from a specific transition, for example, the  $K\alpha$  X-ray for a given element corresponds to a transition of an electron from the L shell to the K shell. Transitions from higher shells to the K shell are then named iterating the greek alphabet as  $K\beta$ ,  $K\gamma$ ,  $K\delta$  and so on. These shells correspond with the principle quantum number  $n$  (Equation 2.11) of the atom, which represents essentially the ground state of the shell, since the angular and magnetic moment quantum numbers are 0. If the resolution of the system is sufficient to detect the non-degeneracies of electrons within each shell, then the characteristic X-rays can be named further according to the specific energy levels involved in the transition. This level of detail was not required for the experiments in this work however, so only a knowledge of the general shell peak positions and amplitudes are important for the determination of composition. Both EELS and EDX provide similar information that is measured differently.

Fundamentally, EDX is less sensitive to light elements, as the chances of an electron-electron collision are reduced, resulting in their characteristic X-ray intensities being reduced. However compositional mapping using EDX is much easier since the X-ray results are directly interpretable through the classification system. X-rays are emitted at many angles from the sample, and are collected as near to the sample as possible for a strong signal. The EDX detector used was a FEI Super-X 4-detector system with 0.3 nm pixel resolution and an energy spectral resolution of less than 140 eV. The EDX mapping is created by the internal software through automatic determination of the elements at a given primary beam location by identifying the energy of the detector signal peak amplitudes.

EELS is known to be more sensitive to light elements, and has a higher spatial and energy resolution, and can provide more information on the material properties of the sample[190]. However interpretation of the results involves identifying the interactions that produce the collected electrons, making analysis difficult. The electrons in EELS are collected in a Gatan GIF Quantum ER spectrometer with 0.1 eV energy spectral resolution. Simultaneous EELS/STEM imaging was not possible as the system was not set up to use the ADF/HAADF detectors in this mode. EELS results were analysed using the Gatan EELS suite, which uses known data to analyse the collected spectra.

## 3.4 Electrical measurement techniques and apparatus

### 3.4.1 Temperature control

Measurements of the temperature dependence of the (current–voltage) I–V curves of MIM devices were performed using different systems. Cryogenic measurements were performed in a cryogen-free variable temperature insert (VTI) supplied by Cryogenic Limited. The ceramic chips containing the devices are loaded onto a connector at the end of the probe stick, called the cold finger, made of a high thermal conductivity metal and containing a temperature sensor. At the top of the probe stick a multi core cable delivers the connections from the devices to a breakout box with coaxial terminations suitable for the measurement equipment. The cryostat is a continuous flow system, which allows minimal consumption of helium and can be maintained at base temperature as long as required. The system is prepared with two external continuous flow systems (water chiller coupled to He compressor) that successively cool the internal magnets (not used), and the cold parts of the helium circuit shown in Figure 3.10. When the temperature of the helium drops below 4.2 K in certain regions of the circuit, the helium condenses into its liquid phase. The sample space, that is separate from the main helium circuit, is filled with a small amount of helium. The oil-free pump pushes helium vapour around the circuit to first be passed through a charcoal filter that removes unwanted contaminants. The helium vapour is then cooled by a 40 K and 4 K stage and collected at the helium pot. When most of the helium vapour has condensed to the liquid phase in the helium pot, the needle valve can be tightened such that the pump pulls the liquid helium through the constriction, allowing it to reach temperatures as low as 1.2 K. The needle valve output feeds directly to a heat exchanger coupled to the sample space, the walls of which can subsequently be cooled to sub-4 K temperatures. When this happens the helium in the sample space condenses onto the walls, and due to the surface tension of the liquid phase, forms a contact with the cold finger, reducing its temperature. The best heat-exchanger temperature achieved during experiments was approximately 1.6 K, whilst the best achieved temperature of the cold finger was approximately 2.6 K. When the heat exchanger has reached base temperature, the probe stick is mounted to the sample space circuit and sealed with vacuum-tight clamping bolts. The sample space is brought under vacuum using a roughing line to roughly match the pressure in the sample space. On disconnecting the roughing line, the gate valve is opened to allow the transfer of the sample to the heat exchanger. The VTI pressure gauge shows negligible pressure when the helium in the sample space has condensed.

The heat exchanger temperature sensor (sensor A), and the coldfinger temperature sensor (sensor B) are fed into a Lakeshore proportional-integral-derivative (PID) temperature controller, which controls a set of internal heaters that warm up the heat

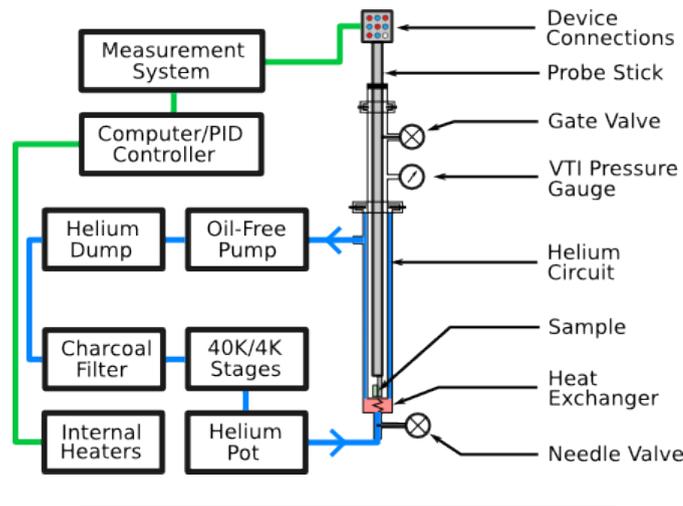


FIGURE 3.10: Schematic of the 4K cryogen-free closed-loop continuous flow system used for low temperature measurements.

exchanger. The PID controller attempts to maintain a set-point temperature of sensor A based on its P, I and D settings, which are set depending on the conditions of the cryostat. The temperature curves of the sensors were calibrated by service engineers and are programmed into the PID controller. The PID controller is connected to a PC through GPIB connectors, and controlled by a Labview program that sets the PID values in the operating temperature range and collects data continuously. A separate Labview program was developed to record sensor A and B measurements at each I–V point measured. Due to the temperature difference between sensors A and B varying with the temperature of the system, the system was operated and inspected manually when changing the setpoint temperature. As there was no on-chip temperature sensor, the measurements were started at least 10 minutes after the setpoint was reached. The system and measurement circuit is grounded and shielded from internal and external noise sources. Temperature dependent I–V measurements ranging from 2.6 K to 290 K were taken for a series of devices.

A high temperature system was developed in-house early during the project to investigate possible thermally activated emission processes occurring in devices with thicker

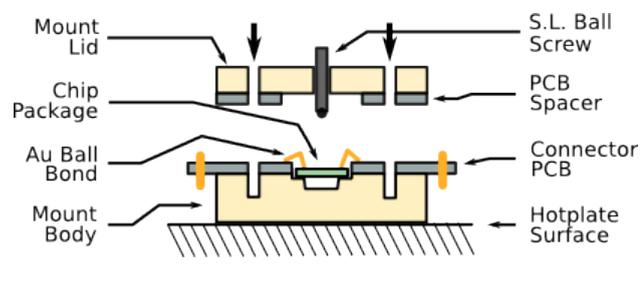


FIGURE 3.11: Schematic of the sample mount which houses the sample and makes a good thermal contact with the hotplate surface.

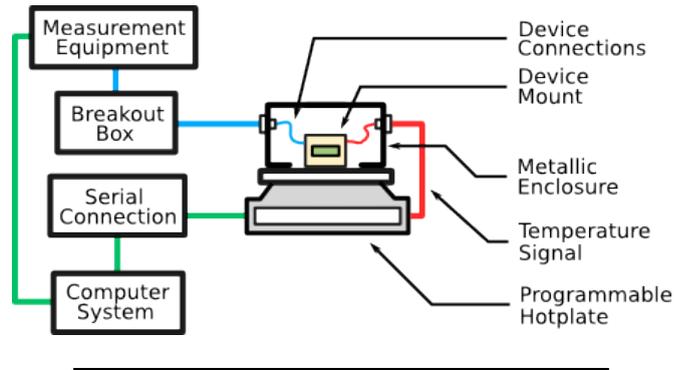


FIGURE 3.12: Schematic of the in-house built high-temperature measurement system, which provides fully shielded connections to the sample mount.

films. It was also used to extend the temperature range of the tunneling I–V measurements of very thin devices to a maximum of 160 °C, or 433 K. The system is based on a programmable IKA RET CV hotplate, which can be operated via a serial RS-232 connection to a PC running Labview. The hotplate has an additional input for an external Pt100/Pt1000 temperature sensor which can be monitored by the internal PID controller, allowing the hotplate temperature to be dictated by the response of an arbitrary external entity rather than the hotplate temperature itself. A cryogenic quality brass device mount was designed and commissioned to a mechanical workshop, to provide a means of making a good thermal contact between the hotplate and the device and simultaneously making low noise connections to the device. Figure 3.11 shows a diagram of the sample mount structure. Two polyimide printed-circuit-boards (PCBs) were designed and commissioned to provide contacts to the device, and a spacer layer that insulates the contacts from the brass lid. The polyimide material provides the highest temperature stability and thermal conductivity combination available, and an electroless-nickel/electroless-palladium/immersion-gold (ENEPIG) plating was used for the PCB traces to provide good resistance to the ball bonder and to improve corrosion resistance. The ceramic chip is placed upside down in a precise indentation on the sample mount, exposing the underside contacts that run from each side. The ball-bonder can then be used to bond the inner PCB edge contacts to the ceramic chip, where using the underside makes removal of the bond wires easier after an experiment. The outer PCB edge contacts were designed to form a plug-pin system with external copper wires that connect to the outer enclosure. Gold plated high temperature resistant plugs were soldered to through holes at the outer edge of the PCB using high-Pb content solder, also resistant to high temperatures. Indentations were placed as near to the devices as possible at the same depth as the chip indentation, allowing Pt100 sensors to be housed, fixed in place using GE varnish thermal contact glue, and soldered to dedicated PCB traces.

Figure 3.12 shows a diagram of the complete hotplate system. The copper wires

are soldered to SMA panel mount connectors (165 °C rated) on one end and gold-plated pins on the other. SMA cables running from the enclosure then connect the temperature sensor and device connections to a breakout box that contains heat sinks (for safety considerations), BNC coaxial connectors for the measurement systems, and a special connector for the temperature sensor feedback cable to the hotplate. Since the top surface of the ceramic chip is gold plated, and the body is made of aluminium oxide, which is widely used for its high thermal conductivity, a good thermal contact can be made with the sample mount body. The mount lid encloses the device and the sensors using threaded screws, and a central spring loaded (S.L. in Figure 3.11) ball screw, which is softly fastened to apply a small pressure to the back of the chip package, holding it in place and ensuring a better thermal contact. The PCB connector board and the brass mount body are cleaned between uses by ultrasonication in acetone followed by a rinse in IPA and N<sub>2</sub> drying. A simple simulation of heat transfer in the sample mount was developed during the design phase in COMSOL, which was used to determine where the most serious temperature gradients occurred. Thermal gradients across different metals induce voltage noise in the circuit due to the Seebeck effect. With a knowledge of the material parameters of the brass mount, the PCBs and the chip package, it was estimated that under an ideal thermal contact the chip package temperature would reach a 100 °C set point in approximately 5 minutes from 20 °C, and the important thermal gradients were mostly along the PCB traces between the devices and the edge plug connectors. The Labview program was made such that setpoints and measurements could be automated. A delay of at least 5 minutes was used between the attainment of a stable setpoint temperature, and the initiation of a measurement cycle. The temperature was recorded for each point as with the cryostat system.

### 3.4.2 Electronic measurement systems

A series of measurement systems were used to electronically characterise the fabricated devices. All connections to the measurement equipment including the temperature control systems were shielded using standard BNC and SMA cables ground at one end, with the ground loops being broken at a specific point in the circuit depending on the mounting system used. For the measurements at room temperature, a series of three small PCBs were created consisting of a soldered chip package connector and BNC connectors at the edges. Each PCB connected two specific chip package connector pins and the remaining 18 connections connected to the grounded coaxial shield. The BNC connectors used had a negligible series resistance (non-50 Ω) in this case, as impedance matching was not necessary at the AC measurement frequencies used. Standard RF coaxial cables were used with a matching impedance of 50 Ω and at most 100 pF.m<sup>-1</sup>

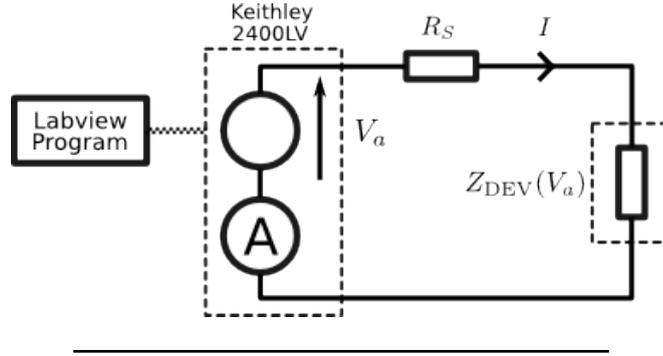


FIGURE 3.13: Circuit diagram of the measurement circuit used to measure the I-V curves of the fabricated devices.

shunt capacitance. The series inductance in all the AC measurement circuits were estimated to be at most on the order of  $100 \mu\text{H}$ , a very small value at the frequencies considered and is thus neglected.

Direct-Current (DC) measurements were performed using a Labview controlled Keithley 2400LV source-measure unit. The equivalent circuit of the measurement system in this case is shown in Figure 3.13. The internal impedance of the 2400LV is less than  $1 \Omega$  and is thus ignored as it is much smaller than all other circuit elements of the system. The applied voltage  $V_a$  induces a current  $I$  that is measured by the internal ammeter of the 2400LV, and the impedance at a given voltage point can be expressed as

$$Z_{DEV}(V_a) = \frac{V_a}{I} - R_S, \quad (3.9)$$

where  $R_S$  is the series resistance due to the device connections, which is due to many contributions along the circuit. The connections in the cryostat system from the measurement end to the chip package connector presented a contribution between  $50$  and  $100 \Omega$ , and in general the fabricated devices had at most a  $100 \Omega$  series resistance between the ceramic chip package edge and the fabricated capacitive junctions. As most devices had a DC impedance much greater than  $R_S \approx 200 \Omega$  over the whole bias range,  $R_S$  was thus ignored in the analysis of the DC properties of the junctions, so that the voltage across the device is taken as the applied voltage. For devices with a high capacitance, a delay between the change in voltage and the current measurement was applied to avoid errors due to capacitive charging and discharging effects. The internal ammeter analog-to-digital converter (ADC) was set to average the measured current over the longest time possible (approximately  $0.5$  s), effectively acting as a low pass filter with a low frequency pole (approximately  $2$  Hz). Further averaging was employed for thick devices showing very high impedances, but generally was not necessary, and was avoided due to risk of damaging the devices under a maintained high voltage.

The first AC measurement system available was a Signal Recovery 7225 lock-in system. Amongst other functions, the lock-in can be used as a voltage signal source synchronised with a phase-sensitive ammeter or voltmeter. The phase-sensitive detector operates by feeding the digitised measured signal to in-phase and quadrature digital-signal-processing (DSP) multipliers. Consider a signal measured in a wide bandwidth system, in response to a source operating at frequency  $\omega_0$ , contaminated by a noise signal at frequency  $\omega_N$

$$f(t) = A_0 \cos(\omega_0 t + \phi_0) + A_N \cos(\omega_N t + \phi_N) \quad (3.10)$$

where  $A_{0,N}$  and  $\phi_{0,N}$  are the amplitude and phase of the response and noise components of the signal  $f(t)$ . This signal is multiplied by demodulation functions  $f_X(t) = 2\cos(n_h\omega_0 t)$  and  $f_Y(t) = 2\sin(n_h\omega_0 t)$  where  $n_h$  is the lock-in harmonic number, always set to unity in this work, and X and Y are the ‘channels’ that represent the signal in the complex plane (Y is the imaginary axis). The resulting signals  $f(t) \times f_X(t)$  and  $f(t) \times f_Y(t)$  are then fed into a low-pass filter with a pole much lower than  $\omega_0$ , such that DC offsets in the signal can be extracted. Using the identity

$$\cos(\omega t + \phi) = \cos(\omega t) \cos(\phi) - \sin(\omega t) \sin(\phi), \quad (3.11)$$

the signal  $f(t)$  can be decomposed, and terms of the form  $\cos^2(\omega_0 t)$  and  $\sin^2(\omega_0 t)$  occur in the X and Y channels respectively. Through the identities  $\cos^2(\omega_0 t) = 0.5(1 + \cos(2\omega_0 t))$  and  $\sin^2(\omega_0 t) = 0.5(1 - \cos(2\omega_0 t))$ , one can see that an offset factor of 0.5 exists due to these terms. The other multiplied components of the waves on the other hand produce terms that on average fluctuate about zero, and given a long enough averaging period, these contributions are suppressed at the low pass output. The value of channel X in response to  $f(t)$  is then given by

$$A_X = A_0 \cos(\phi_0), \quad (3.12)$$

and for channel Y

$$A_Y = -A_0 \sin(\phi_0). \quad (3.13)$$

The complex form of the signal can then be reconstructed from  $A_X$  and  $A_Y$  as

$$f_r(t) = \sqrt{A_X^2 + A_Y^2} \exp \left[ i \arctan \left( \frac{A_Y}{A_X} \right) \right] \exp(i\omega_0 t) = A_0 \exp(i\phi_0) \exp(i\omega_0 t). \quad (3.14)$$

Therefore the amplitude and phase of the response signal can be determined very accurately, and is limited by the integration time constant, that should be set to a value greater than  $20\pi/\omega_0$  s to obtain a stabilised value.

The lock-in amplifier was used in the ammeter mode, and the equivalent circuit in this case is shown in Figure 3.14. In this case the values of  $A_X$  and  $A_Y$  correspond to the real and imaginary components of the measured current. The bandwidth of the ammeter input amplifier was limited to about 2 kHz (in the wide-bandwidth mode), as beyond this frequency the gain depends on the load as well as the frequency, and therefore measurements with a small, frequency-dependent load are very difficult to interpret. At low frequencies, the voltage source is offset by a constant phase that needs to be accounted for in the measurements to obtain correct signal phase measurements. These phase offsets remain constant over time, so they were entered into a lookup table in the Labview program that operates the SR7225. The offsets were measured using a very low parasitic capacitance, 0.01% tolerance metal-foil 100 k $\Omega$  resistor, placed where the ceramic chip package would be in the measurement circuit, so as to simulate an ideal real impedance in the measurement circuit. The phase values obtained at each measurement frequency were then used as the phase offsets corrections for an arbitrary impedance device placed in the measurement circuit. The measurement of this arbitrary impedance is then limited by the frequency response of the measurement circuit itself. From the circuit in Figure 3.14, the internal impedances of the voltage source and the ammeter are given as  $r_s = 50 \Omega$  and  $r_A \approx 250 \Omega$  at 1 kHz respectively[191]. The measurement circuit is then not limited by the stray capacitances unless their impedances become comparable to  $r_s$  and  $r_A$ . Starting from the voltage source side, as long as  $r_s \ll (\omega C_s)^{-1}$ , then the voltage across the left stray capacitor is  $v_a(\omega)$ . From the right side, as long as  $r_A \ll (\omega C_s)^{-1}$ , then the current measured is that coming from the device branch. This is only true if  $r_s, r_A \ll Z_{\text{DEV}}(\omega)$  of course, however  $Z_{\text{DEV}}(\omega)$  is not required to be much greater than  $(\omega C_s)^{-1}$ , as long as the previous conditions are held, therefore the effect of the stray capacitance can be neglected under these conditions. In this case, the impedance derived from the X and Y r.m.s. current values and the known r.m.s. voltage is representative of the device on the ceramic chip.

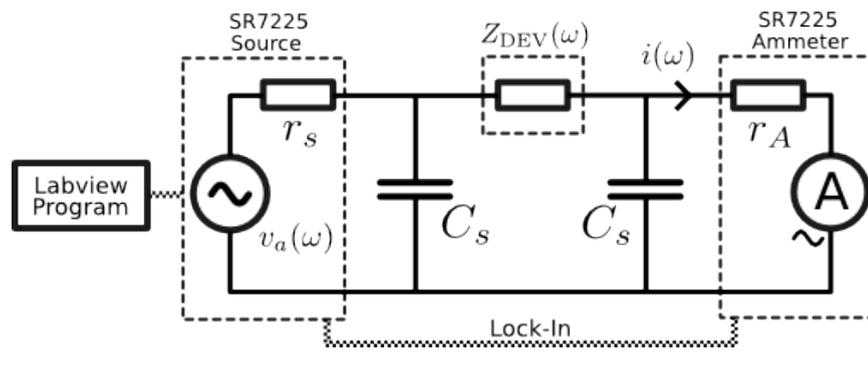


FIGURE 3.14: Circuit diagram of the measurement circuit used to measure the complex impedance of the fabricated devices as a function of frequency.

A different system with a wider bandwidth and load correction capabilities became available at a later time. A Rohde&Schwarz 8118 LCR meter was used to perform AC measurements from 20 Hz to 200 kHz, and to perform capacitance voltage (C–V) measurements. This system provides the same information as the SR7225 lock-in, but with a greater frequency range and flexibility, and the ability to apply a DC offset voltage. The circuit diagram of this measurement circuit is shown in Figure 3.15, configured as a four-wire measurement. The 8118 has open, load and short compensation capabilities, whereby the load of the measurement circuit can be accounted for in the determination of the impedance[192]. To perform this compensation, an empty ceramic chip package was placed in the measurement system to perform a measurement of the open circuit, and for the short circuit, a ceramic chip package with a gold wirebond shorting the relevant contacts was used. The load setting was ignored as it was not necessary. Coaxial T-junctions were used as the PCB mount was very small, so that a separate PCB with four connectors was not necessary. The data was collected by a Labview program in the magnitude-phase configuration, which provides  $A_0$  and  $\phi_0$ . The minimum AC r.m.s. voltage applicable is 50 mV, and the impedance is determined internally. The slow measurement mode was used for maximal accuracy, which similarly to the lock-in, controls the time constant used in the averaging. The consistency between the 8118 and the SR7225 lock-in was verified using a range of capacitors, and the systems were found to have negligible differences up to 1 kHz.

A Keithley 2400LV was used to provide a computer controlled DC offset voltage which enables C–V measurements, however the connection is not direct, rather the 8118 has an internal source matched to the voltage seen at the 2400LV terminals. This prevents the 2400LV from being used in its source-measure mode, which should be configured to only source a voltage. In a C–V measurement, a constant voltage  $V_a$  is superimposed with the AC signal  $v_a(\omega)$  is applied to the device, and the AC current is measured similarly

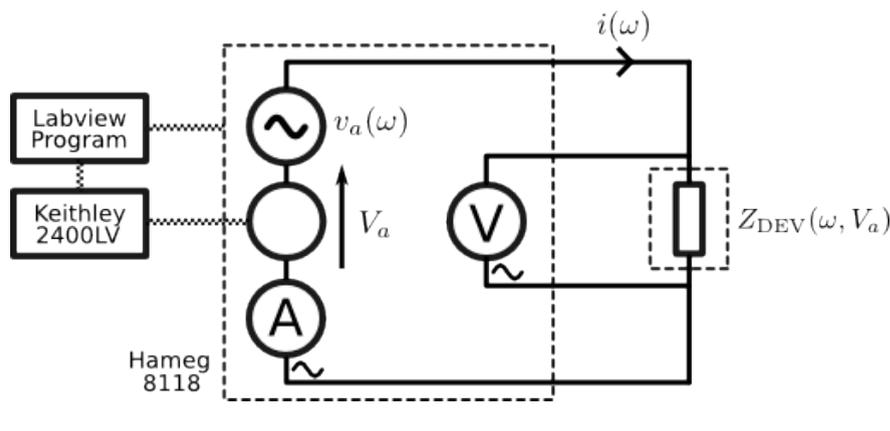


FIGURE 3.15: Circuit diagram of the measurement circuit used to measure the complex impedance of the devices as a function of both frequency and an applied bias simultaneously.

---

to the lock-in method. The voltage applicable in a C–V measurement is limited by the impedance of the device, i.e. when the DC offset current becomes too large for the internal ammeter and voltmeter. This limited the range of the C–V measurement for the more conductive devices, as well as the upper frequency used to determine the AC impedance. Furthermore only a positive voltage can be applied, so that the device PCB had to be flipped to obtain both polarities.



# Fabrication Results and Characterisation

---

## 4.1 Introduction

The fabrication and results of the fabrication of the metal-insulator-metal (MIM) devices are detailed in this chapter. An analysis and discussion of each critical step in the fabrication process is presented in a chronological order. The first section treats the design and fabrication of MIM capacitors and tunnel junctions. These devices are built specifically for investigating the material and electronic properties of these thin films over a large surface area. The results of the characterisation of the fabricated devices are used in Chapter 5 to estimate the performance of a metallic SET with multiple islands. The two critical parameters needed for the estimation of the calculation of the capacitance and current density are the area of the device and the thickness of the insulator film. It is first shown that using a standard photolithography and metal deposition process, geometrically consistent and repeatable metallic electrodes can be created for a wide variety of junction areas, ranging from  $1300 \mu\text{m}^2$  to  $724000 \mu\text{m}^2$ . Next it is shown that the ALD film grows linearly with the number of cycles on both quartz and Au through thickness measurements over a wide range of values from 1.4 nm up to 30 nm. The final section details the final steps of the device fabrication and presents a summary of all devices fabricated.

## 4.2 MIM capacitor design and fabrication

MIM devices were designed to study the electronic properties of ultra-thin  $\text{Al}_2\text{O}_3$  layers deposited by ALD, when in contact with similar and dissimilar metals. The purpose of these devices was to determine the capacitance and the tunneling properties as a function of surface area, and thickness of the  $\text{Al}_2\text{O}_3$  layer over a large area. In such devices the aspect ratio of the critical electrode dimensions to the thickness of the device is on the order of  $10^4$ . As there is a lack of literature that targets the fabrication of SETs using ALD[102, 103] using methods suggested in Chapter 2, although promising, there appears to be little understanding of the material properties of the junctions with respect to the SET parameters extracted from the measurements. In order to close this gap, these large-area MIM devices provide a method of investigating the consistency of the material parameters as the geometry of the junctions is changed, and at  $\text{Al}_2\text{O}_3$  thicknesses of relevance to an SET device. The primary benefit of using large-area

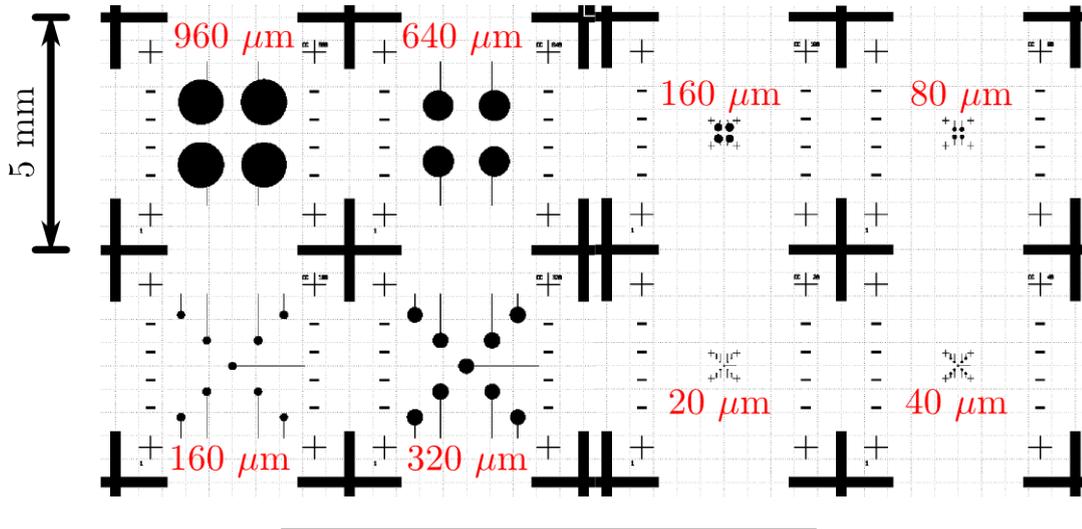


FIGURE 4.1: Mask patterns (polarity inverted) for forming the bottom electrodes. The diameter of each disc in each cell is indicated by the red font.

devices is the ease with which they can be fabricated and measured, compared to complex multi-layer electron-beam lithography fabrication processes and very high sensitivity measurements. Another important benefit is that significant inconsistencies in the film, such as pin-holes, should be rather obvious in an electronic measurement. Of more interest are the subtle inconsistencies in the junctions, which are expected to produce discernible electronic effects that are more easily qualified, or identified. However these effects are not readily quantifiable due to the associated uncertainty in the distribution of defects or inconsistencies, and the difficulty associated with accurately characterising materials at the nanoscale over large surface areas. This issue is addressed in Chapter 6, with regard to the dependence of the capacitance and tunneling properties on the geometry of the fabricated MIM devices.

#### 4.2.1 Design details

The fabrication process is based on a series of photolithography patterning steps and subsequent depositions by electron-beam evaporation (EBE) or ALD, and an  $\text{Al}_2\text{O}_3$  etch step to expose contacts to the coated bottom electrodes. Figure 4.1 shows the inverted mask pattern for creating the bottom electrodes, which is similar to the top electrode mask. The patterns on the mask were designed to align in groups of four, so that the minimum substrate dimensions are 1 cm by 1 cm, each containing 26 devices with a set of four different values of surface area. The stacked MIM structure is formed by depositing  $\text{Al}_2\text{O}_3$  films with a specific thickness by ALD, and subsequently depositing top electrodes of approximately equal dimensions to the bottom electrodes. The junction area is therefore defined by the overlapping area enclosed by the bottom and top electrodes. In the case that the electrodes are well aligned, the smallest diameter disc

(those shown in Figure 4.1 by design) defines the junction area such that

$$A_J = \frac{1}{4}\pi D_c^2 \quad (4.1)$$

where  $D_c$  is the critical diameter, and it is assumed the interfaces are perfectly planar. There is also a very small additional contribution to  $A_J$  due to the overlap of the top electrode connecting wire, which was generally ignored in subsequent calculations, but acknowledged as a small source of uncertainty in the precise area  $A_J$ . The wires were made very narrow on the order of 2 – 10  $\mu\text{m}$  depending on the device, to reduce the stray capacitance. Due to the very large aspect ratios of the structures, a parallel plate model of the MIM structure capacitance is a very good approximation. Circular electrode structures were used to minimize the fringe field strength due to sharp edges. Simulations using COMSOL of the contributions to the capacitance due to the stray fields at the edges of the structure showed that the error is much less than 1% for the smallest area devices. The capacitance of the junction can then be written as

$$C_J = \epsilon_r \epsilon_0 \frac{A_J}{t_J} = \pi \epsilon_r \epsilon_0 \frac{D_c^2}{4t_J} \quad (4.2)$$

where  $t_J$  is the thickness of the junction. The capacitance can be extracted from measurements of the impedance and a suitable equivalent circuit model, which, given an accurate knowledge of  $t_J$ , allows the relative permittivity of the film to be calculated.

Figure 4.2 shows an image of the etchmask pattern (one of four identical patterns) used for devices with electrode diameters from 20  $\mu\text{m}$  to 160  $\mu\text{m}$ . A significant amount

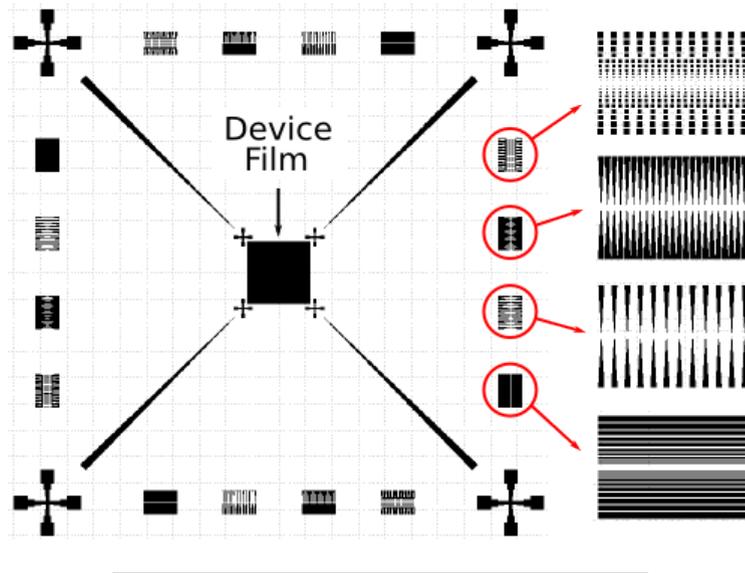


FIGURE 4.2: Mask patterns for forming the etch mask, which protects the deposited ALD film in the region of the electrodes exposing only the tip of the connecting wire. Small structures between 1 and 10  $\mu\text{m}$  were designed around the edges for the purpose of ALD film thickness measurements.

of space (at least  $50 \mu\text{m}$ ) was left between the edge of the bottom electrode disks and the edge of the central etch mask pattern to stop the creeping of a wet etchant under the resist from reaching the device region. Small patterns with critical dimensions ranging between  $1$  and  $10 \mu\text{m}$  are laid out around the edge of the etch mask, which were intended to leave small patterns of the deposited  $\text{Al}_2\text{O}_3$  film that are more easily imaged by AFM. As can be seen in Figure 4.1, small bars are deposited around the edges of the devices which overlap with a portion of the etch mask patterns. This was intended to allow thickness measurements of the  $\text{Al}_2\text{O}_3$  layer on the deposited bottom metal layer, and on the neighbouring quartz surface within a small area for AFM imaging. The deposited films are optically transparent and thus the top electrodes are easily aligned with the bottom electrodes using the mask aligner system. The top electrodes were made separate from the bulk electrodes that connect to the contact pads for the ball bonder. This allowed different top electrode metals to be deposited, whilst retaining gold contact pads with a chromium adhesion layer, suitable for the ball-bonding process.

Figure 4.3 shows a logarithmic plot of the range of capacitance values achievable in principle over the range of areas and thicknesses. The capacitance values were calculated using Equations 4.1 and 4.2 with a value  $\epsilon_r = 9$ , a commonly reported value for ALD deposited  $\text{Al}_2\text{O}_3$ . The chosen values of  $A_J$  result in a significant range of area, and consequently a significant practical range of capacitances on the order of  $10 \text{ pf}$  to  $10 \text{ nF}$ .

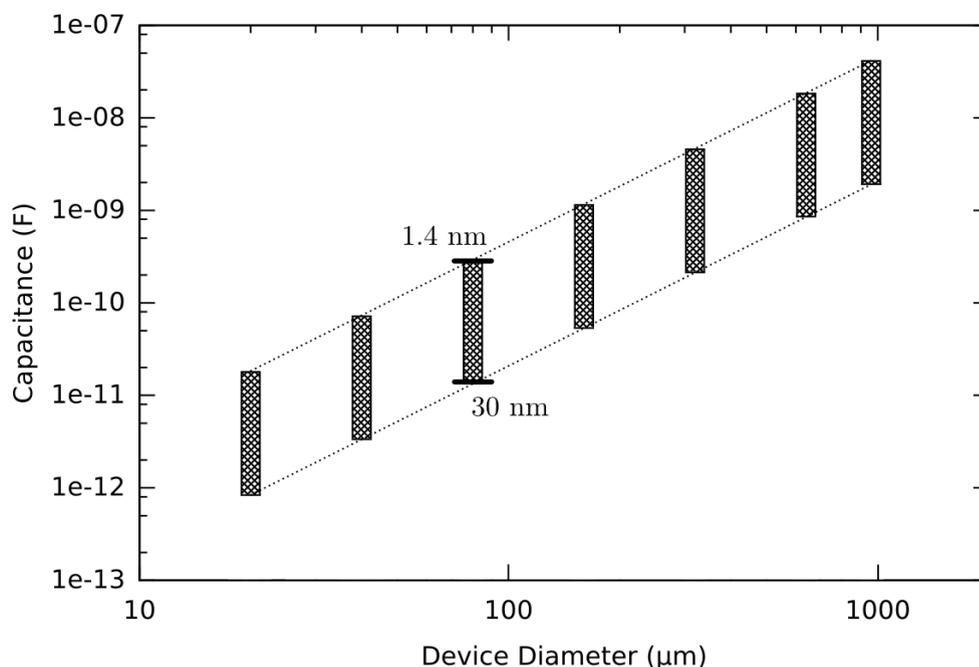


FIGURE 4.3: Logarithmic plot of the possible range of capacitance values provided by each electrode diameter, where the height of the box represents the range provided by the change in thickness. A relative dielectric constant value of  $\epsilon_r = 9$  was used.

The fabrication of devices was performed in batches of up to 8 substrates at once, and a certain range of values of  $t_J$  and  $D_c$  were chosen for each batch. Capacitors with diameters greater than  $160\ \mu\text{m}$  were generally only used for thick films ( $t_J \geq 10\ \text{nm}$ ), and thus were usually used only for measurements of the dielectric constant as a function of thickness. Devices with diameters  $160\ \mu\text{m}$  or less, and with thin films were used for I–V measurements as well as capacitance measurements. Although thicker film devices were characterised, the conduction mechanisms at such length scales are not directly of interest to SET devices and thus less effort was invested in quantifying the conduction mechanisms. These devices are more suited to determining the bulk properties of the film[193], which may or may not show a consistent behaviour with the thin-film devices.

### 4.2.2 Fabrication process

The fabrication process was developed early on in the project and did not change significantly between produced batches, therefore the general process is described below in sufficient detail to be reproduced, and the important details are treated in further sections. The MIM devices were fabricated using  $500\text{-}\mu\text{m}$ -thick quartz, single-side polished wafers, diced into  $1.7\ \text{cm}$  by  $1.7\ \text{cm}$  square substrates using the wafer saw. The substrates were then cleaned by ultrasonication in water to remove large particulates (at least 5 minutes), in acetone to dissolve organic compounds and water residue (at least 5 minutes), and finally in IPA to remove acetone residue (at least 1 minute). The substrates are dried using a  $\text{N}_2$ -gun for at least 20 s. The substrates surfaces were observed using the dark-field microscope to check for damage or optically visible residue.

The bottom electrode resist layer was formed using optical lithography with a LOR-3A/S1813 photoresist bilayer to ensure a good undercut is produced. Prior to LOR-3A deposition, the surface must be treated with HMDS to ensure good adhesion between the resist layer and the substrate for contact lithography. The substrates are placed on a hotplate at  $200\ ^\circ\text{C}$  to evaporate remaining solvent and water molecules on the surface. Substrates are then placed (serially) on a spinner chuck with the vacuum turned on, and HMDS solution is applied to the substrate surface using a small pipette. The chuck is spun at  $5\ \text{krpm}$  for 30 s to finely distribute the solution on surface, after which it is placed on the  $200\ ^\circ\text{C}$  hotplate for at least 1 minute. A short rinse in IPA then cleans the surface of HMDS solution residue, and the substrates are again left on the  $200\ ^\circ\text{C}$  hotplate to evaporate IPA residue. The substrates are then placed (serially) into a different spinner, and coated with a LOR-3A layer at  $500\ \text{rpm}$  for 5 s to coat the substrate evenly, and  $5\ \text{krpm}$  for 45 s for a uniform coating, approximately  $300\ \text{nm}$  thick. This layer is then baked between  $180$  and  $200\ ^\circ\text{C}$  for 3 minutes. It was found at this stage that a good thermal contact between the hotplate surface and the substrate is indispensable, since a poor contact under-bakes the resist and causes it to dissolve much

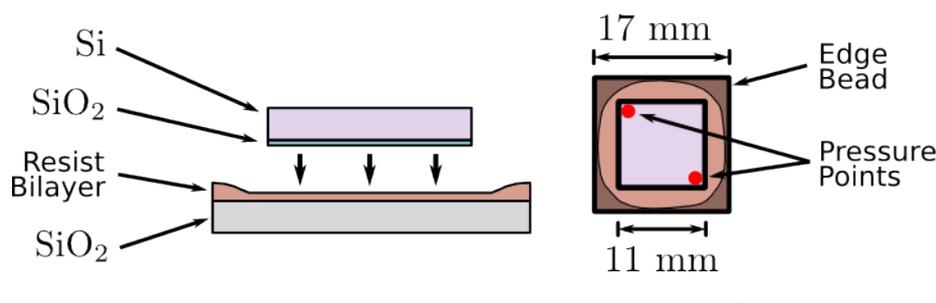


FIGURE 4.4: Schematic of the EBR process for ensuring a good contact with mask during contact lithography processing.

too fast, which subsequently causes the resist bilayer to collapse, or delaminate from the substrate upon development in MF-319. For the S1813 layer, the substrates are placed on a spinner chuck, and S1813 is dispensed and spun at 500 rpm for 5 s followed by 5 krpm for 45 s, yielding a resist thickness of at most  $1.2 \mu\text{m}$ . The layer is then baked at  $115^\circ\text{C}$  for 2 minutes.

As the wires connecting the devices in some cases are very small ( $< 5 \mu\text{m}$ ), an edge-bead-removal (EBR) process was used on such substrates to ensure the distance between the mask and the resist was minimized. Edge-beads arise at the corners of the substrate due to the surface tension of the S1813, which causes these regions to have a resist thickness on the order of three times the nominal thickness, preventing a good contact between the mask and the central resist region. Figure 4.4 shows a diagram of the EBR process, where Si/SiO<sub>2</sub> 1.1 cm by 1.1 cm square substrates, prepared and cleaned identically to the quartz substrates, are turned upside down and gently placed on the resist surface. A small amount of pressure is then applied to the back of the Si/SiO<sub>2</sub> substrate using tweezers, to hold it in place during transferral to the mask aligner. The substrates are placed up to four at a time on the mask aligner chuck, and exposed for 4 s without a mask placed between the final collimating lens and the substrate. The substrates are then gently agitated in MF-319 developer for 45 s to clear the exposed edges of resist, and rinsed in a large beaker of DI water for at least 20 s, followed by N<sub>2</sub> drying until the surface is water free. The remaining resist is then exposed for up to 8 s in the mask aligner, now with the patterned mask placed under the collimating lens. A good contact is achieved between the central resist area and the mask patterns by applying a high pressure to the bottom surface of the substrate, such that the mechanical control is finger-tight. The resists are then developed in a large beaker of MF-319 for 45 s, or until the thinnest wires have developed which can be up to 1 minute depending on the exposure time and quality of the contact, and finally rinsed in DI water for at least 20 s. The resist patterns are inspected in the bright and dark-fields to ensure that there is no resist residue, and that the exposed quartz surfaces are free of visible contaminants. The substrates are then placed in an O<sub>2</sub>-plasma asher for 40 s, including a 10 s ramp,

and a sustained 50 W plasma power for 30 s, to clean the surface of adsorbed organic compounds, and immediately loaded into the EBE system and placed under vacuum.

Processes were planned such that the substrates could be left in the EBE system vacuum chamber overnight, to ensure the substrates have had plenty of time to outgas. A chromium layer at least 5 nm thick was always deposited for the bottom electrodes, to ensure good adhesion with the substrate. This layer was deposited between 0.1 and 0.5 Å.s<sup>-1</sup>. A gold layer between 25 nm and 45 nm thick is then deposited to form the bottom electrode surface. This layer was deposited at a rate no greater than 5 Å.s<sup>-1</sup> to minimize surface roughness due to uncontrolled build up of certain crystal planes over others, but kept at a high enough rate to ensure the deposition time is short to minimize adsorption of impurities in the chamber. The lift-off of the resist is then performed in a solution of 1165 remover placed on a hotplate at 75 °C, for at least 10 minutes. The substrate is then removed from the solution with tweezers, and the acetone wash bottle is used to spray the surface, removing 1165, resist and metal residue as much as possible. The substrates are then rinsed in IPA and blow dried with the N<sub>2</sub> gun, to leave a clean surface on the top electrodes. The electrode surfaces are then observed in the bright and dark fields of the optical microscope to check for defects. In some cases, resist residue can remain, thus an additional round of heated 1165, acetone and IPA rinses may be needed to ensure the surface is clean. The substrates are ready for deposition of the Al<sub>2</sub>O<sub>3</sub> film by ALD at this stage.

The ALD O<sub>2</sub>-plasma source was used to clean the samples in the chamber, for 2 minutes at 200 °C and 300 W, after which the dielectric layer was deposited by ALD using the TMA/O<sub>2</sub>-plasma method at 200 °C (process parameters are detailed in Table 3.2). The growth-per-cycle (GPC) of this process was initially measured using a blank Si/SiO<sub>2</sub> substrate exposed to 100 ALD cycles. An S1813 resist layer was then deposited and patterned using an optical microscope for exposure (approximately 2 minute exposure at full brightness and focused on the resist surface), which served as the etch mask. The Al<sub>2</sub>O<sub>3</sub> film was then wet-etched in a solution of orthophosphoric acid (85%), commonly used for Al<sub>2</sub>O<sub>3</sub>[194], heated to 100 °C for a much longer time than needed. The formed step was then measured by AFM and the GPC was determined to be approximately 0.114 nm.cycle<sup>-1</sup>. The next section details the film thickness characterisation. This rate was then used to deposit devices with a specific design thickness, detailed in Table 4.1. The film is deposited everywhere on the substrate and thus an etch step is required to expose the bottom electrode connecting wires. S1813 was found to be a suitable masking material for short etch times, as minimal damage was incurred, as long as the substrate was not left at bottom of the beaker. Acid resistant tweezers (PTFE was suitable) were used to handle the substrates. An approximate etch rate was determined in orthophosphoric acid by observing the progressive change in colour of the Si/SiO<sub>2</sub> substrate between 1 minute exposures of the test substrate to the heated acid solution.

It is well known that the colour of the Si/SiO<sub>2</sub> substrate is strongly dependent on the thickness of the SiO<sub>2</sub> layer and additional dielectric layers. This fact is commonly used to determine the number of monolayers in graphene sheets deposited on Si/SiO<sub>2</sub> substrates[195, 196]. Using this method, an etch rate of approximately 2 nm.min<sup>-1</sup> was determined by finding the etch time at which the substrate colour no longer changes, and using the film thickness value determined by AFM. An accurate etch time was generally not required, and in most cases the substrates were gently agitated in the acid solution for longer than required by up to a few minutes, to ensure the electrode contact wire is clean of any insulating material. Cleaned blank Si/SiO<sub>2</sub> substrates were placed in the ALD chamber alongside the devices for each thickness point, which serve as monitor substrates, used to easily measure the thickness of the film.

The Al<sub>2</sub>O<sub>3</sub> etch mask was formed by first priming the surface with HMDS, then depositing S1813 spun at 500 rpm for 5 s, 4 krpm for 45 s and subsequently baked at 115 °C for 2 mins. A shorter exposure time up to 4 s was used as this mask pattern is mostly transparent, and the small features are easily over-exposed such that they are not well resolved after development. The patterns are developed in MF-319 for 45 s, rinsed in DI water, and baked again at 115 °C for up to 5 minutes to evaporate remaining solvent, reflow (diffuse) the resist edges, and increase the resistance of the S1813 to the etchant solution. The substrate can then be placed in the etchant solution for the desired amount of time. It was found that for etch times exceeding 5 minutes, the resist would incur significant damage, and the acid would creep a significant distance under the resist, making AFM measurements of the step impossible. Despite this, the distance between the etch mask edge and the electrodes was still sufficient to prevent exposure to the acid, since the regions attacked by the etchant were readily observable optically. The post-bake step was found to be essential in maintaining the adhesion and resistance of the S1813 layer when in the etchant for a long time (only for films with  $t_J \geq 10$  nm). Following the etch, the resists are stripped using 1165-remover solution placed on a hotplate heated to 75 °C, and subsequently rinsed in acetone and IPA, and blow dried.

Once the dielectric films were deposited, cleaning by ultrasonication was avoided as the films are fragile when deposited on gold. This was determined in the very first batch of devices made, which all showed short circuits. Subsequent batches were fabricated almost identically, without cleaning by ultrasonication and the number of short circuited devices was drastically reduced. To form the top electrodes, a PMMA/S1813 bilayer was required to protect the underlying Al<sub>2</sub>O<sub>3</sub> film from the MF-319 developer, which is a

<b>Cycles</b>	12	18	26	44	88	176	264
<b>Thickness</b>	1.4 nm	2 nm	3 nm	5 nm	10 nm	20 nm	30 nm

TABLE 4.1: Number of cycles used for the corresponding desired thickness, according to the determined GPC, and assuming linear growth.

strong etchant at room temperature. MF-319 was not used as the etchant although this is simpler in practice, as significant residue was observed at the  $\text{Al}_2\text{O}_3$  edges when a test was performed, possibly due the formation of a precipitate, and thus was not used as it made AFM imaging more difficult. The PMMA/S1813 bilayer can be achieved in a number of ways, however in each case the results were not significantly different, thus the most reliable process used is described, which was used for the most important devices presented in this work. A 495k A8 PMMA layer was deposited, by spinning the substrate at 500 rpm for 5 s, and 4 krpm for 45 s, which yields a thickness of approximately 400 nm. The PMMA layer is then baked at 150 °C for 8 minutes. The substrate is then placed in the S1813 spinner, and S1813 is deposited at 500 rpm for 5 s and 5 krpm for 45 s. The time between the application of the S1813 and the initiation of the spinning routine should be minimized to avoid excessive reaction between the PMMA and S1813 layers, and similarly should be baked as soon as spinning has finished. The S1813 layer is baked at 115 °C for 2 minutes. The EBR process can be performed at this stage, identically to the first layer. Afterwards, the S1813 is patterned by aligning the top electrode mask with the bottom electrodes, exposing for at least 10 s, and developing the resist in MF-319 for 45 s. The substrates are then placed on a hotplate at 140 °C for at least 30 s, to improve the resistance of the S1813 layer to the UVO cleaner, and the developer used for the PMMA under-layer. The PMMA regions opened up by the S1813 development are then exposed in a UVO cleaner for approximately 20 minutes. The UVO tray was moved as near as possible to the DUV bulbs, to make the best use of the power output available. To develop the PMMA layer, a solution of 1:3 MIBK:IPA was prepared during the UVO exposure, after which the substrates are gently agitated in for 30 s, and rinsed in IPA for at most 20 s, and promptly blow dried. Undiluted MIBK, or a 7:3 IPA:H<sub>2</sub>O mixture may also be used as developers, however it was found that the 7:3 IPA:H<sub>2</sub>O developer attacks the S1813 more, and thus limits the development time. The surface is then inspected under the bright and dark-fields to check that the PMMA layer has been removed. If the surfaces are not clear, the substrates can be placed in the UVO for an additional time, and the development repeated until the surfaces are clear, which is made possible by the hardening bake step. Finally, the substrates are placed in the O<sub>2</sub>-plasma asher for 40 s, as for the bottom electrode process.

The first metal deposited as to form the top electrode will influence the properties of the junction. Devices formed of a chromium/gold layer of thickness 5 nm/65 nm were formed by EBE, using the same deposition rates as for the bottom electrode, forming an Au- $\text{Al}_2\text{O}_3$ -Cr junction. These devices, featuring a Cr top electrode, will be referred to as type-A devices or junctions. In a separate batch, devices formed only of a single gold layer 80 nm thick were created, again at the same deposition rate as other electrode layers, forming an Au- $\text{Al}_2\text{O}_3$ -Au junction. These devices are referred to as type-B devices or junctions. The large thickness was chosen to ensure continuity

over the step caused by the bottom electrode and the deposited dielectric. The lift off is then performed in 1165 remover solution, identically to the other lift-off processes used. The quality of the undercut of the PMMA/S1813 was found to be quite poor and unpredictable, but sufficiently good for the desired regions to lift off properly. The effect of such an undercut can be observed in the dark-field microscope, which reveals a highly reflective edge, due to the so called ‘lilly-padding’ of the metal edges, which can also be observed and measured by AFM. This effect was only a concern for the deposition of the contact pad layer and continuity of the contact with top electrode wire, since the metal edges can be raised quite significantly, shadowing the evaporation source. Some devices showed open circuits, which could be attributed to this, however there was no easy way of testing this. Since a sufficient amount of devices were functional, the process was not improved, although it is acknowledged as a source of uncertainty in the reliability and reproducibility of the process. A more suitable solution to tuning this process would be to use a resist bilayer designed to work in this way, such that the  $\text{Al}_2\text{O}_3$  layer can be protected, and a good undercut achieved.

To form the contact pad layer, the LOR-3A/S1813 bilayer was used. The substrates are first primed with HMDS, and coated with LOR-3A at 500 rpm for 5 s, 2 krpm for 45 s to provide a larger undercut. The substrates are placed on a hotplate at 200 °C for 3 minutes to bake the layer. S1813 is then deposited at 500 rpm for 5 s, 5 krpm for 45 s, and baked at 115 °C for 2 minutes. The substrates are then aligned in the mask aligner with the correct pad patterns, exposed for 4-5 s, developed in MF-319 for 1 minute, rinsed in DI water and blow dried. The development time is increased to increase the undercut, as the metal pad layer is quite thick. After routine inspection under the optical microscopes, the substrates are placed in the  $\text{O}_2$ -plasma asher for 40

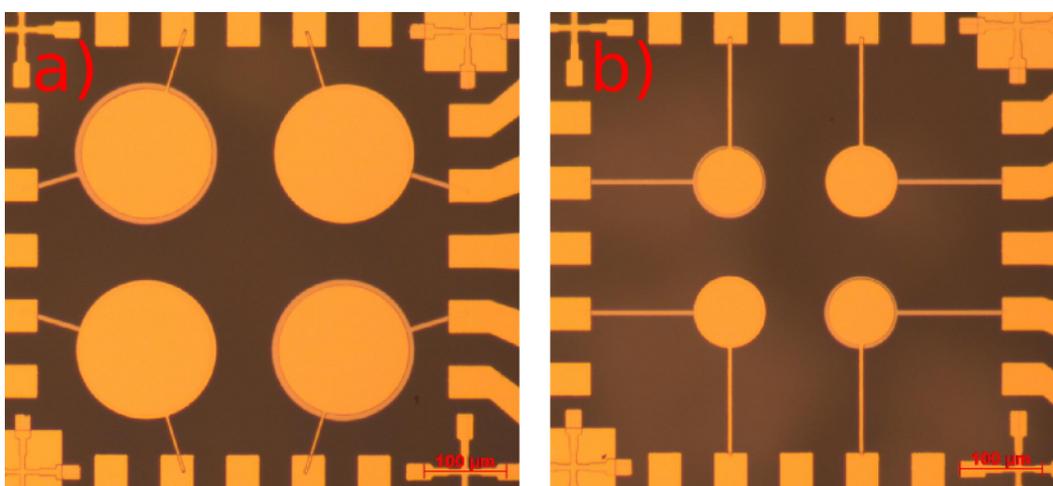


FIGURE 4.5: Images of 160  $\mu\text{m}$  (a) and 80  $\mu\text{m}$  (b) diameter finished devices, taken using an optical microscope.

s, before being placed in the EBE chamber, and held under vacuum overnight. Contact pads were then deposited, formed of 5 nm Cr/120 nm Au for all devices. The deposition rates used for Au were higher ( $\approx 1 \text{ nm}\cdot\text{s}^{-1}$ ) in this case to reduce time, as roughness was not a concern for this layer. The lift-off step is then performed identically to the previous lift-off steps.

Prior to dicing the devices, the substrates are coated in S1813 to protect the devices from material ejected by the wafer saw. The devices can then be placed upwards on the wafer saw chuck, so that the saw is easily optically aligned with the required cut lines. A series of 8 cuts are required to separate the devices completely, with their final size being on the order of 4 cm by 4 mm. The devices are then sealed to a chip package using a small amount of GE varnish, and baked at 100 °C. The pads are connected to the chip package with gold wires, using the ball bonder. Finally the devices are inspected to determine obvious faults, and recorded in a notebook for referral when performing measurements. Discerning between the bottom and top contacts is essential in interpreting the measurements.

Figure 4.5 shows images of finished devices taken using an optical microscope. Measurements of the dimensions of the electrodes using the optical microscope showed that an error up to 5% in excess of the target value was obtained, due mostly to the long exposures required for the narrow wires connecting the electrode discs. This translates into an amplified error in the surface area of the electrodes, since it depends on the square of the diameter. This is discussed in more detail in the next chapter.

### 4.3 Electrodes and ALD film characterisation

The most critical dimension in the fabricated devices is the thickness of the ALD film. In this section measurements of different film thicknesses, using AFM measurements and high resolution TEM images are presented and discussed. The linearity of the growth as a function of thickness is discussed, and the consistency between the results obtained by AFM and TEM is assessed. The last subsection discusses the geometry of the bottom electrode surface, through an analysis of the statistical properties of AFM images of the gold surface, and comparing these results with the TEM cross section images obtained.

#### 4.3.1 AFM ALD-film characterisation

As discussed previously, Si/SiO<sub>2</sub> substrates were used to monitor the ALD depositions, and small structures around the edges of the devices were included to make AFM analysis more straightforward. The monitor substrates allow a measurement of the thickness of the Al<sub>2</sub>O<sub>3</sub> film without interfering with the device, however, it is not entirely representative of the film thickness on the gold regions of the device. The edge patterns provide a means of measuring the thickness of the Al<sub>2</sub>O<sub>3</sub> film, and hence the

growth rate, on both gold and the  $\text{SiO}_2$  surface. In many cases however, due to the variability of the wet etching process and the extra time used to ensure the contact is properly exposed, many of the small etch mask patterns lifted off in solution, making the remaining patterns difficult to find and interpret. A similar limitation was encountered when imaging the thin film monitor substrates, where the edges were difficult to find and could not be imaged in a sufficiently small area to capture the whole step, which is diffuse due to the wet etching process. As discussed in Chapter 4, the growth of  $\text{Al}_2\text{O}_3$  on an oxide substrate is expected to be linear, i.e. a constant GPC, since a sufficient quantity of OH groups should exist on the surface[98], allowing saturation of the surface with TMA molecules. Furthermore, the *in-situ* plasma exposure prior to the deposition cycle was always used, so that it is very likely that OH groups can be formed from the oxygen radicals and parasitic hydrocarbon compounds always present due to the vacuum pumping system. However gold is known to be highly inert, and thus there is a

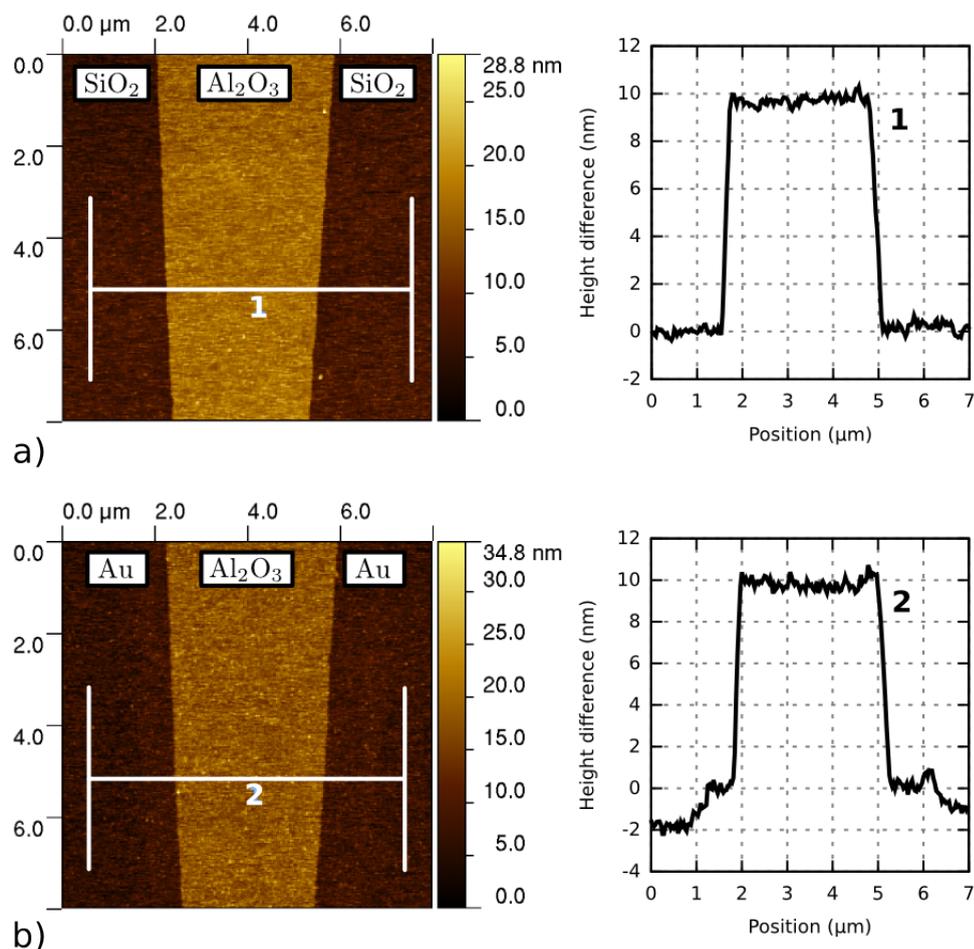


FIGURE 4.6: AFM images of  $\text{Al}_2\text{O}_3$  strips on Au (a) and  $\text{SiO}_2$  (b) surfaces, and the corresponding plots of the profiles delimited by the solid white numbered lines, showing an average thickness of the film of approximately 10 nm. The averaging region is the area delimited by the solid white lines.

concern as to the linearity of the growth in the first cycles, before the surface has been covered in an oxide. This is called a nucleation delay, where coverage of the surface by TMA groups is not complete, and has been observed on some metallic or semiconducting substrates, such as hydrogen passivated Si[197] although very small, on the order of 2-3 ALD cycles. In contrast, on an Al substrate, there is generally an initial overgrowth in the initial cycles due to the oxidation of Al surface during the oxygen based precursor exposure[198]. Generally the growth is reported to be linear on many substrates including gold[199]. This is discussed in more detail in the next section concerning the TEM images, which provide more conclusive evidence as to the behaviour of the growth on gold.

AFM images were analysed using the open-source program Gwyddion, which provides the necessary capabilities for basic analysis of the image properties. Figure 4.6 shows AFM images taken of some of the small  $\text{Al}_2\text{O}_3$  patterns (see Figure 4.2) created by etching on  $\text{SiO}_2$  (a), and Au surfaces (b), of a device fabricated with 88 ALD cycles, corresponding to approximately 10 nm desired thickness. As can be seen in the profile plot to right of the AFM images, the average thickness measured by AFM is very close to 10 nm on both the  $\text{SiO}_2$  and Au surface, which shows that the growth appears to be linear to within very few ALD cycles. The averaging region is the area delimited by the solid white lines. The profile plot in Figure 4.6 (b) shows what appears to be a 2 nm step either side of the  $\text{Al}_2\text{O}_3$  strip, however this is due to the levelling process used on the surface of the strip during image processing. The original image has wave-like background which is likely due to some larger scale fluctuation of the Au surface. To ensure such structures are real, the scan speed of the AFM is reduced to prevent the tip from jumping from the surface and gliding significantly, and the tapping force can be increased to close the gap between the trace and retraces further. When the traces and retraces of the tip are observed to overlap, this is a good indication the observed texture is real.

Figure 4.7 shows AFM images of the step height of monitor substrates used in a 5 nm (a) and 3 nm (b) process. The profile plots shown to the right of the images correspond to the numbered solid white lines in the image, and show that the average thickness of the film is very close to the targeted thickness, using the constant GPC value. The thickness values obtained from different monitor substrates used at different times were generally very similar, showing that the ALD process is highly reproducible and repeatable to a remarkable accuracy. These results are summarised at the end of this section including the measurements made with TEM images. Monitor substrates with films less than 3 nm in thickness were generally not imaged, due to the difficulty in finding the etched edges and analysing the resulting images, as the thickness in this case approaches the scale of the surface roughness of the  $\text{SiO}_2$  surface. Due to similar difficulties, images of sub-10 nm films on gold surfaces could not be produced due to

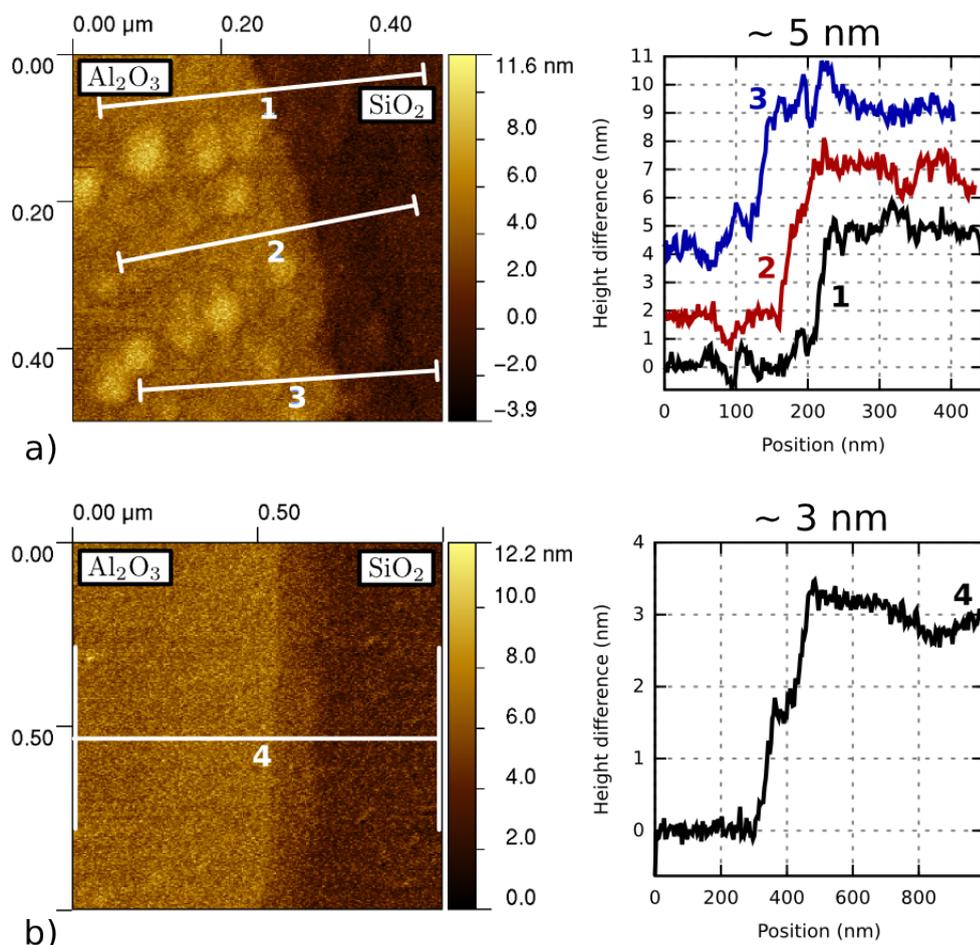


FIGURE 4.7: AFM images of 5 nm (a) and 3 nm (b) process monitor substrates, and the corresponding profile plots to the right, showing that the average thickness is very close to the targeted thickness. The profiles in plot (a) are offset for clarity.

the non-ideal performance of the etch mask. Despite this, these thickness measurements using AFM strongly suggest that the ALD growth is linear within very few cycles. The z-scale accuracy of the AFM system under ideal conditions is on the order of 0.1 nm, limited by the detector or height motion actuator resolution, however many factors can influence this value, including the geometry of the tip and the difference between the tip- $\text{SiO}_2$  and tip-Au interactions. The use of tapping-mode minimises these effects to a great extent, therefore the accuracy of these measurement is possibly on the order of 2-3 cycles, corresponding to at most a 3.3 Å error. This is made more evident in the thickness measurement plot (Figure 4.10 (a)), discussed in the next subsection.

### 4.3.2 TEM cross section characterisation

Cross section images of a sample from a device featuring a 2 nm  $\text{Al}_2\text{O}_3$  film were taken using the TEM system described in Chapter 4, to observe the morphology of the interfaces between all the materials of the MIM structure, and to measure the thickness

to within 0.1 nm. This particular device was studied at length alongside another device, the details of which are discussed in Chapter 6, and was characterised electronically, including against temperature using the systems described in Chapter 3, ranging from 2.6 K to 433 K. Despite these extensive measurements, the device remained intact, and provided consistent electronic characteristics over a period exceeding a year. Due to this, and the fact that the fabrication process was not changed much, and only the thickness of the  $\text{Al}_2\text{O}_3$  film was significantly changed, the material properties of this single device were assumed to be representative of the batch it was part of during fabrication, since the electrode depositions were done in parallel in each batch, and the ALD process is highly repeatable and performed identically for each set of given thickness devices. Additionally, the properties observed for the ALD deposited  $\text{Al}_2\text{O}_3$  film are likely to be representative of films of different thickness, and those deposited in different batches of devices. The only conditions that make a difference at the nanoscale between batches are the variability of the bottom electrode gold deposition by EBE, which cannot be easily controlled, since this depends on many factors in addition to the deposition rate detected by the crystal monitor. The variability between devices is then attributed to extrinsic and intrinsic defects incurred during the fabrication of the devices, but not to variability in the  $\text{Al}_2\text{O}_3$  film. The geometrical properties of the  $\text{Al}_2\text{O}_3$  film observed in this device are found to be conformal to the bottom electrode surface, and highly uniform and homogeneous. The spatial resolution of the TEM system is within one ALD cycle, which corresponds to approximately a half of the average atomic spacing between an O and Al atom[200] in the amorphous phase. Figure 4.8 (a) shows an SEM image of the imaged device, and the region that was extracted by Ga-ion milling. This area was chosen as it contains what appeared to be a macroscopic defect, but was found to reside

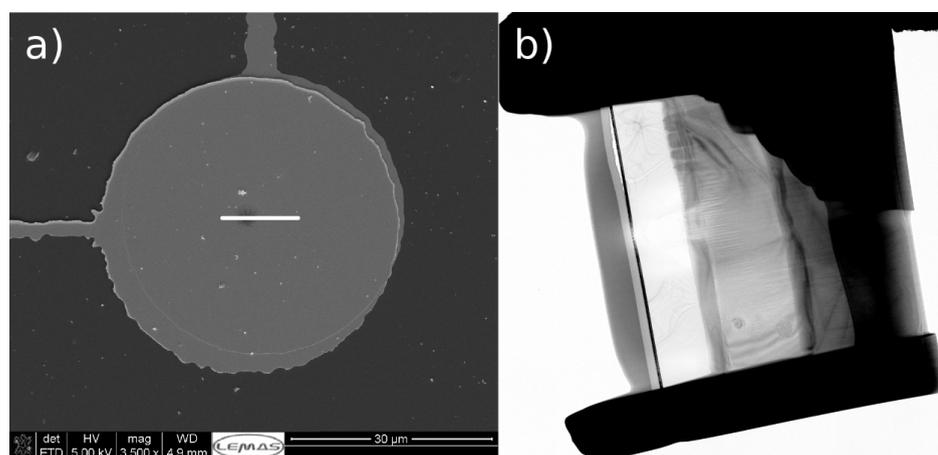


FIGURE 4.8: (a) SEM image of a MIM device with a 2 nm  $\text{Al}_2\text{O}_3$  film, taken prior to the cross section fabrication process. The region selected for analysis is shown by the white bar. (b) TEM image of the finished cross-section-sample, taken at low magnification.

on top of the MIM structure, likely corresponding to resist residue from the final sawing process. Figure 4.8 (b) shows a TEM image of the resulting device cross section. The images obtained were ultra high resolution 4096 pixel by 4096 pixel raw images in a dm4 format, converted to tiff images using Gatan image processing software available online. The image file contains an extensive information section, including the TEM system parameters used and the scale information. The images were analysed both in the Gatan suite and in the open-source program ImageJ, which features extensive image processing capabilities.

Figure 4.9 (a) shows a TEM image containing all the materials in the MIM structure. The contrast between each material is mostly due to the atomic mass of the elements present in the imaging region, such that the higher atomic mass materials appear darker. The thickness of the electrodes is measured to be very close to the targeted thickness value, between 30 nm and 35 nm for the bottom electrode, and around 60 nm for the top electrode. The chromium adhesion layers are clearly visible, and are measured to be very close to 5 nm in thickness, as intended. The interface between the gold bottom electrode surface and the  $\text{Al}_2\text{O}_3$  layer is visibly sharp due to the high contrast between the two layers. The  $\text{Al}_2\text{O}_3$  layer shows the expected conformal, homogeneous and uniform structure, and measurements of the thickness are very close to 2 nm, well within the 0.1 nm error of the system. Figures 4.9 (c) and (d) show TEM images of two different regions at a high magnification. The polycrystalline structure of both the chromium and gold layers is made clear in these images. The gold layer clearly shows many crystal planes, and the grain structure appears to be columnar. The grain size cannot be seen clearly, however there appears to be diffuse boundaries that run through of the bottom gold layer, as seen to the left in Figure 4.9 (c) where a grain boundary can be seen at the Au- $\text{Al}_2\text{O}_3$  interface, and in the bottom and top gold layers in Figure 4.9 (a). The Au- $\text{Al}_2\text{O}_3$  interface is clearly very sharp, as there is a clear departure from crystallinity and a strong contrast between the materials. The Cr- $\text{Al}_2\text{O}_3$  is less clear but appears to be equally sharp in most regions. In Figure 4.9 (d) there appears to be amorphous Cr regions at the interface with  $\text{Al}_2\text{O}_3$ , however this could be due to curvature of the structure. As the Au surface is not perfectly flat, and the grains appear to curve outwards, there are regions of materials that overlap each other in the images. In these regions the  $\text{Al}_2\text{O}_3$  film is not easily discerned from other materials, particularly Cr, due to the lower contrast between the materials. In some cases, if the overlapping is small, sub-nanometer scale fluctuations of the gold surface can be observed, which are discussed in Chapter 6.

Figure 4.9 (b) shows 2D fast-fourier-transforms (FFTs) of each region of Figure 4.9 (d), which highlights the polycrystallinity of the Au and Cr layers, and the highly amorphous structure of the  $\text{Al}_2\text{O}_3$  layer. The dotted rings denote certain periods of the crystal structure, where a single point on the ring specifies the direction the plane is

repeating. The signal intensities of the Au and Cr layers are concentrated around each ring, showing that there are crystal planes present in many different orientations. Indeed single-crystal regions can be found by reducing the area used in the FFT calculation, which shows a series of points whose symmetry about the origin is tied to the orientation of the unit cell of the structure relative to the electron beam. The absence of periodic structure in the  $\text{Al}_2\text{O}_3$  layer is also highlighted, where the FFT results in a spherical probability density.

Figure 4.10 (a) summarises the measurements of the  $\text{Al}_2\text{O}_3$  film thickness made using AFM and TEM, which shows the growth is on average linear to within a single ALD cycle ( $\approx 0.1$  nm). The conformal, uniform and linear growth, and the amorphous structure of the  $\text{Al}_2\text{O}_3$  film is consistent with extensive literature [98, 168]. The growth on gold

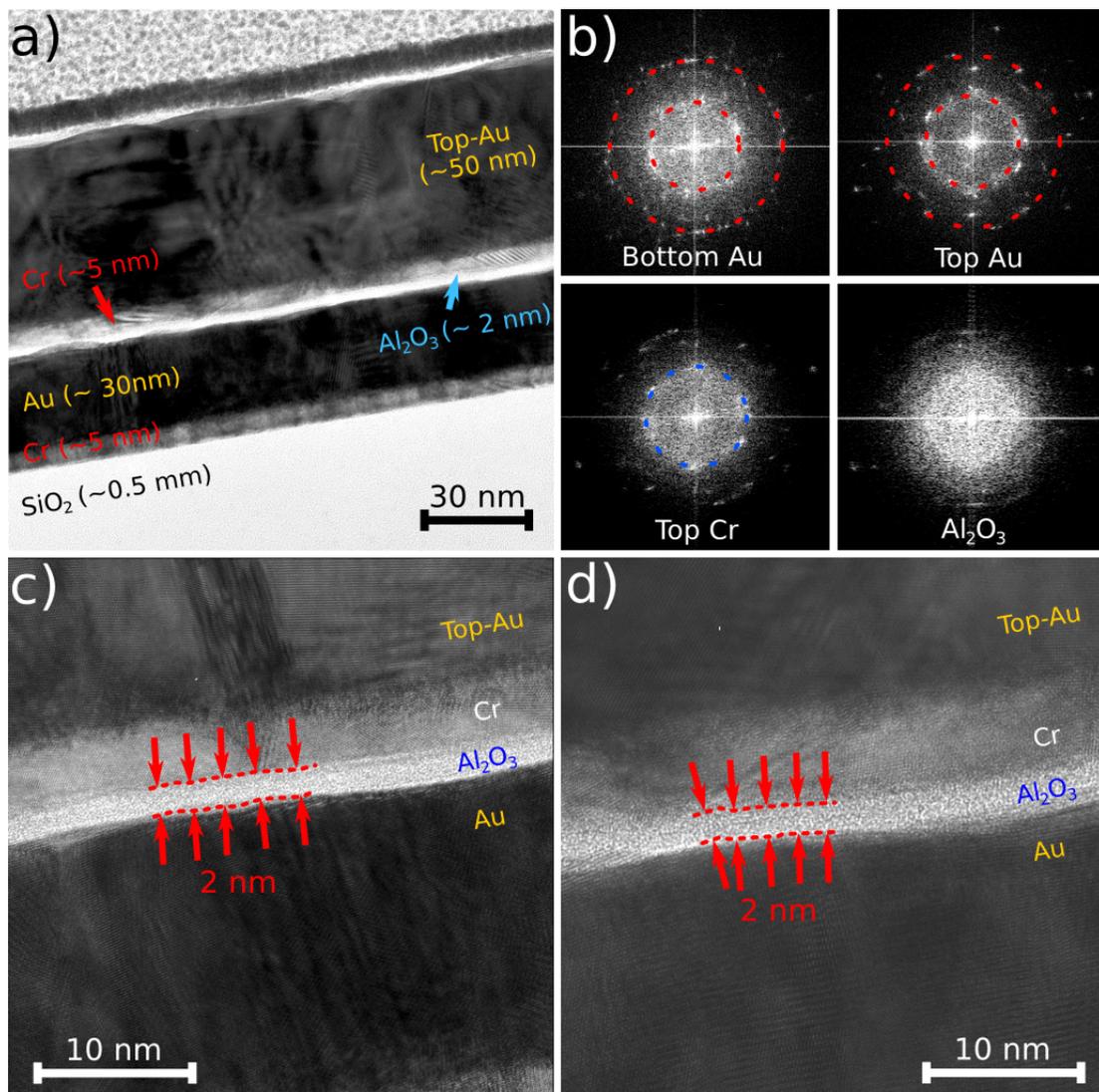


FIGURE 4.9: (a) TEM image of the MIM structure, showing all the materials. (b) FFT transformations of each material region in image (a). (c) and (d), TEM images showing the interfaces with the  $\text{Al}_2\text{O}_3$  layer in detail.

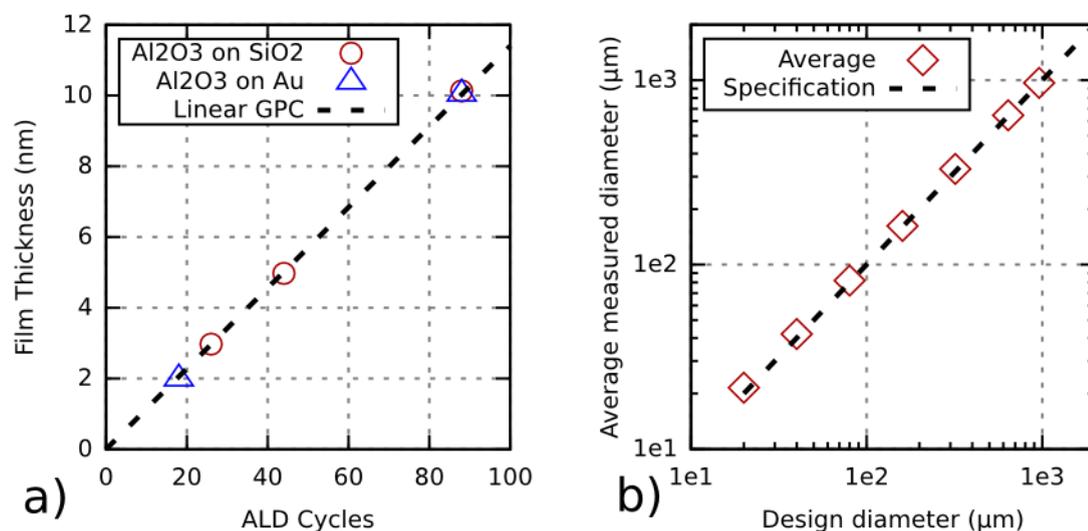


FIGURE 4.10: (a) Summary of the average values of the thickness of Al<sub>2</sub>O<sub>3</sub> measured on SiO<sub>2</sub> and Au surfaces throughout the fabrication work. (b) Summary of the average values of the electrode diameters obtained throughout the fabrication work.

is linear, and therefore there must have been sufficient reactive sites for saturation of the surface by TMA molecules to occur, leading to linear growth. Studies suggest that atomic oxygen can be adsorbed preferentially onto different planes of an Au surface and remain stable[201, 202]. As the bottom electrode surface is disordered, presenting many different crystal planes within a small area, atmospheric oxygen and plasma oxygen radicals may readily adsorb onto the surface. The presence of a finite amount of light hydrocarbons in the atmosphere and the ALD chamber may provide a pathway to the creation of OH groups on the Au surface, through combustion-like reactions that release CO<sub>2</sub>.

Figure 4.10 (b) shows the average of the measurements of the electrode diameters. The smallest electrodes with a 20 μm diameter were at most 5% in excess of the design value, and this percentage is halved for each increasing diameter, and becomes negligible for devices with a diameter greater than 160 μm. The excess of the diameter on average is due to the long exposures used in the patterning of the LOR-3A/S1813 and PMMA/S1813 bilayers, required to adequately expose the small connecting wires. This led to the broadening of the electrode disks during development in MF-319, which is also required to be long enough to resolve the small wires. The PMMA/S1813 bilayer hardening step caused many of the smallest wires connecting the 20 μm diameter devices not to be resolved, and thus very few of these devices were measurable.

### 4.3.3 AFM bottom electrode characterisation

The surface features of the Au bottom metal were analysed by AFM, to corroborate the results obtained from the TEM images. Figure 4.11 (a) shows a high resolution image of an Au layer approximately 30 nm thick. The scan was performed at low speed and with a high tapping force to track the surface as accurately as possible. A small area was used to maximize the spatial resolution of the surface, and then was scanned a number of times after the main image was obtained, to ensure all the features of the surface stable on the surface. The image was analysed statistically using Gwyddion. A clear granular structure of the surface can be observed, with an average grain size between 30 nm and 50 nm. Figure 4.11 (b) shows the slope intensity image corresponding to the image in (a), which is calculated using the derivative of the height as a function of position on the surface. This highlights the boundaries between grains. The small values of the slope of the features in image (a) is consistent with the TEM images, where a small wave-like background fluctuation can be clearly observed. Figure 4.11 (c) shows a plot of the trace indicated by the white solid line in image (a), which crosses through

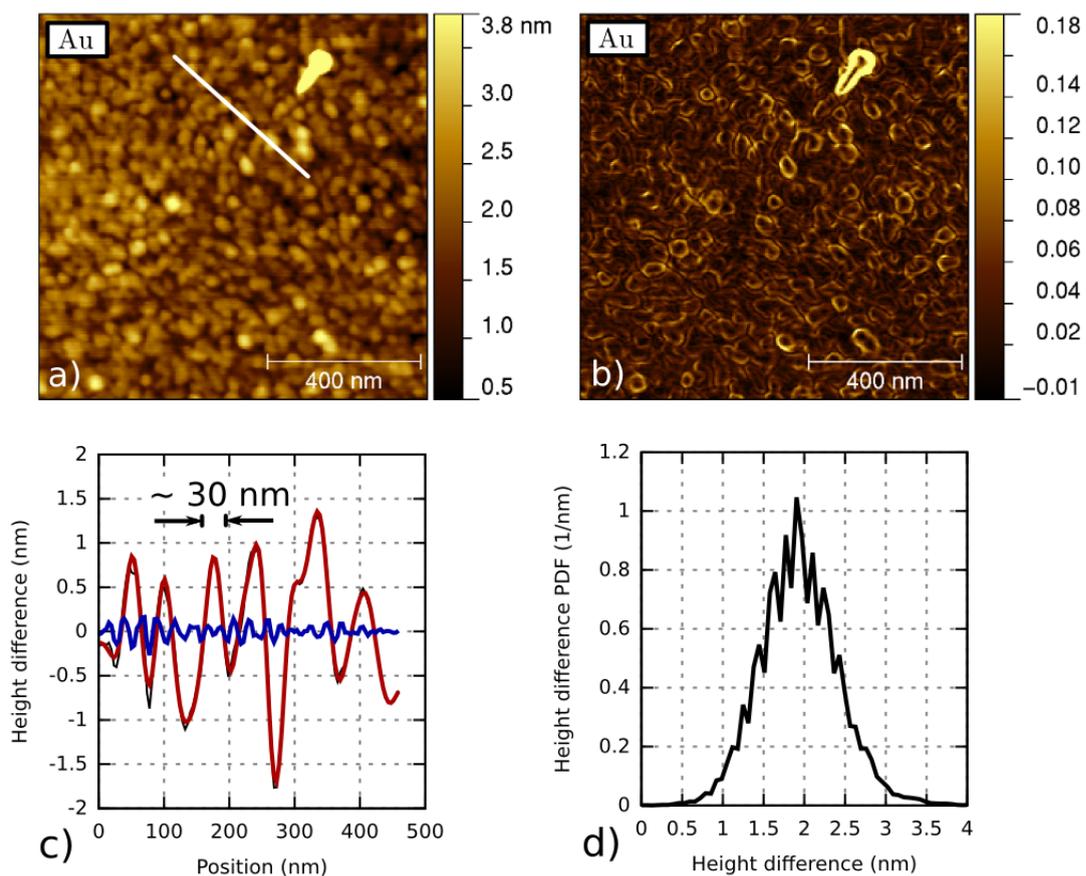


FIGURE 4.11: (a) High resolution image of an Au surface. (b) Slope of the topography of image (a). (c) Height fluctuation along the trace indicated by the solid white line in image (a). (d) Probability distribution function of the height fluctuations.

the middle of certain grains. The height amplitude curve can be decomposed into a noise contribution, shown in blue, and a smooth contribution shown in red, where the noise contribution represents the residual from the fitting to a high order polynomial. The noise contribution could be due to sub-atomic fluctuations of the surface, such as few-atom terraces or small protrusions. However these fluctuations are on the order of the resolution limit of the AFM system, and therefore can not be directly interpreted. The grain sizes estimated from the width of the peaks are on the order of 30 nm to 50 nm in diameter. The maximum peak to peak amplitude of the surface in the whole image is found to be less than 4 nm, ignoring the large feature in the top right of panel (a). This is highlighted by considering the probability distribution of the height fluctuations, shown in Figure 4.11 (d), which yields a Gaussian distribution centred at 2 nm, with a spread at half-maximum of about 1 nm. The results obtained from the TEM images and such high resolution AFM images of the Au surface show that the roughness is small on the scale of the electrodes dimensions, but not on the scale of the thinnest  $\text{Al}_2\text{O}_3$  films, which due to their conformal structure, introduce bends in the junction. Since the Au bottom electrode does not oxidise, the Au surface topography is preserved during the growth of the  $\text{Al}_2\text{O}_3$  layer, resulting in a sharp interface. The topography of the  $\text{Al}_2\text{O}_3$  surface is then a mould for the surface of the top electrode metal, which, due to the conformal nature of the growth, provides a smoothed version of the Au topography, which has been termed ‘asymmetric’ roughness[203, 204].

It is well known that bends and sharp features of metallic surface lead to the local enhancement of the electric field surrounding those regions. As the growth is uniform and conformal, as evidenced by the AFM and TEM measurements, the effect of bends and sharp features can be effectively approximated in a planar geometrical model, allowing the use of finite-element methods to solve the potential and derive the local electric field. This is discussed in detail in Chapter 6. From the data in the AFM image, each (single-channel) pixel can be divided into a series of points, which when connected to points of other pixels define facets with a given orientation, slope and surface area. This method is implemented in Gwyddion and was used to estimate the probability distributions of the slopes and inclinations angles of these facets. Figure 4.12 shows a plot of the slope probability distribution function (PDF) (a) and the inclination angle PDF (b). The small average values and the narrow distribution of the slope and inclination angles are consistent with the morphology observed in the TEM images. The contribution to the total surface area due to the topography in Figure 4.11 (a), estimated from the sum of the surface area of each facet, is approximately 1%, which is a negligibly small value. Nonetheless it is important to consider the effects of field enhancement, which are shown to be considerable in the interpretation of the tunnel current in the present work. The tail in the distributions suggests that small numbers of, and thus well isolated, sharp features must exist on the surface. The sharp spike seen in the top right of Figure

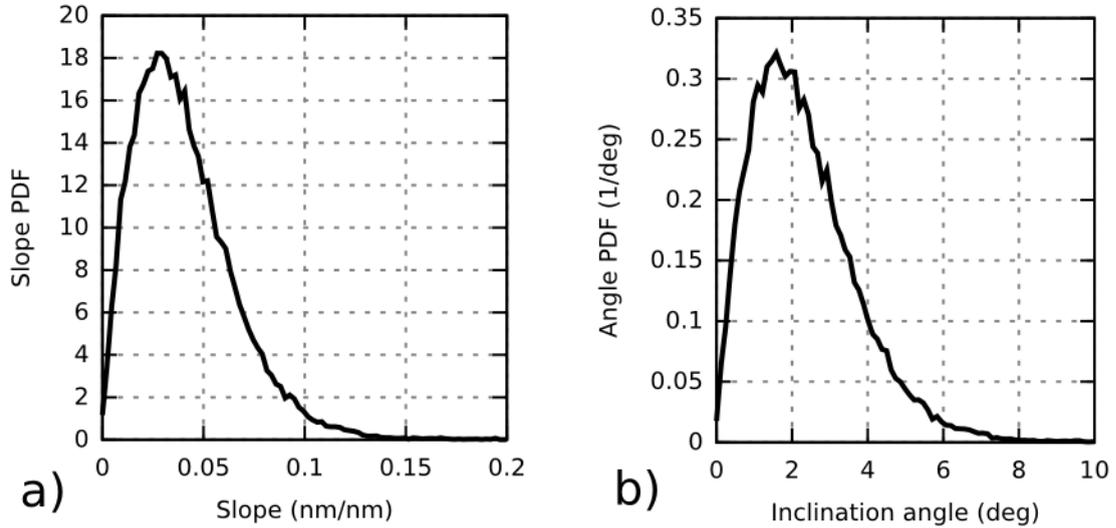


FIGURE 4.12: Probability distributions of the slope (a) and inclination angle (b) present in the topography of Figure 4.11 (a).

4.11 (a) however is a comparatively irregular fluctuation of the surface that could cause significant field enhancement from the bottom electrode. Fluctuations such as these with a slope or inclination angle much greater than the mean of the distribution are not represented well due to the small area used. A larger area is likely to include a greater number of such irregular fluctuations, which could produce a second distribution with a higher mean than the general surface topography distribution. This is limited by the spatial resolution required to create an accurate representation of the topography of the surface. The local field enhancement is not expected to play a significant role in the capacitance of the devices, as long as the response of the dielectric remains proportional to the electric field. The C–V measurements presented in Chapter 5 show that a significant externally applied electric field is required to modulate the permittivity of the material. The increased area due to the non-planar topography of each surface introduces a negligibly small contribution to the capacitance if the parallel plate model approximation is retained, which is wholly applicable to the slow fluctuations of the topography on a larger scale than the thickness of the  $\text{Al}_2\text{O}_3$  layer.

#### 4.4 Compositional characterisation

Although the evidence presented so far suggests the materials in the MIM structures are of high quality and show many of the properties expected of them, EDX/STEM imaging was performed in a few small regions of the TEM cross section sample to confirm the atomic composition is as intended to a suitable spatial resolution (approximately 5 Å). Figure 4.13 shows a STEM image (HAADF detector) and corresponding EDX data,

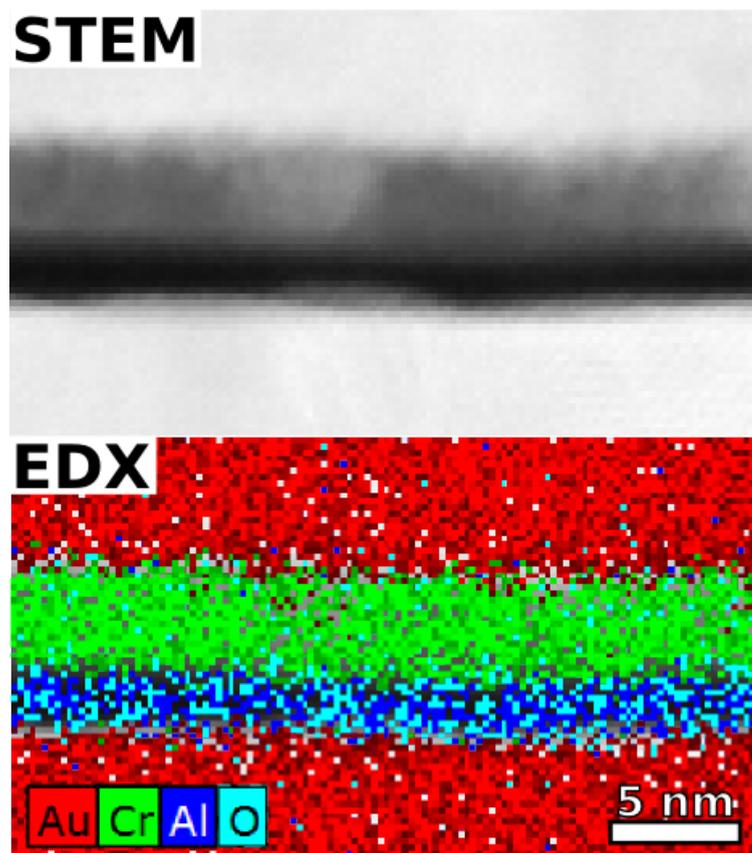


FIGURE 4.13: STEM/EDX mapping of the composition of the MIM structure, showing all the elements present in significant quantities.

showing a high homogeneity in the composition between each layer. The EDX signal is strongly dependent on the mass of the element, and limited in spatial resolution, and thus the aluminium and oxygen signals are particularly weak within the MIM structure around the  $\text{Al}_2\text{O}_3$  layer. The gold and chromium signals are very strong, indicating the structure is dense, as confirmed by the polycrystalline structure observed in the TEM images. The  $\text{Al}_2\text{O}_3$  layer also appears highly uniform, although ultra small features could not be resolved in this mode. The oxygen signal was noisy and appears prominently not only in the  $\text{Al}_2\text{O}_3$  region, but also in the entire Cr layer. This is likely due to the significant overlap of the O-K peak centred at 532 eV and the edge of the Cr-L<sub>3</sub> peak centred at 575 eV that has a delayed maximum (EELS Atlas data[188]), leading to false signals for oxygen in the Cr layer. We note that while it is conceivable that oxygen may diffuse from the  $\text{Al}_2\text{O}_3$  into the Cr layer, this would result in a visible change of the Cr structure predominately at the interface, which is inconsistent with the homogeneous distribution of O in the Cr layer as suggested by the EDX results. However, some regions of the Cr interface with  $\text{Al}_2\text{O}_3$  do show an amorphous structure, as noted previously in Figure 4.9 (d), which may be formed, for example, due to the low initial temperature of the substrate in the electron-beam evaporator chamber and the high surface energy

of the ALD film due to residual stress[165, 205] amongst other effects that influence ordering of deposited atoms on a substrate.

From the other STEM/EDX images obtained, small regions where gold from the top electrode appears deep in the chromium layer can be observed. This can be seen in Figure 4.9 (a), where the top Cr region appears darker in certain regions. This suggests that the work function in these regions may vary, due to the varying composition of the apparent alloy. The work function corresponds to the energy required to ionise an electron at the Fermi energy in the metal, and is inherently a surface property that is dependent on the atomic and electronic structure of the metal. This is discussed further in Section 6.2. In the chromium regions that appear to have an amorphous structure, the work function may also vary due to the departure from crystallinity, which affects the electronic properties. Variations in the work function are likely to be more important for the Au bottom electrode however, due to the presence of different crystal faces extended over larger areas, although the differences are expected to be small and on the order of 10 meV[206].

## 4.5 Device summary and conclusion

A series of batches of devices were fabricated using the process described in the first section. Table 4.2 summarises the performed fabrication runs, and the yield, calculated by considering devices that showed neither a short nor an open circuit, and had an interpretable electronic response shortly after fabrication completion. All of these devices showed highly consistent behaviour as a function of the geometry. In the devices for which a tunneling current could be measured, there were no significant deviations unless the devices were permanently damaged by application of high voltages near breakdown and high temperatures, which generally led to short circuits or devices with a significantly larger and qualitatively different tunnel current. The first Au-Al<sub>2</sub>O<sub>3</sub>-Au batch suffered from a poor yield, and some devices showed a small dielectric loss which was attributed to devices that were contaminated during fabrication of the top electrode layer. This is discussed in more detail in Section 5.2.1.

A wide range of diameters and thicknesses were used to measure the AC response properties of the Al<sub>2</sub>O<sub>3</sub> film as function of these parameters, and the devices with

Date	Junction	Al <sub>2</sub> O <sub>3</sub> thick.	E Diam.	Yield
19th Feb 2015	Au-Al <sub>2</sub> O <sub>3</sub> -Cr	2,3,5,10 nm	40,80,160 $\mu\text{m}$	36% of 67 (13 chips)
22nd Jun 2015	Au-Al <sub>2</sub> O <sub>3</sub> -Au	2,3,5,10,20 nm	40,80,160 $\mu\text{m}$	25% of 67 (11 chips)
7th Jun 2016	Au-Al <sub>2</sub> O <sub>3</sub> -Au	1.4,2,3 nm	20,40,80,160 $\mu\text{m}$	55% of 73 (12 chips)
22nd Mar 2017	Au-Al <sub>2</sub> O <sub>3</sub> -Cr	10,20,30 nm	160,320,640,960 $\mu\text{m}$	76% of 78 (12 chips)

TABLE 4.2: Summary of the fabricated devices, where two different MIM structures were created, using Cr and Au as top electrodes.

the thinnest films were used to study the tunneling properties of the MIM structure. Chromium and gold were used as top electrodes, to form MIM junctions with different properties. The fabrication duration was on average two and a half weeks, and the measurements were performed immediately on completion of the fabrication, beginning with capacitance measurements, followed by room temperature I–V measurements. The low yield values are due primarily to devices showing an open circuit, and relatively few short circuits were detected. The open circuits occurred randomly across all types of devices, and is thought to be caused by the poor undercut of the PMMA/S1813, leaving raised metal edges that may prevent a good contact between the top electrode and the corresponding contact pad. Short circuits were also distributed across the different devices, however they did occur more commonly in the thinnest devices with films less than 10 nm. These short circuits are possibly due to damage incurred during fabrication, or large particles trapped under the comparatively thin  $\text{Al}_2\text{O}_3$  film, which are subsequently desorbed from the surface in the processes immediately following the ALD step, leaving the bottom electrode exposed in small regions.

It was shown that the growth of  $\text{Al}_2\text{O}_3$  by ALD is proportional to the number of cycles on both  $\text{SiO}_2$  and Au surfaces, and that there is therefore no significant nucleation delay, allowing highly reproducible and uniform thickness films to be deposited. Due to the consistency of the ALD and photolithography processes, accurate values of  $t_J$  and  $A_J$  could be obtained for each device, allowing them to be used as parameters in the modelling of the electronic properties. The use of ALD therefore provides a significant advantage over native oxidation processes, since the thickness can be accurately known, given prior characterisation of the growth properties on the substrate concerned. The characteristic surface geometry of the bottom electrode Au layer was assessed and compared with the results obtained by TEM imaging, and showed a good consistency. The geometrical errors and immediate implications associated with the surface topography were discussed.

---

# Measurement Results and Analysis

---

## 5.1 Introduction

The electronic characterisations performed on the devices described in Chapter 4 are detailed in this chapter. The two critical properties for the design of metallic SET systems are the dielectric constant of the insulating thin film, and the tunneling properties of the junction. Through low-frequency measurements the static dielectric constant of the film was determined using standard circuit models of the impedance. Using C–V measurements it was found that the permittivity of all MIM devices increases with applied voltage, consistent with the behaviour of ceramic insulators. It is shown that below a critical thickness the dielectric constant decreases with thickness and appears to be roughly independent of the type of the device, i.e. the top electrode metal Cr (type-A) and Au (type-B). DC measurements of the I–V characteristics of a range of thin film thickness devices were performed to determine the dependence of the tunnel current on the geometrical parameters of the junctions. It is shown that in general the devices have a response consistent with the geometry, particularly for the thinnest films.

## 5.2 Capacitance measurements

The measurement of the capacitance of each device was performed initially using a lock-in system, which was limited in bandwidth. The LCR system was not available until much later in the project, which allowed for a much greater flexibility in the measurements, as described in Section 3.4.2. The first batch of Au-Al<sub>2</sub>O<sub>3</sub>-Cr devices, referred to as type-A, featuring thin films were initially only measured using the lock-in system. Many of these devices were subsequently tested at high voltages and high temperatures following these initial measurements, which often destroyed the devices hence leading to only a small number of remaining working devices by the time the LCR system was available. Nonetheless, some of these devices were measured again with the LCR meter at a later date. The higher fabrication yield of the Au-Al<sub>2</sub>O<sub>3</sub>-Au (type-B) batch provided an extensive range of ultra-thin film (3 nm, 2 nm and 1.4 nm) MIM capacitors/diodes featuring a wide range of surface areas. In general, the quantity of interest in this section is the impedance of the device as a function of the material properties and the device geometry  $Z_J(\omega, V_a, t_J, A_J)$ , which is treated as a ‘black box’ in this expression, and where  $\omega$  is the angular frequency of the AC excitation,  $V_a$  an

applied bias (used in the C–V measurements). The capacitance is determined by fitting the unknown impedance to an equivalent circuit model. The permittivity of the film can then be extracted using the junction geometrical parameters  $t_J$  and  $A_J$  with the parallel plate model of the capacitance (Equation 4.2). The low frequency measurements of  $Z_J(\omega, t_J, A_J)$  are presented in the first subsection, and in the following subsection the C–V measurements of a few selected devices (160  $\mu\text{m}$  diameter, and 1.4, 2, 3 thickness for type B and 10, 20 and 30 nm thickness for type A) are presented, and compared with reports in the literature on the subject.

### 5.2.1 Low-frequency measurements

Figure 5.1 shows the impedances of a series of type-A devices measured using the lock-in system, between 1 Hz and 1 kHz. The impedance of each device was determined from the measurement of the r.m.s. current detected in channels X (in-phase) and Y (quadrature), using the circuit represented in Figure 3.14. Using the complex number representation, the applied voltage, and the complex current measured can be written as

$$\underline{v}(\omega) = \sqrt{2}v_{rms}e^{j\omega t}e^{j\phi_v} \quad \text{and} \quad (5.1)$$

$$\underline{i}(\omega) = \sqrt{2}i_{rms}e^{j\omega t}e^{j\phi_i}, \quad (5.2)$$

where  $\phi_{v,i}$  are the phase values of the voltage and current, and  $j$  is the complex number. The complex voltage expression can be obtained by measuring the amplitude and phase of the voltage source, which was only required once as it did not vary between experiments. In each case, the r.m.s. AC voltage amplitude was chosen to be sufficiently small to avoid causing damage to the devices due to excessive electric field strengths and to avoid a non-linear dielectric response as a result. The C–V measurements discussed in the next subsection show that the MIM device dielectric is well within the linear response regime at voltages below 100 mV r.m.s. for 10 nm thick films. It was found that the measured impedance indeed varies very little below this value of r.m.s. AC voltages. The impedance is determined using the complex voltage and current expressions as

$$\underline{Z}_J = \frac{v_{rms}}{i_{rms}}e^{j(\phi_v-\phi_i)} = |Z_J|e^{j\phi}. \quad (5.3)$$

From this expression it can be seen that when the phase of the voltage source and measured current are the same then the load is purely real, and otherwise has a finite imaginary part. As can be seen from the impedance and phase curves in Figure 5.1, devices with films greater than 2 nm show an impedance that decreases with increasing frequency over the whole frequency range (1 Hz to 1 kHz), which is the expected behaviour of a single capacitor. For the 2 nm films, the impedance plateau at low frequencies suggests there is a significant real conductance across the MIM structure,

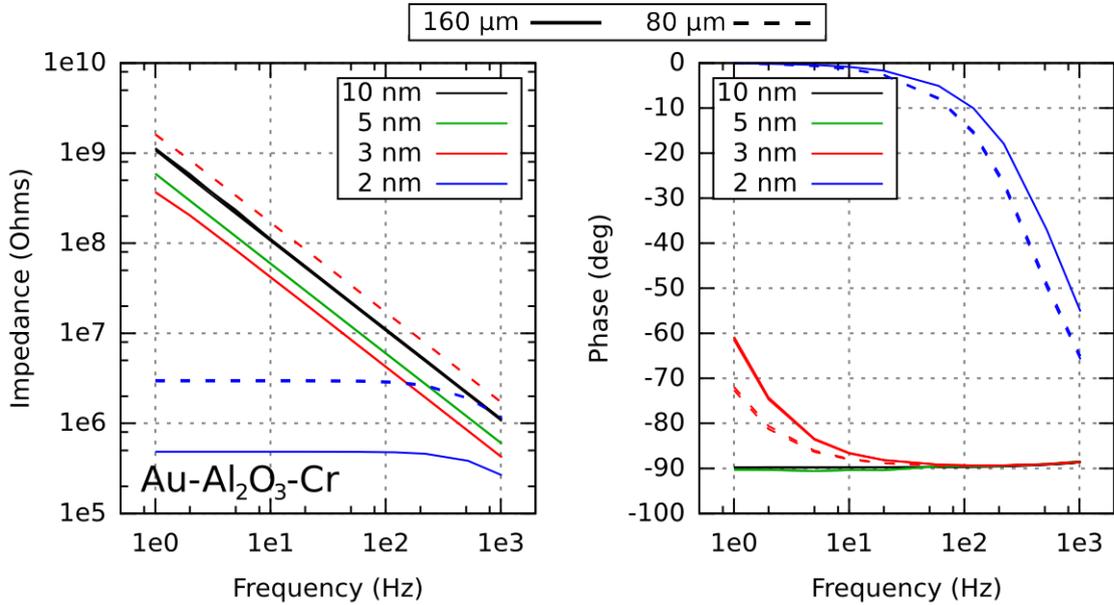


FIGURE 5.1: Plot of the magnitude (left) and phase (right) of the measured impedance of a series of type-A devices, of varying thickness and surface area, using the lock-in system. The top legend shows the electrode diameters, and the inset legends show the film thicknesses used.

evidenced by the low phase values in this range. In the purely capacitive region of the impedance, the phase values are very close to  $-90^\circ$ , which suggests there are minimal power losses in the dielectric. The dielectric loss, or absorption, is discussed in more detail in Section 6.4.

Figure 5.2 shows the impedance of a large number of type-B devices, measured using the LCR system between 50 Hz and 200 kHz. The impedance is determined internally from the measurement of the voltage across the device and current measured in the four-probe system, represented in Figure 3.15, where the bias voltage is set to zero. The minimum AC voltage of 50 mV r.m.s. was used in all device measurements. The broader frequency range provides a much broader range to measure the decreasing impedance of thinner-film and larger-area devices, allowing the dielectric properties to be determined more precisely. The plot in Figure 5.2 shows two devices for each  $(t_J, A_J)$  combination, except for the 1.4 nm thickness device with a  $160 \mu\text{m}$  diameter, as only one such device was successfully fabricated. The dependence of the impedance on the geometrical parameters  $t_J$  and  $A_J$  is much clearer here due to the significant number of devices available, where the spacing between the curves of equal  $t_J$  in the y-axis are very similar, as expected from Equation 4.2. There is also a roughly even spacing between the curves of equal  $A_J$ , showing that the capacitance does appear to increase with decreasing thickness. The plateau of the impedance for the 1.4 nm curves at low frequencies suggests a significant real conductance shunts the capacitive impedance, similarly to the 2 nm devices in Figure 5.1. In addition, the spacing between the plateaus for the 1.4 nm

devices is less consistent than in the capacitive regions, which suggests the tunneling conductance is more susceptible to inconsistencies, as expected.

Collectively these results suggest a simple model of the MIM structure can be used to extract the capacitance. There are clear contributions from a capacitive impedance in all the measurements presented, and in the thinnest devices a resistance parallel to this capacitance dominates at low frequencies. Although a second plateau at high frequencies cannot be seen in these measurements, a small series resistance on the order of  $100 \Omega$  exists due to the small wires connecting the electrodes, and was measured primarily in the short circuited devices in each batch. The impedance can then be written as

$$\underline{Z}_J(\omega) = R_S + \frac{R_J}{1 + j\omega R_J C_J(t_J, A_J)} \quad (5.4)$$

where  $R_S$  is the small series resistance,  $R_J$  is the parallel resistance of the junction, and  $C_J(t_J, A_J)$  is the junction capacitance according to Equation 4.2. This impedance can then be fitted to the measured impedance, using  $R_S$ ,  $R_J$  and  $C_J(t_J, A_J)$  as fitting parameters. As there are distinct regions in the impedance spectrum, Equation 5.4 provides a unique solution  $\{R_S, R_J, C_J(t_J, A_J)\}$  with a high certainty. The permittivity of the film can then be extracted using Equation 4.2, which under the current approximations is considered to be constant at all measurement frequencies. As the phase is very close to  $-90^\circ$  and varies little with frequency in the purely capacitive regions, this

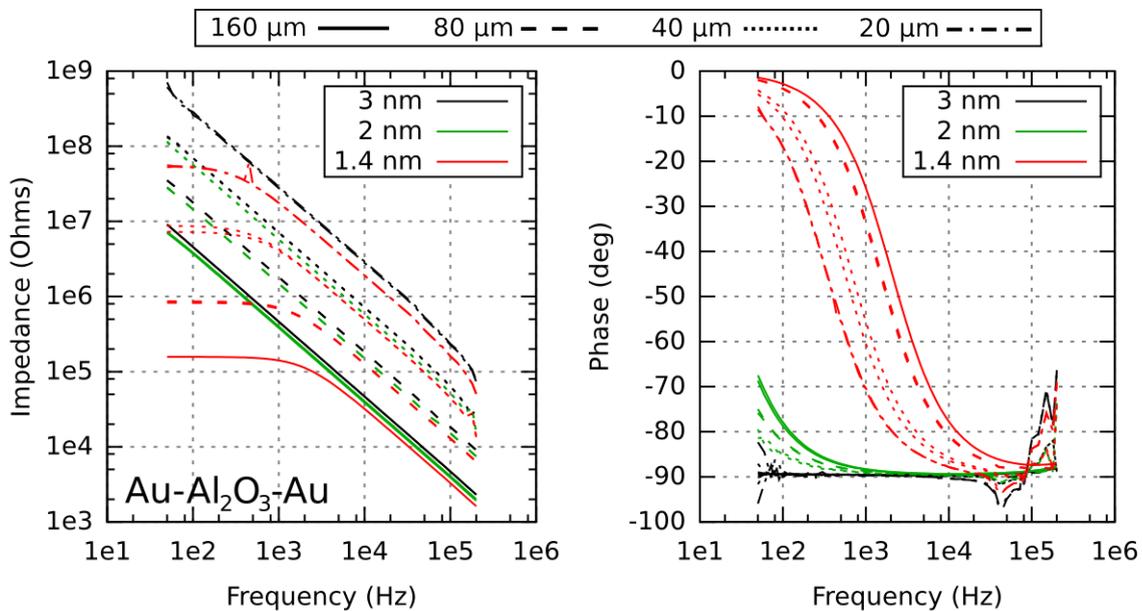


FIGURE 5.2: Plot of the magnitude (left) and phase (right) of the measured impedance of a series of type-B devices, of varying thickness and surface area, using the LCR system. In the measurements shown there are two devices for each  $(t_J, A_J)$  combination, except for the 1.4 nm thickness device with a  $160 \mu\text{m}$  diameter where only one such device was successfully fabricated.

is in fact a good approximation. The resistance  $R_J$  can be associated with the tunnel conductance of the barrier near equilibrium, termed the zero-bias conductance, since the amplitude of the AC voltage is small such that the in-phase response is approximately linear. The I–V curves of the thin-film devices are presented in the next section, and do show an approximate linear behaviour at such low applied voltages, which gives a nearly constant  $R_J$  in this range.

Figure 5.4 (a) shows a diagram of the equivalent circuit, illustrating the physical correspondence with the circuit elements  $R_S$ ,  $R_J$  and  $C_J$ . Figure 5.4 (b)-(d) shows the results of fitting to Equation 5.4, where three devices with an equal electrode diameter  $D_J = 160 \mu\text{m}$  and  $\text{Al}_2\text{O}_3$  film thicknesses of 1.4 nm, 2 nm and 3 nm are shown. The crosses denote the measured points, and the solid black line is the result of the model with the best fit parameters, shown within each plot. The quality of the fit was better than  $R^2 = 0.9999$  in each case, when each point is weighted by their relative magnitude in the fitting routine. In general, the response of many devices of different thickness and surface area are well described at least qualitatively by this simple circuit model, and as such the same model was used for each device. In most cases the  $R_S$  was less than  $100 \Omega$ , and in some rare cases reached up to  $10 \text{ k}\Omega$  in some devices with thicker films. Nonetheless the simple circuit model used permits reliable extraction of the capacitance even in those rare cases. Figure 5.3 shows a summary of the values of  $C_J$  obtained from fitting the impedance of all working devices to the circuit model. As can be seen, the spread in the capacitance values for a given geometry is generally very small, even across different batches. There is a slight discrepancy between the type-A and type-B thin film devices, this is addressed in Section 6.4.

The resistance  $R_J$  is omitted in devices where the phase at low frequency is near  $-90^\circ$  (in general for  $t_J \geq 5 \text{ nm}$ ), since it cannot be determined to any accuracy.  $R_J$  in these cases is much larger than the capacitive impedance, which is the expected behaviour of a high quality insulating thin-film. In the first type-B batch, a few devices consistently showed a phase offset and a small frequency dependence in the capacitive impedance range, where the phase would approach, but not reach  $-90^\circ$ , and the difference increases slowly with frequency. The magnitude of the impedance in these regions is not perfectly proportional to  $(\omega C_J)^{-1}$ , which suggests that dielectric relaxation mechanisms are present in these particular devices. As this was not observed in any devices obtained from other batches, it was attributed to the fabrication process of this particular batch, which suffered due to sub-optimal cleanliness of the resist patterning steps, resulting in a low yield. Contaminating particles that lie between the top and bottom electrodes at the interface with the  $\text{Al}_2\text{O}_3$  layer may introduce a low frequency loss component in the AC response, resulting in a perceptible frequency dependence of the ‘black-box’ dielectric susceptibility[207]. In this case the capacitance itself can be expressed as a complex quantity  $\underline{C}_J(\omega, t_J, A_J)$ , which results in a small contribution to the real part of

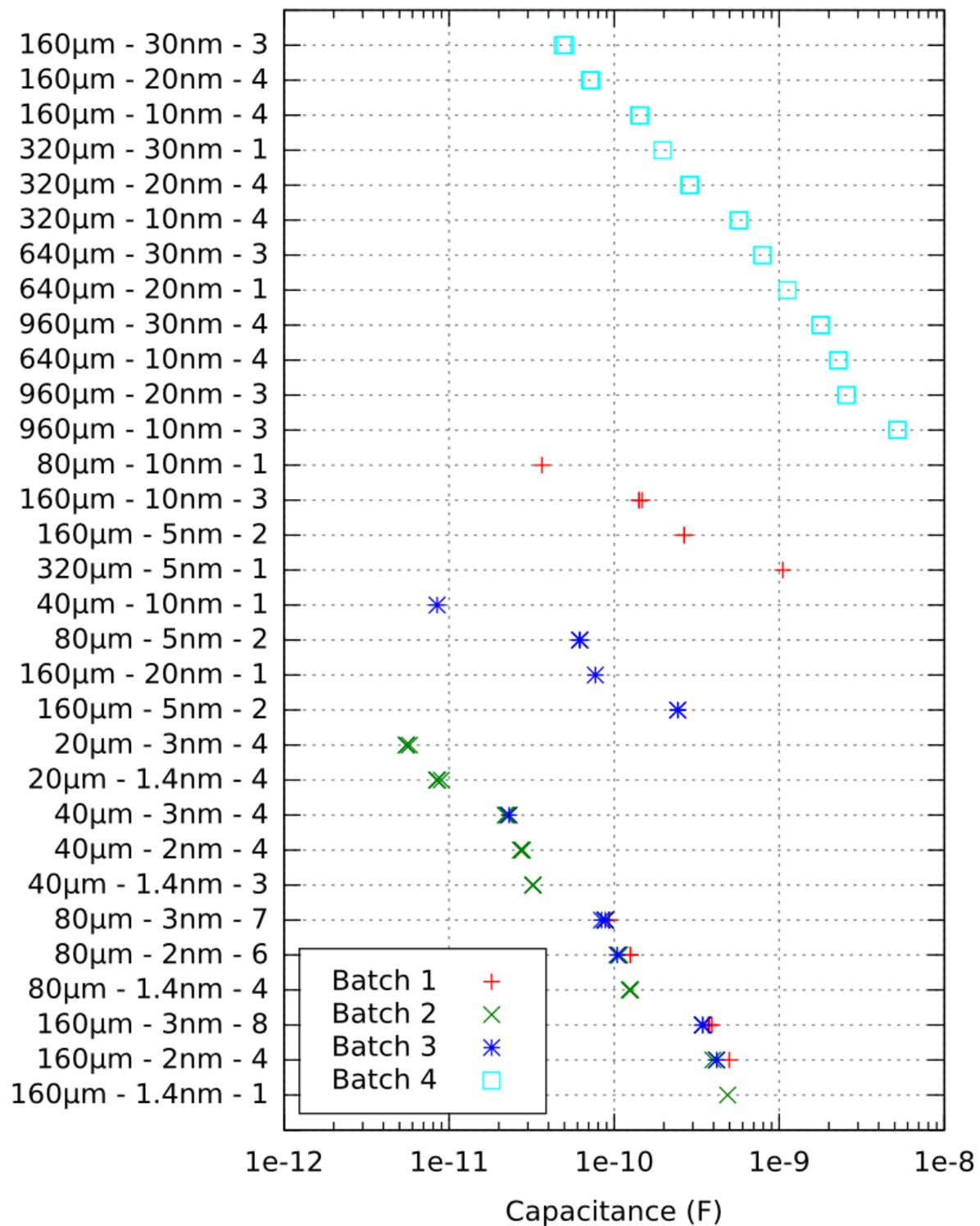


FIGURE 5.3: Summary of capacitances extracted from the impedance spectra of all the devices. The labels on the y-axis show the electrode diameter, the thickness of the Al<sub>2</sub>O<sub>3</sub> film, and the number of devices in that set.

the complex impedance in Equation 5.3[208]. This is not discussed in greater detail as these effects are avoidable when the fabrication process is maintained at a satisfactory

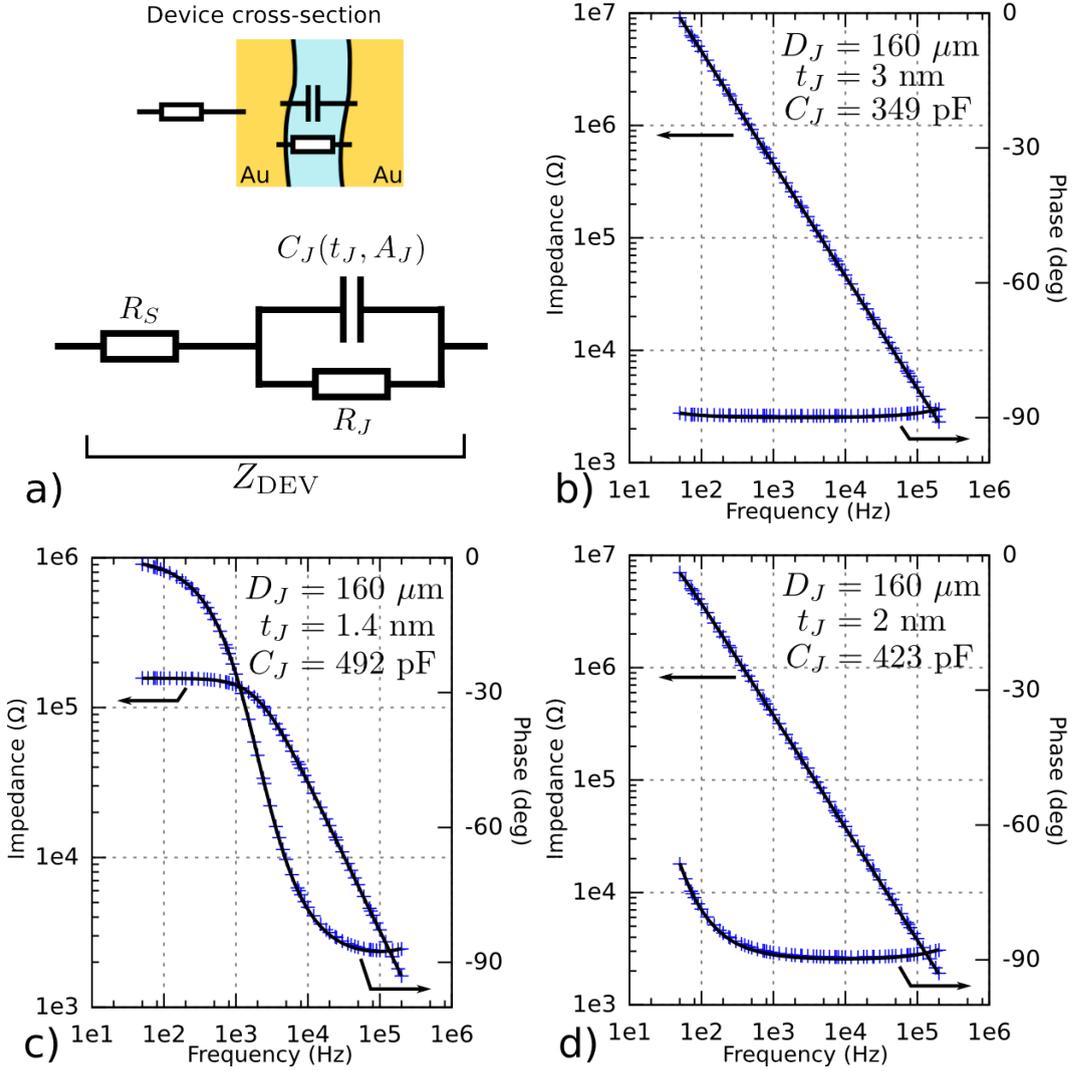


FIGURE 5.4: (a) Equivalent circuit model, where each element is associated with a certain layer in the MIM device. (b)-(d) Results of fitting the complex impedance to Equation 5.4, where type-B devices with  $D_J = 160 \mu\text{m}$  and a range of ultra-thin thickness films are shown.

cleanliness throughout each step. The devices that show this effect are not as well represented by Equation 5.4, and are thus omitted in the analysis of the capacitance.

The consistency between the geometry and the  $C_J$  values is assessed by computing the static permittivity  $\epsilon_r$  of the  $\text{Al}_2\text{O}_3$  film, according to each  $(t_J, A_J)$  combination using Equation 4.2 as

$$\epsilon_r = \frac{C_J t_J}{\epsilon_0 A_J} = \frac{4 C_J t_J}{\pi \epsilon_0 D_J^2}. \quad (5.5)$$

Figure 5.5 (a) shows the values of the permittivity obtained for each batch as a function of  $t_J$ , and the curves connect the average values obtained for each top electrode material Cr and Au. As can be seen, a remarkable consistency between the permittivity values at a given thickness and the surface area  $A_J$  is obtained for a given top electrode type,

where the scatter in the symbols at each thickness value are assumed to be entirely representative of inconsistencies in  $A_J$  between each device. The permittivity of the film is found to decrease significantly below a certain critical thickness less than 10 nm, according to the parallel plate model of the capacitance. To deduce if this value is truly representative of the permittivity of the film, the errors incurred from Equation 5.5 due to inconsistencies in the geometry must be considered. The values of  $\epsilon_r$  obtained for each device were computed using the average electrode diameters obtained from optical microscopy measurements, summarised in Figure 4.10 (b), where the smallest diameter devices were on average larger than the design value. The uncertainty in the measured diameter is  $\pm\delta D \approx \pm 0.5 \mu\text{m}$ , due to the difficulty in finding the true edges of the circular electrodes. The true diameter of the perfectly planar electrode is then expressed as  $D_J = D_M \pm \delta D$ , where  $D_M$  is the measured diameter, which takes the values shown in Figure 4.10 (b). This uncertainty then translates to an uncertainty in the calculated surface area  $\pm\delta A \approx \pm 0.5\pi D_M \delta D$ . The true (continuous) surface area is then written as

$$A_J = A_M + \delta A_N \pm \delta A, \quad (5.6)$$

where  $\delta A_N$  represents the additive contribution due to the deviation of the surface geometry from an ideal planar surface, which was estimated to be on the order of 1% of the total area in section 4.3.3 on the bottom electrode characterisation.  $A_M$  is the measured area, calculated using  $D_M$  through Equation 4.1. Using Equation 5.5, the uncertainty in the dielectric constant due the area  $A_J$  can then be expressed as

$$\pm \delta\epsilon(A_J) = \frac{C_J t_J}{\epsilon_0} \left( \frac{1}{A_M + \delta A_N \mp \delta A} - \frac{1}{A_M + \delta A_N} \right), \quad (5.7)$$

which depends on  $\delta A$ , as well as the measured area  $A_M$  and the contribution  $\delta A_N$ . The uncertainty in the  $\text{Al}_2\text{O}_3$  thickness is assumed to be on the order of  $\pm\delta t \approx \pm 0.1 \text{ nm}$  in the thin film devices ( $t_J \leq 10 \text{ nm}$ ), as assessed by the AFM and TEM measurements

$D_M$	21 $\mu\text{m}$	41 $\mu\text{m}$	81 $\mu\text{m}$	161 $\mu\text{m}$	320 $\mu\text{m}$	640 $\mu\text{m}$	960 $\mu\text{m}$
$A_M$ ( $\text{m}^2$ )	$3.5 \times 10^{-10}$	$1.3 \times 10^{-9}$	$5.2 \times 10^{-9}$	$2 \times 10^{-8}$	$8 \times 10^{-8}$	$3.2 \times 10^{-7}$	$7.2 \times 10^{-7}$
$\pm\delta A$	$\pm 5\%$	$\pm 2.5\%$	$\pm 1.25\%$	$\pm 0.63\%$	$\pm 0.31\%$	$\pm 0.16\%$	$\pm 0.1\%$
$C_J$ max.	10 pF	30 pF	120 pF	480 pF	570 pF	2.3 nF	5.2 nF
$\pm\delta\epsilon(A_J)$	$\pm 0.21$	$\pm 0.08$	$\pm 0.04$	$\pm 0.02$	$\pm 0.02$	$\pm 0.01$	-
$t_M$	1.4 nm	2 nm	3 nm	5 nm	10 nm	20 nm	30 nm
$\pm\delta t$	$\pm 7.1\%$	$\pm 5\%$	$\pm 3.33\%$	$\pm 2\%$	$\pm 1\%$	$\pm 0.5\%$	$\pm 0.33\%$
$C_J$ max.	480 pF	400 pF	340 pF	250 pF	5.2 nF	2.5 nF	1.8 nF
$\pm\delta\epsilon(t_J)$	$\pm 0.27$	$\pm 0.22$	$\pm 0.19$	$\pm 0.14$	$\pm 0.08$	$\pm 0.04$	$\pm 0.03$

TABLE 5.1: Summary of the errors associated with  $A_J$  (top section) and  $t_J$  (bottom section) values in each column. The ‘ $C_J$  maximum’ row represents the (approximate) maximum capacitance measured for the given geometrical parameters, so that the largest errors are represented for the extracted dielectric constant.

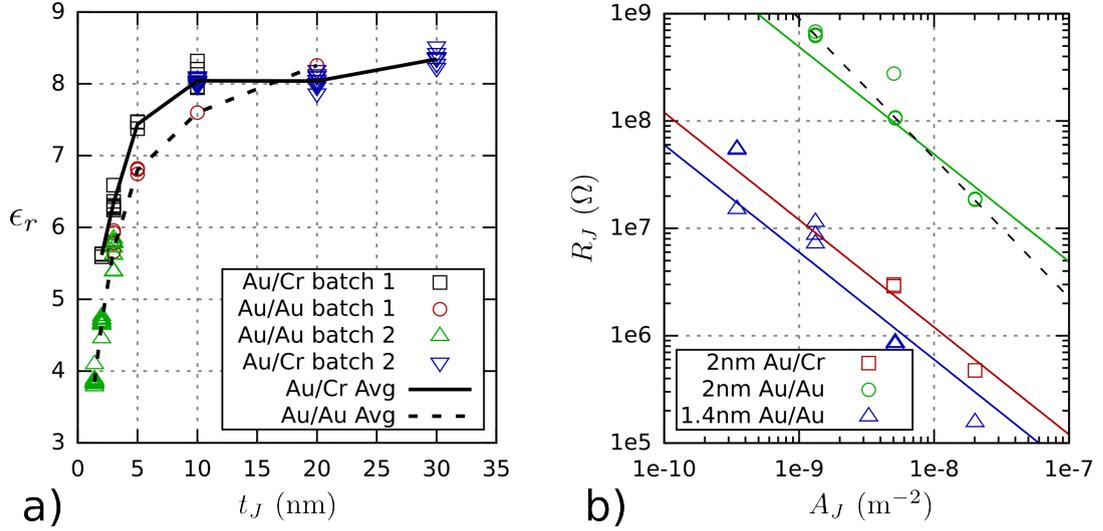


FIGURE 5.5: (a) Thin film permittivity of all measured devices from both the chromium and gold top electrode batches, extracted using Equation 4.2, plotted as a function of  $t_J$ . (b) Selected values of  $R_J$  determined from fitting the impedance to Equation 5.4, plotted as a function of  $A_J$ . The solid lines represent ideal scaling of  $R_J$  with  $A_J$ , and the dashed line represents non-linear scaling.

presented in Section 4.3.2. The uncertainty in the permittivity due to the thickness can then be expressed as

$$\pm \delta\epsilon(t_J) = \pm \delta t \frac{C_J}{\epsilon_0 A_J}, \quad (5.8)$$

which depends only on  $\delta t$ . Table 5.1 summarises the errors discussed, where the largest values for a given case are shown. As can be seen, the error in the permittivity is on the order of  $\pm 10\%$  for the devices with the smallest electrode diameters  $D_M$  and the smallest films. The combined error  $\delta\epsilon(t_J, A_J)$  can then reach on the order of  $\pm 20\%$  of the measured permittivity value. Thus the permittivity of the film appears to strongly decrease with the thickness of the film, within both extremes of the estimated errors. Plausible physical explanations for this trend are discussed in Section 4.3.2. The curves obtained for each top electrode type are roughly similar, although at small  $t_J$  values, the permittivity appears larger in the chromium top electrode devices, which is likely partly due to the difference between the electrode work functions, which results in a considerable electric field within the  $Al_2O_3$  layer that increases with decreasing thickness. This is discussed in the next subsection concerning the C–V measurements. The value of the permittivity appears to vary little as  $t_J$  is increased beyond 10 nm, and approaches a value between 8.0 and 8.5. These values are slightly smaller than reported values for  $Al_2O_3$  in general[209], usually on the order of 9, however they are consistent with similar MIM devices fabricated by ALD in a CMOS-compatible process[99], where values on the order of 8.2 were obtained. See also the references in section 3.2.3, which report similar values also for the process used to fabricate the present devices.

Figure 5.5 (b) shows a plot of selected values of the junction resistance  $R_J$  obtained from fitting Equation 5.4 to the impedance data. The value of  $R_J$  cannot easily be expressed as a function of  $t_J$  and  $A_J$  without considering the tunneling characteristics and the surface geometry in detail, which is performed in Chapter 6. The scaling of  $R_J$  with  $A_J$  in an ideal device, i.e. with planar geometry and uniform material properties, would be expected to scale linearly as

$$R_J = \frac{\rho_J(t_J)}{A_J}, \quad (5.9)$$

where  $\rho_J(t_J)$  is an empirical surface resistivity associated with electron tunneling, that sensitively depends on the thickness and energetic potential barrier of the  $\text{Al}_2\text{O}_3$  film, as discussed in Chapter 2. This ideal scaling law is represented by the solid lines in Figure 5.5 (b). As can be seen,  $R_J$  can be interpreted as changing slightly non-linearly with  $A_J$ , and as a result can be empirically described as

$$R_J = \alpha_c \frac{\rho_J(t_J)}{A_J^{1+x}}, \quad (5.10)$$

where  $\alpha_c$  is a proportionality factor and  $x$  determines the non-linearity of the scaling. This scaling law is represented by the dashed line in Figure 5.5 (b), where  $x = 0.3$ . According to either of these scaling laws, scaling  $A_J$  down to values relevant to SET systems leads to values of  $R_J$  far greater than desired. For example, Equation 5.9 with  $\rho_J = 6 \times 10^{-3} \text{ } \Omega \cdot \text{m}^2$  obtained at  $t_J = 1.4 \text{ nm}$  (solid blue line) corresponds to an empirical surface resistivity on the order  $6 \text{ G}\Omega \cdot \mu\text{m}^2$ . The values of  $R_J$  obtained for the 2 nm chromium top electrode devices are in general much smaller than the equivalent thickness gold top electrode devices, which is due in part to the smaller work function of Cr, resulting in an overall smaller potential barrier. Furthermore the values of  $R_J$  for films greater than 2 nm are highly uncertain for the type-B devices due to the limited impedance range. A limited range of ultra-thin film, type-A devices were measured, insufficient to observe a trend, and the determination of  $R_J$  was further limited due to the frequency range of the lock-in system used at the time those devices were measured. Despite the scatter and uncertainty, there clearly is scaling of the values of  $R_J$  with  $A_J$ , and the deviations from the ideal scaling law are likely due to the sensitivity of the total tunneling current to fluctuations of the geometry of the structure, and the material properties of the  $\text{Al}_2\text{O}_3$  films. The uncertainty in  $R_J$  is smallest for the type-B 1.4 nm thick film devices, which shows the closest match to ideal scaling of the resistance with area. The scaling of  $C_J$  with  $A_J$  was shown to be nearly ideal, since the derived values of  $\epsilon_r$  have a minimal spread for each thickness value. Therefore, the tunneling properties are clearly more sensitive to imperfections and irregularities between devices, since the scaling of  $R_J$  with  $A_J$  at a given thickness shows a greater deviation from the expected

scaling law.

### 5.2.2 Capacitance-voltage measurements

The behaviour of the capacitance as a function of a DC applied bias is measured using the LCR system described in Figure 3.15, where the DC bias is controlled using the voltage output of a Keithley 2400LV, which is readily controlled by a computer running Labview. A current due to both the DC voltage source and AC voltage source will exist in the branch connecting the device, which can be expressed as

$$I_{\text{DEV}}(V_a, v_a(t)) = I(V_a) + \underline{i}(v_a(t)) \quad (5.11)$$

where  $I(V_a)$  is the current in response to the applied DC voltage  $V_a$ , and  $\underline{i}(v_a(t))$  is the current in response to the applied AC voltage  $v_a(t)$ , which in general is a complex quantity. The LCR meter internally separates the AC component  $\underline{i}(v_a(t))$  from the current  $I_{\text{DEV}}$  using techniques similar to the lock-in method, and provides the additional benefit that a significant DC offset current can be effectively rejected, unlike the lock-in system available, which could not handle significant DC offsets. Additionally, the AC voltage across the device is measured close to the terminals of the device, again using a phase sensitive detector that rejects DC offsets. The LCR meter was calibrated with open and short corrections performed as detailed in section 3.4.2. The quantity measured is then the dynamical impedance of the junction in a state determined by the applied voltage  $Z_J(\omega, V_a, t_J, A_J)$ . It was found that if  $R_J$  is too small, the LCR meter can no longer resolve the AC current, as it is too small compared to the DC offset current. This was primarily an issue for the thinnest films and largest area devices and therefore imposed limits on the applied bias, which varied from device to device. In order to obtain the most accurate measurement of the capacitance possible, the largest area devices available were used for the thin film devices, which provided large  $C_J$  values and slightly larger  $\rho$  values according to Figure 5.5 (b). The frequency of the measurement was chosen such that the reactance dominates the impedance spectrum, so that the capacitance could be easily calculated from the measured impedance as a function of  $V_a$ , using

$$C_J(V_a) \approx \frac{1}{\omega Z_C(\omega, V_a)} \quad (5.12)$$

where  $Z_C$  is the measured impedance at a given frequency  $\omega$  and applied voltage  $V_a$ , and was assumed to be entirely due to the capacitance of the junction. The range of suitable frequencies could easily be determined by inspecting the impedance spectrum of the device.

This type of measurement is commonly used to characterise the so called voltage coefficient of the capacitance (VCC) in MIM structures with thin films and is an important

test for the reliability of MIM capacitors for RF applications[210, 211], where precision filters and strong DC offsets are often required. The voltage coefficient  $\gamma(V_a)$  represents the normalised difference between the nominal capacitance  $C_J(0)$  and that at a given voltage

$$\gamma(V_a) = \frac{C_J(V_a) - C_J(0)}{C_J(0)}. \quad (5.13)$$

The form of  $\gamma(V_a)$  for most insulating materials used in MIM devices can then generally be described empirically as

$$\gamma(V_a) = \alpha(\omega, t_J)V_a^2 + \beta(\omega, t_J)V_a, \quad (5.14)$$

where  $\alpha$  and  $\beta$  represent a quadratic and linear factor respectively, that each depend on frequency and thickness. These coefficients also depend on temperature, however only measurements at room temperature were performed in these experiments. For most high-k materials, including  $\text{Al}_2\text{O}_3$ ,  $\gamma(V_a)$  tends to increase with applied voltage, such that at least  $\alpha$  is positive. In some cases the voltage coefficient can be negative such as for  $\text{SiO}_2$ [212] and  $\text{TiO}_2$ [213]. In MIM devices for RF applications, the smallest possible value of  $\gamma(V_a)$  is usually required, which is an important driver of MIM research[212, 214, 215]. As it is well known that  $\gamma(V_a)$  increases in magnitude with decreasing thickness[99], ultra-thin devices such as those presented here are generally not of great interest in RF applications, and therefore there is no specific research targeted towards the characterisation of the permittivity in such thin films. Figure 5.6 (a) and (b) show the calculated VCC for a 3 and 2 nm device at a series of frequencies, and measured in a cycle between -1.5 and 1.5 V, applied to the top electrode of the MIM structure. The connections to the device had to be switched around to measure the opposite polarity of the VCC. Both devices have a diameter of 160  $\mu\text{m}$ , and the values of  $C_J(0)$  obtained from Equation 5.12 are approximately 340 pF for the 3 nm device, and approximately 400 pF for the 2 nm device. The impedance spectrum of the 3 nm device is that shown in Figure 5.4 (b), and the nominal capacitance is in close agreement with the capacitance derived by fitting the impedance to the circuit model. A slight hysteresis was observed at low voltage, which is possibly due to trapped charges incorporated in the film during higher voltage measurements, which may increase the permittivity slightly due to the formation of additional dipoles. The noise in the higher frequency measurements was less prominent due to the much larger AC current available.

There are a number of important features in the curves of Figure 5.6 that suggest the dielectric properties of the film are consistent with an  $\text{Al}_2\text{O}_3$  layer in a MIM capacitor. Firstly the VCC is positive and increases with thickness, which is the expected behaviour for  $\text{Al}_2\text{O}_3$ [99]. Furthermore, the increase of  $\gamma(V_a)$  with decreasing frequency is known in ALD deposited  $\text{Al}_2\text{O}_3$  films[216]. Both of these properties occur in many

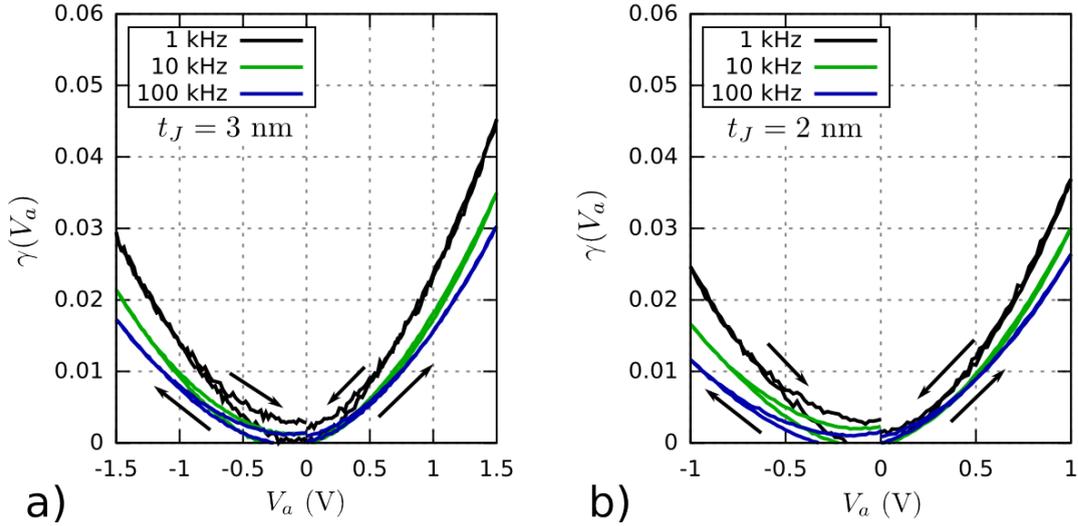


FIGURE 5.6: C–V measurement results, expressed in the form described by Equation 5.13, for a 3 nm device (a), and a 2 nm device (b). A series of frequencies are shown.

materials besides  $\text{Al}_2\text{O}_3$  however, particularly in the high-k oxides, which tend to have mostly polar (ionic) bonds. The resulting ionic dipoles induced under an applied electric field are understood to be responsible for the high static permittivity of these materials[209], which depends on both the electronic and atomic structure of the material. The STEM/EDX analysis of the  $\text{Al}_2\text{O}_3$  film presented in section 4.4 showed that Al and O dominate the X-ray spectrum. Although the precise composition of the film could not be determined from these measurements, electronic measurements of the static relative permittivity  $\epsilon_r$  in thick films are consistent with widely reported values for disordered, ALD deposited  $\text{Al}_2\text{O}_3$  thick films ( $\geq 7$  nm). Another important feature of the  $\gamma(V_a)$  curve is the clear asymmetry in the polarities, and the offset of the minimum from the origin. This asymmetry can also be observed in Bécu *et al.*'s C–V experiments with  $\text{Al}_2\text{O}_3$  MIM structures, formed of TiN bottom and top electrodes in a CMOS compatible fabrication process. There is little specific literature that treats the physical origin of the slight capacitance enhancement in ceramic MIM structures, and therefore the asymmetry is difficult to interpret. A series of plausible explanations are presented in Chapter 6, based on the understanding of the capacitance enhancement in the literature and the modelling of the tunneling results, which provides some insight into the properties of the metal layers.

The offset voltage from the origin appears similar in both devices, which suggests this effect could be due to a difference between the top and bottom electrode work functions[217], since the internal electric field  $E_i$  can be large enough to offset the  $C_J(V_a)$  value by a discernible amount. If  $\Phi_B$  is the bottom electrode work function, and  $\Phi_T$  is the top electrode work function, there exists a potential difference across the insulating

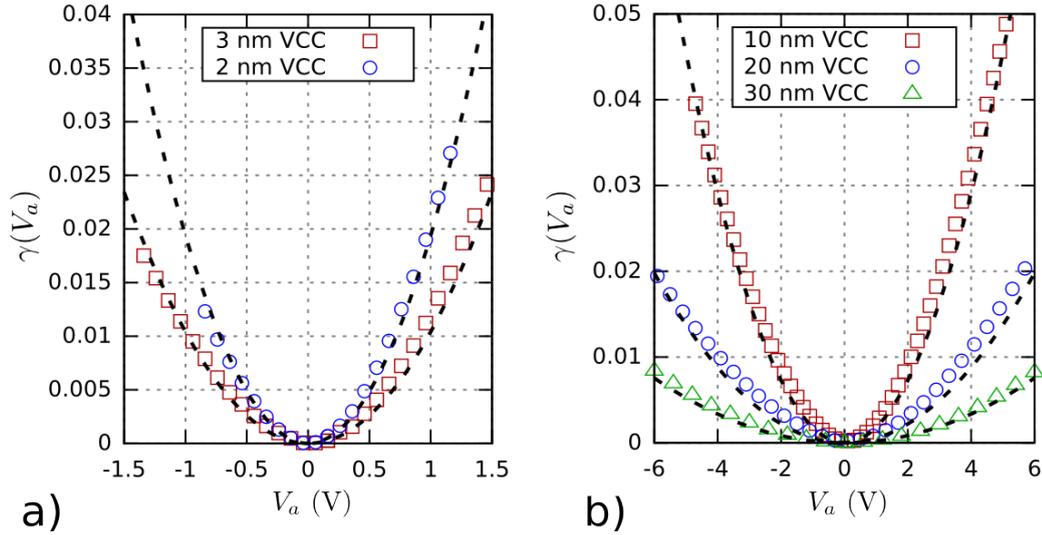


FIGURE 5.7: (a) Fitting of the 2 and 3 nm device VCC data at 100 kHz, using Equation 5.14 with  $\beta = 0$ , where only the up-sweep was considered for both polarities. (b) Fitting of the 10, 20 and 30 nm device VCC data at 100 kHz, using the same method as for (a). The curves in (a) and (b) were centred by applying offsets discussed in the text.

region

$$V_i = \frac{\Phi_B - \Phi_T}{e} = \frac{\Delta\Phi}{e}, \quad (5.15)$$

which is required for the junction charge to reach equilibrium. In the parallel plate model, this internal potential difference results in a uniform internal electric field  $E_i = V_i/t_J$ , which has the same effect as an applied field. Under this approximation, the magnitude of the (macroscopic) total electric field is due to the superposition of the applied field and the internal field

$$E_T = \frac{1}{t_J} \left( V_a + \frac{\Delta\Phi}{e} \right), \quad (5.16)$$

normal to the metal surfaces, where the internal contribution can be constructive or destructive to the applied electric field. This is discussed in more detail in Chapter 6 concerning the modelling of the tunneling I–V measurements. This internal electric field is a very plausible explanation for the offsets observed in the VCC curves, which appears to increase with decreasing thickness, consistent with Equation 5.16. Equation 5.14 is then better expressed as

$$\gamma(\omega, V_a) = \alpha(\omega, t_J) \left( V_a + \frac{\Delta\Phi}{e} \right)^2 + \beta(\omega, t_J) \left( V_a + \frac{\Delta\Phi}{e} \right), \quad (5.17)$$

such that the work function difference  $\Delta\Phi$  is accounted for.

Figure 5.7 (a) shows the results of fitting the 100 kHz up sweep  $\gamma(V_a)$  curves of the 2 and 3 nm type-B devices to Equation 5.17, represented by the dashed lines, and

the data points are denoted by the symbols. An offset of approximately 165 mV was obtained for both curves, which has been accounted for in the plot to centre the curve. Only the quadratic factor  $\alpha$  of Equation 5.14 was used as a free parameter in the fitting routine, so that any linear contribution is reflected in  $\Delta\Phi$ . In this case, according to Equation 5.17, the offset caused by the work function difference can be expressed as the  $\beta$  parameter

$$\beta = 2\alpha \frac{\Delta\Phi}{e} \quad (5.18)$$

and the corresponding offset in the VCC itself at  $V_a = 0$  is then  $\alpha(\Delta\Phi/e)^2$ , which increases with decreasing thickness due to  $\alpha$ . A value of  $\alpha \approx 0.01 \text{ V}^{-2}$  was obtained for the 3 nm device, which represents roughly a 1% increase in the capacitance at 1 V applied voltage. For the 2 nm device,  $\alpha \approx 0.02 \text{ V}^{-2}$  was obtained, which represents approximately a 2% increase at 1 V applied voltage. The adsorption of oxygen onto the bottom Au surface has been shown to increase the work function up to 5.35 eV[218], 250 meV greater than the usual value reported for evaporated thin films on the order of 5.1 eV[219]. The bottom Au electrode is exposed to oxygen plasma and a small amount of oxygen gas during the ALD process, and growth was observed to be linear in every case, which suggests that there are sufficient oxygen reactive sites on the surface. However Au deposited on top of ALD deposited  $\text{Al}_2\text{O}_3$  is known to have poor adhesion, and has been used in fabrication processes to create plasmonic nanostructures, where top gold layers can be easily lifted off[220–222]. The bond between the  $\text{Al}_2\text{O}_3$  layer and the top electrode is therefore weaker than that with the bottom electrode, and may result in a work function difference on the order of the observed shift. The direction of the shift in the C–V measurements is consistent with this observation. Such small shifts are not readily observable in devices with thicker films, since the electric field produced reduces significantly with increasing thickness. A device with a 1.4 nm thick film and similar electrode diameter as the 2 and 3 nm devices was also characterised at 100 kHz, however the device was damaged during the experiment which resulted in a permanent change in the measured capacitance. Despite this damage incurred during the measurement, an offset of approximately 165 mV consistent with the other devices could still be determined from the portion of the curve that was measured before the device was damaged.

C–V measurements of thick film devices were performed to compare with the thinner films and with VCC values reported in the literature, which are usually performed at high frequencies. Voltages ranging from  $\pm 5 \text{ V}$  to  $\pm 15 \text{ V}$  were applied to the devices to obtain electric fields on the same order as those in the thin film devices (approximately  $5 \text{ MV}\cdot\text{cm}^{-1}$ ). In these devices the entire impedance range at low frequencies is dominated by the reactance and large DC voltages are required to produce a large electric field. Figure 5.7 (b) shows the results of fitting the VCC curves obtained for 10 nm, 20 nm

and 30 nm thick film type-A devices with a 160  $\mu\text{m}$  diameter, shown between -6 V and 6 V. Small offsets on the order of  $\pm 150$  mV are present that are not consistent with the difference between the top and bottom electrodes, which is expected to be on the order of 500 mV, since the Cr work function is smaller than the Au work function by approximately this amount[223, 224], and is expected to be unique in polarity as observed in the thin film devices. The observed offsets are possibly due to other extrinsic effects that contribute to  $\beta$ , or noise in the measurements at low voltages. The values of  $\alpha$  were found to decrease with thickness as expected, yielding  $1.819 \times 10^{-3} \text{ V}^{-2}$ ,  $5.5 \times 10^{-4} \text{ V}^{-2}$  and  $2.1 \times 10^{-4} \text{ V}^{-2}$  for the 10 nm, 20 nm and 30 nm films respectively. The nominal capacitance  $C_J(0)$  obtained for each thickness device was on the order 140 pF, 70 pF and 50 pF respectively.

Figure 5.8 (a) and (b) shows the C–V results re-expressed in terms of the total electric field  $E_T$ , in order to highlight the dependence of the dielectric constant itself on the electric field. The change in the capacitance under an applied bias is fundamentally linked to the dependence of the permittivity of the film on the total electric field, which can be expressed as

$$C_J(V_a) = \epsilon_0 \epsilon(E_T) \frac{A_J}{t_J}, \quad (5.19)$$

where  $A_J$  and  $t_J$  are assumed independent of the applied voltage. An empirical expression similar to Equation 5.14 can then be found for the dielectric constant

$$\epsilon(E_T) = \epsilon(0)(1 + aE_T^2), \quad (5.20)$$

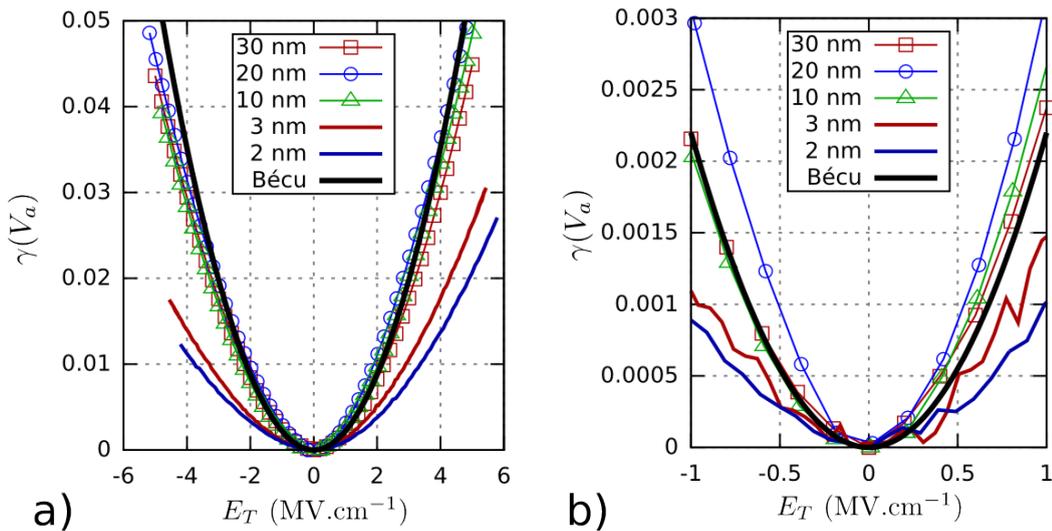


FIGURE 5.8: (a)  $\gamma(V_a)$  expressed as a function of the total electric field  $E_T$ , computed using Equation 5.16. The dotted line in corresponds to Bécú *et al.*'s result for similar MIM devices, obtained at 100 kHz[99]. (a) Same data as in (b) displayed in the same electric field range as in Bécú *et al.*'s experiments.

where  $\epsilon(0) = \epsilon_r$  is determined from the impedance measurements, and  $a$  is an empirical fitting factor similar to  $\alpha$ . Equation 5.20 is valid as is in a homogeneous electric field, and thus  $a$  is related to  $\alpha$  through

$$a(\omega) = \alpha(\omega, t_J)t_J^2, \quad (5.21)$$

where  $a$  is a parameter more representative of the behaviour of the permittivity and is independent of the geometry of the MIM capacitor as long as the electric field is homogeneous. Bécu *et al.* have developed an insightful microscopic model to explain the quadratic dependence of the VCC[225], and which incorporates the magnitude of the permittivity in an empirical way. This model was used in their analysis of the thickness dependence of  $\gamma(V_a)$  in MIM devices with  $\text{Al}_2\text{O}_3$  film thicknesses ranging from 7 nm up to 20 nm[99], and will serve as a starting point in Chapter 6 for modelling of the C–V results presented here. They found that  $a$  was independent of thickness in their devices and their result is represented by the thick solid line in Figure 5.8 (a) and (b). The electric field applied to their devices was not as strong as that used in this work, and thus the curve shown in Figure 5.8 is a projection of their result based on the fitting parameters they obtained with Equation 5.20. Figure 5.8 (b) represents the data in the range of electric fields used by Bécu *et al.* in their devices. As can be seen, the value of  $a$  for the thick film devices presented in this work are highly consistent with their result. With the present ultra-thin film devices, a decrease in  $a$  is observed with thickness apart from the observed decrease of the static permittivity with thickness. Since both of these trends are representative of the permittivity of the film, it is likely that they are linked. These results therefore suggest that a more complete empirical description of the permittivity, ignoring the frequency dependence, can be expressed as

$$\epsilon(E_T, t_J) = \epsilon(t_J)(1 + a(t_J)E_T^2), \quad (5.22)$$

where  $a(t_J)$  reflects the behaviour of the permittivity under an applied electric field and  $\epsilon(t_J)$  reflects the thickness dependence of the nominal value of the permittivity. Using Equation 5.21, the values of  $a$  obtained for each thickness were calculated as approximately  $7.3 \times 10^{-4} \text{ cm}^2.\text{MV}^{-2}$  and  $9.4 \times 10^{-4} \text{ cm}^2.\text{MV}^{-2}$  for the 2 nm and 3 nm thick films respectively and Bécu *et al.* obtained  $2.2 \times 10^{-3} \text{ cm}^2.\text{MV}^{-2}$  for each thickness which is consistent with the values obtained for the 10 nm, 20 nm and 30 nm film devices.

From the present results, it is also clear that the difference between the dielectric constants obtained for the type-A and B devices cannot be explained by the larger difference in work function alone. The work function of Cr is commonly reported to be close to 4.5 eV[223, 224] such that the work function difference is on the order of

0.6 eV, which would result in only a very small change of the permittivity. Chromium is known to form a good chemical bond with  $\text{Al}_2\text{O}_3$ , evidenced by its good adhesion properties[226]. In a ALD-based MIM structure that employs Cr as the top electrode, both the top and bottom electrodes form a reliable chemical bond. In the type-B (Au top electrode) MIM structure the poor chemical bond may be responsible for the increased reduction in the dielectric constant. This is discussed in Chapter 6 concerning the modelling of the permittivity thickness and voltage dependence.

### 5.3 DC conductance measurements

In the previous section, the AC properties of the junctions was determined, in which a strongly thickness-dependent parallel resistance  $R_J$  was observed in films with  $t_J \leq 3$  nm. In this section, this parallel resistance is investigated against an applied DC bias  $V_a$ , which is discussed in the first subsection, and against temperature  $T$ , discussed in the following subsection. The dependence of  $R_J(V_a, T)$  against these parameters and knowledge of the material properties of the junction serves to identify the conduction mechanisms involved in manifesting the recorded current[193]. The measurement circuit shown in Figure 3.13 was used in all DC characterisation experiments. Both the source and the ground wires are shielded using coaxial cables with the outer connection made to the earthed chassis of the Keithley 2400LV. This section will only focus on the properties of the thinnest film devices, which are of greatest interest to SET devices. In the previous section, thick film devices were required to confirm that the permittivity of the film becomes independent of thickness, and to check the consistency of the C–V measurements. All working MIM devices with  $t_J > 5$  nm were highly insulating, and significant electric fields were required to produce a measurable current. In most cases, these devices were damaged shortly after the current became measurable, which provided a rough estimation of the breakdown field strength of  $8 \text{ MV}\cdot\text{cm}^{-1}$ , which is on the order of reported measurements of the breakdown field[181, 199] for the  $\text{Al}_2\text{O}_3$  PEALD process. Measurements of the thin film devices showed a measurable current at an electric field significantly below the breakdown field. Electric fields up to  $6 \text{ MV}\cdot\text{cm}^{-1}$  were used to characterise certain devices, to gain as much information on the barrier parameters as possible, which is discussed in more detail in Chapter 6 where models are used to estimate the barrier parameters. This section focuses on the consistency of the measurements as a function of the geometrical parameters of the junctions.

#### 5.3.1 Tunneling thickness and area dependence

When a potential difference is applied between the top and bottom electrodes, a current  $I$  flows that sensitively depends on the material properties and the geometry of the junction. Under the assumption of ideal planar electrodes and a uniform thickness,

the current density can be expressed as

$$J(t_J) = \frac{I(t_J)}{A_J}, \quad (5.23)$$

where the underlying assumption is that the current density is independent of position in the volume between the electrode surfaces. Similarly, the applied electric field can be expressed as

$$E(t_J) = \frac{V_a}{t_J}, \quad (5.24)$$

which is also assumed to be independent of position. The values of  $A_J$  and  $t_J$  used are those shown in Figure 4.10. Using the generalised form of Ohm's law, the conductivity is expressed as

$$\sigma(t_J, V_a) = \frac{J(t_J)}{E(t_J)}, \quad (5.25)$$

where  $\sigma(t_J, V_a)$  is the conductivity of the junction at a given film thickness and applied voltage, and where it is implied that the current density is generally not proportional to the applied electric field. The junction resistance at a given voltage and thickness can then be expressed as

$$R_J(t_J, V_a) = \frac{1}{A_J \sigma(t_J, V_a)} = \frac{\rho(t_J, V_a)}{A_J}, \quad (5.26)$$

where  $\rho(t_J, V_a)$  is the surface resistivity expressed in Equation 5.9 in the previous section. Figure 5.9 (a) shows a summary of the type-A device current density—electric field (J–E) curves, in which a series of 2 nm, 3 nm and 5 nm thick film devices with different surface areas were characterised. Figure 5.9 (b) shows a summary of the type-B device J–E curves, in which a series of 1.4 nm, 2 nm and 3 nm thick film devices with different surface areas were characterised. Equations 5.23 and 5.24 were used to calculate  $J$  and  $E$ . From these J–E characteristics it is clear that  $J$  is a non-trivial function of  $E$ , especially for the type-B devices, shown in panel (b). In both cases there is a clear exponential dependence of the current density as a function of thickness, at high electric fields. This is the first evidence that electron tunneling likely dominates in the thinnest films. The smallest measurable currents were on the order of 1 pA, which is achieved using long averaging times and accounting for the constant offset current present in each Keithley 2400LV, which was within  $\pm 50$  pA, depending on the system used. This limit prevented accurate measurements of the current at low electric fields in the thicker devices, as can be seen by the spread of the 3 nm and 5 nm device curves at low electric fields. The shape of the J–E curves of the thinnest devices ( $\leq 2$  nm) are highly consistent at a given thickness, and there are offsets that reflect errors in  $A_J$  rather than  $t_J$ . The type-B devices are found to be much less conductive than the type-A devices, which is expected to be due to the larger effective barrier height in these devices, since the

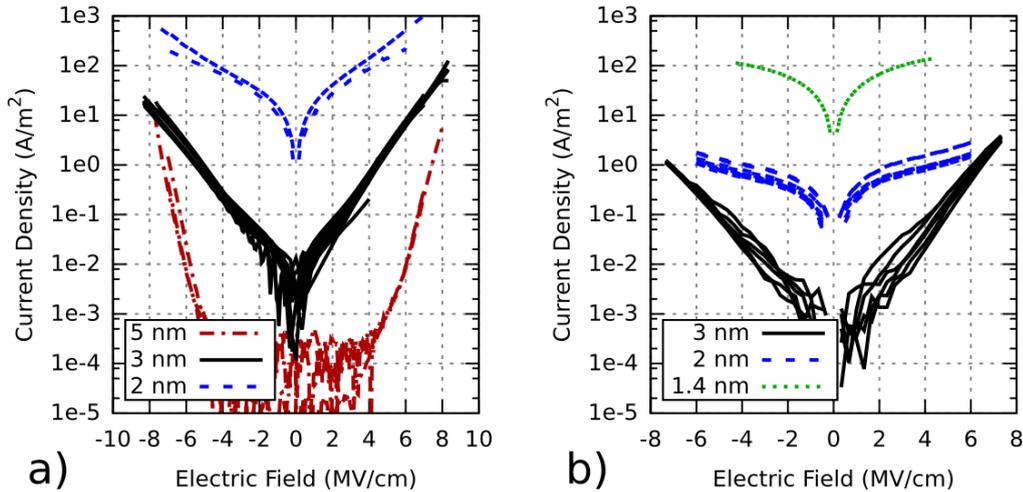


FIGURE 5.9: (a) Summary of the J–E measurements of 5 nm, 3 nm and 2 nm thick type-A devices with different diameters. (b) Summary of the J–E measurements of 3 nm, 2 nm and 1.4 nm thick type-B devices.

Au work function is larger than that of Cr, as discussed previously. However, it will be shown in further paragraphs that this cannot be the only factor that causes this difference.

Figure 5.10 (a) and (b) shows the J–E data computed using Equation 5.25 and 5.26 to express the resistivity as a function of thickness and applied voltage. As can be seen in plot (a), the resistance of the type-A devices decreases non-linearly and monotonically. In the type-B thin-film devices however, the resistance changes non-monotonically for the 2 nm devices and below. Peaks in  $R_J(V_a)$  can be observed in the 2 nm type-B devices in both polarities. An asymmetry of the current between forward and reverse polarity can be observed in all measured devices. All devices show a slightly higher current when the top electrode is positively biased, such that electrons flow from the bottom electrode to the top electrode. The qualitative shape of the J–E and resistivity curves in the type-A devices closely resembles the typical shapes described in the literature for electron tunneling[100, 181, 193, 227–229], whereas for the type-B devices there appears to be subtle effects not commonly encountered in MIM devices in general. An effective way to identify the conduction mechanism in the devices is to measure the temperature dependence of the current. In thin insulating film devices in general, a series of different conduction mechanisms are commonly observed. There exist classical and semi-classical thermally activated conduction mechanisms, such as thermionic (Schottky) emission[230], an electrode limited process, and Frenkel-Poole conduction[231, 232], a bulk limited process. These processes are generally highly temperature dependent, and can be described by a

thermal activation law of the form[233]

$$I \propto \exp\left(-\frac{W_a}{k_B T}\right), \quad (5.27)$$

where  $W_a$  is an activation energy related to the underlying physical mechanism responsible for the current, and may depend on the applied voltage also[234]. None of the I–V measurements performed followed this law at all temperatures considered, which suggests these mechanisms are not present, or at least they are not discernible from the measured current. Frenkel-Poole type conduction mechanisms have previously been identified in thick ( $> 5$  nm) ALD deposited  $\text{Al}_2\text{O}_3$ [181, 235], from which it is possible to extract the energy levels of electron trapping centres within the insulator band gap[236], since the electron transfer across the structure occurs by field-assisted thermal hopping between trapping centres. Variants of the Frenkel-Poole conduction mechanism exist that treat the hopping between trapping sites as tunnel-assisted, as well as thermally-assisted[237, 238], however these mechanisms also have a significant temperature dependence again following Equation 5.27. The temperature dependence measurements are discussed in detail in the next section, focusing on the ultra-thin film devices.

Due to the disordered nature of the ALD deposited  $\text{Al}_2\text{O}_3$  film, as seen in the TEM images presented in Figure 4.9 in Chapter 4, it is expected that electron traps exist intrinsically as a result of the long range random fluctuations of the atomic potential in such structures. In these structures the band theory of electrons in solids cannot be applied exactly, since there is no symmetry in the material. N. F. Mott showed that in such structures the conduction and valence band edges are not well defined, rather, the quantity of trap energy levels for electrons increases as the average maximum of the

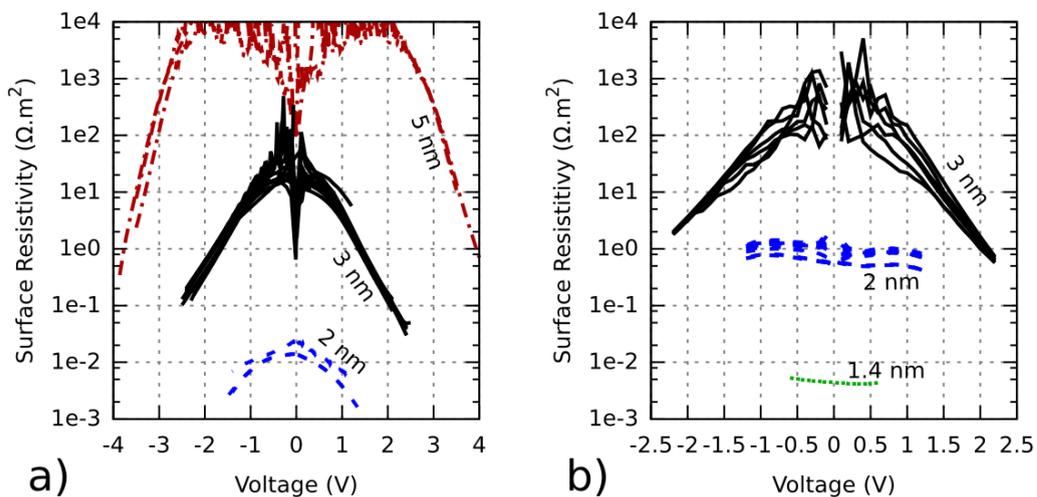


FIGURE 5.10: Measurements from Figure 5.9 (a) and (b) expressed as the resistivity of the junctions as a function of voltage.

atomic potential energy is approached, beyond which a conduction band is formed[142]. The same occurs with holes to form a valence band, and the resulting band gap consequently contains intrinsic traps near the conduction and valence band edges. The band gap of thick ALD  $\text{Al}_2\text{O}_3$  films is estimated to be between 6 and 7 eV[239, 240], which represents a very good insulator. The permittivity measurements discussed in the previous section suggest that the material properties of the  $\text{Al}_2\text{O}_3$  layer in our device must have similar electronic parameters. The reduction of the permittivity with thickness also suggests that the band gap could increase, according to the commonly ascribed trend that permittivity is intrinsically inversely related to the band gap[209], although this is usually for thin films representative of the bulk properties, in the present case for  $t_J \geq 10$  nm. In such large band gap disordered materials of high purity[241], the random fluctuations of the atomic potential energy are much smaller in amplitude than the band gap itself, and thus virtually no intrinsic trap levels exist near the Fermi level. The Fermi level in the insulator is determined by the thermal equilibrium of the electrons and holes both extrinsic and intrinsic to the material, but generally lies near the middle of the band gap in such a wide-gap insulator[242, 243]. Since the thermal energy at room temperature is much smaller than the band gap and the occupancy of the electron traps is governed by Fermi statistics (Equation 2.26), virtually no electron trap levels can be occupied at equilibrium. Therefore these traps play no direct role in the conduction of electrons under near-equilibrium conditions in the junction, such as low voltage measurements. Electron traps that exist due to extrinsic factors have been detected in ALD  $\text{Al}_2\text{O}_3$ , and are estimated to lie between 1.4 eV and 1 eV below the conduction band using Frenkel-Poole measurements in thick films as a function of temperature and using EELS techniques[240, 244]. These trapping levels have been analysed mostly in metal-insulator-semiconductor (MIS) junctions, due to the considerable interest in the use of high-k materials and ALD techniques for CMOS transistor gate insulators. In such systems, the alignment of the conduction and valence bands between the semiconductor and insulator layers is a critical design consideration, as charges inside the material layers and in the interface region cause an effective offset in the alignment of the bands to reach equilibrium, which affects the barrier height of the gate material, and consequently the leakage current[245].

Figure 5.11 (a) shows a simplified energy band diagram of the MIM structures, where the electric field throughout the insulator is assumed to be constant.  $\Phi_{\pm}$  represents the metal electrode work functions, and  $\chi$  represents the electron affinity of the  $\text{Al}_2\text{O}_3$  layer, which describes the energy required to bring an electron at bottom of the  $\text{Al}_2\text{O}_3$  conduction band to the vacuum energy level  $E_{\text{vac}}$ . The band diagram is formed by assuming that before contact is made between the materials,  $E_{\text{vac}}$  is the same for each material and the Fermi energy  $E_{\text{F}}$  differs for each material. When the MIM junction is formed, charges may flow in, out and through the insulator to establish equilibrium,

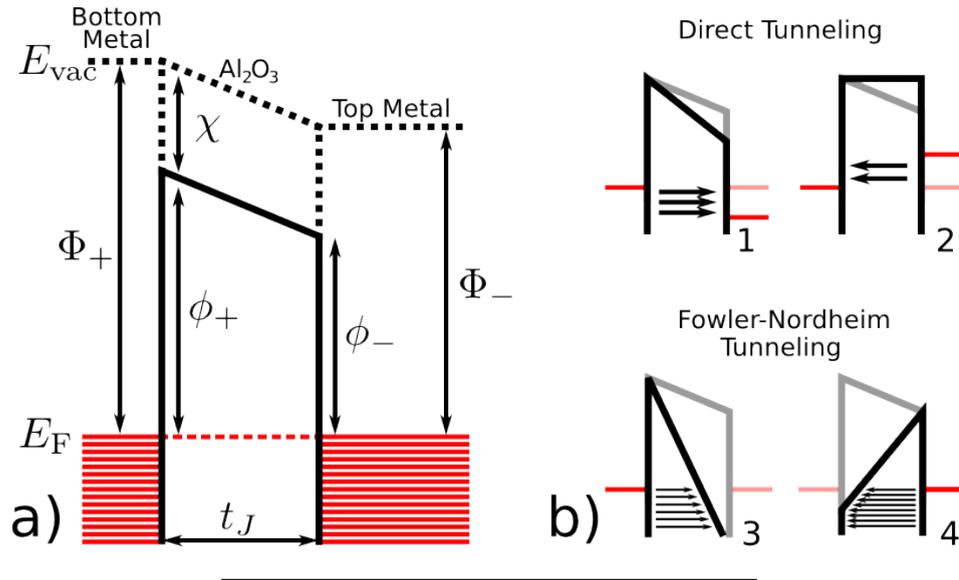


FIGURE 5.11: (a) Simplified energy band diagram of the MIM structures used in this work, showing the important junction parameters discussed in the text. (b) Energy band diagrams depicting the junction in the forward (1) and reverse (2) direct tunneling regime, and similarly for the FNT regime, (3) and (4). The arrow density in (b) indicates the relative magnitude of the current in each configuration.

so that the Fermi level  $E_F$  becomes the same in each material, which results in an internal potential difference when  $\Phi_+ \neq \Phi_-$ , as discussed previously. Depending on the trapped charge density and energy, and the thickness of the film, the insulator conduction band bends due to the screening of the metal plane by the trapped charges, resulting in different types of metal-insulator contacts[242, 243]. In thin films such as those used in this work, a significant trapped charge density is required to cause discernible band bending, since the metal contacts are so close together they cannot be screened effectively, and hence does not need to be considered in this work. However, trapped charges that remain in the thin insulating layer after equilibrium is established will cause a so called space-charge to exist in the layer, which modulates the potential energy seen by electrons, either enhancing or suppressing electronic conductivity due to positive, and negative space-charge respectively[246]. Static negative space-charge has been reported in  $\text{Al}_2\text{O}_3$  usually in the context of MIS junctions[181, 247], and is usually not specifically reported in modern MIM tunnel diode literature, which is probably due to the much higher quality insulating films that can be created with modern thin film deposition techniques (see ALD related references). Therefore the effect of space charge is not treated explicitly in Chapter 6 on the modelling of the tunneling characteristics of the junctions. The top portion of Figure 5.11 (b) shows the junction under forward (1) and reverse (2) bias in the direct tunneling regime. The current due to electron emission from the higher work function metal is slightly more conductive in this regime, and is discussed in detail in Chapter 6.

The electron tunneling process is known to be relatively temperature independent, since energy is not dissipated or gained during the tunneling transition itself, i.e. the kinetic energy of electrons is not changed throughout the transition across the insulator. In early experimental work on MIM structures, due to limitations in fabrication technology, only thick film (on the order of 10 nm) devices could be reliably produced, and consequently electron tunneling was first discovered experimentally by application of very high electric fields[248]. Under these conditions the conduction is entirely electrode limited, since the rate limiting process is electron tunneling from the cathode to a region in the insulator with a considerably lowered potential due to the high electric field, which is depicted in the lower portion of Figure 5.11 (b). This conduction regime is known as the Fowler-Nordheim tunneling (FNT) regime, where the potential energy barrier appears triangular to electrons in the cathode, and thus appears considerably reduced in height and width so that the tunneling probability becomes significant and highly sensitive to the applied voltage. In the simplest case of this regime, the current density obeys a law of the form

$$J_{\text{FN}} = P_1 E_T^2 \exp\left(-\frac{P_2}{E_T}\right), \quad (5.28)$$

where  $P_{1,2}$  are fitting parameters and  $E_T$  is the total electric field as defined by Equation 5.16. The study of FNT is today a highly advanced field, driven by various applications such as electron emitters in electron microscopes, and the interpretation of the parameters  $P_{1,2}$  must be carefully undertaken to obtain accurate results[249, 250]. Figures 5.12 (a) and (b) show the J-E data for three 5 nm film devices expressed in the form of Equation 5.28, using the applied electric field  $E_a = V_a/t_J$ , and centred onto the high electric field region of the curve. The diameter of two of the devices was 80  $\mu\text{m}$  and 320  $\mu\text{m}$ , respectively. Surprisingly, despite the great difference in surface area, the current density appears consistent for two of three devices. In both plots the highest electric field regions of the curves appear to be reasonably linear, particularly in the forward characteristic, which throughout this work is referred to as emission from the bottom Au electrode, which was almost always the more conductive polarity in the voltage range of the measurements. The forward characteristic is then obtained when a positive (negative) potential is applied to the top (bottom) electrode.

FNT conduction is commonly identified in MIM devices with thick ( $t_J \geq 5$  nm) ALD  $\text{Al}_2\text{O}_3$  films. The data corresponding to the green curve in Figures 5.12 (a) and (b) is fitted to Equation 5.28. Following Cowen *et al.*'s method, the FNT equation can be expressed in terms of the forward (+) and reverse (-) current as[251]

$$I_{\pm} = \pm \frac{A_J P'_1}{t_J^2 \phi_{\pm}} (V_a + V_i)^2 \exp\left(\mp P'_2 \sqrt{m^* \phi_{\pm}^3} \frac{t_J}{V_a + V_i}\right), \quad (5.29)$$

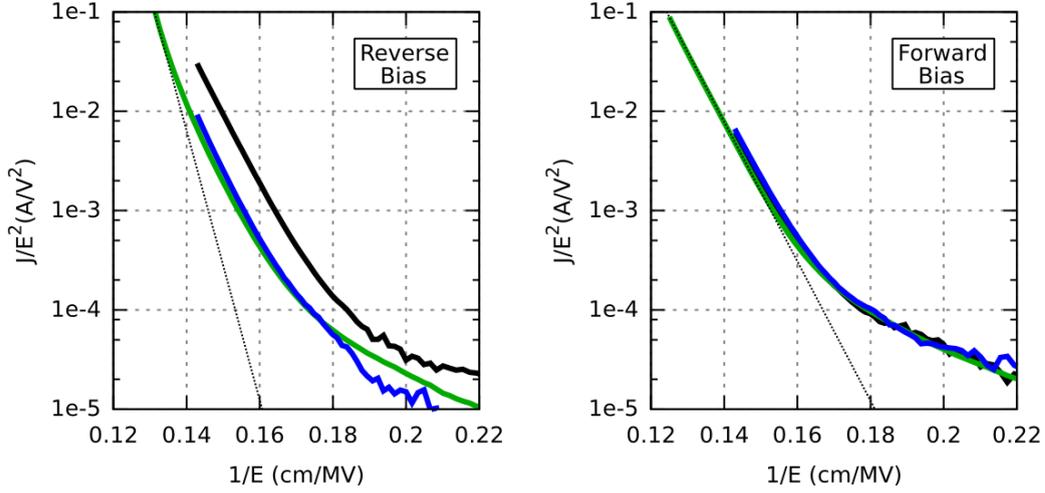


FIGURE 5.12: Fowler-Nordheim plot of the 5 nm thick film device, the linear regions in the curves can be fitted to Equation 5.28.

where  $P'_{1,2}$  represent fitting parameters that are related to physical constants in the analysis of planar  $\text{Al}_2\text{O}_3$  MIM junctions[246, 249–251], and  $V_{a,i}$  are the applied and internal potential difference.  $\phi_{\pm}$  represents the energetic barrier height (shown in Figure 5.11 (a)) in eV, and is given by

$$\phi_{\pm} = \Phi_{\pm} - \chi, \quad (5.30)$$

which shows that a metal with a small work function results in a smaller potential barrier at equilibrium, as depicted in Figure 5.11 (b). The electron affinity  $\chi$  of the insulator determines the height of each interfacial barrier. The effective mass  $m^*$  represents the mass that electrons appear to have when inside the insulating layer, and is the result of approximating electrons at the  $\text{Al}_2\text{O}_3$  conduction band edge as having a parabolic dispersion relation, which was discussed briefly in Section 2.2.1, and will be discussed in more detail in Chapter 6. Each of  $P'_{1,2}$  are defined as[246]

$$P'_1 = \frac{e^3}{8\pi\hbar} \approx 1.541 \times 10^{-6} \text{ A.eV.V}^{-2}, \quad (5.31)$$

and

$$P'_2 = \frac{4\sqrt{2}}{3e\hbar} \approx 6.83 \times 10^9 \text{ eV}^{-\frac{3}{2}}.\text{V.m}^{-1}. \quad (5.32)$$

In the fitting routine,  $P'_{1,2}$  are kept constant, and  $V_i$  and  $\phi_{\pm}$  are used as the fitting parameters. A value of  $m^* = 0.44m_e$ [252, 253] was used for the effective mass of bulk  $\text{Al}_2\text{O}_3$ . Due to the curvature in the reverse characteristic, a conclusive fit could not be obtained, and both polarity curves could not be fitted simultaneously, which suggests that the reverse characteristic has not fully reached the FNT regime. This would explain the symmetry observed between the devices measured at a lower electric field, shown

by the blue and black lines in Figures 5.12 (a) and (b), since at the onset of FNT, the asymmetry in the current, defined as

$$\eta_{\text{asy}} = \frac{I_+(V_a)}{I_-(V_a)}, \quad (5.33)$$

approaches unity, and in the FNT regime, becomes very small[193]. The asymmetry calculated for each of the curves in Figures 5.12 (a) and (b) is illustrated in Figure 5.13 (a). The values obtained from fitting the forward characteristic to  $I_+$  are  $\phi_+ \approx 3$  eV,  $\phi_- \approx 2.3$  eV and  $V_i \approx 0.7$  V, and consequently the  $\text{Al}_2\text{O}_3$  layer electron affinity is determined as  $\chi = \Phi_{\pm} - \phi_{\pm} \approx 2.1$  eV, which is consistent with literature values[100, 219, 239]. The value of  $V_i$  is also consistent with the difference between the Cr and Au work function, as discussed in the previous section. These results show that thick film devices fabricated during this work are consistent with reports of similar work on similar devices, and that devices with a thinner film should show direct tunneling at voltages below the barrier height  $\phi_+$ , which should be similar in the bottom electrodes for both the Au and type-A devices. The dashed lines in Figures 5.12 (a) and (b) represent the result obtained using Equation 5.29 for each polarity.

The tunneling characteristics of selected ultra-thin film devices (2 nm and 3 nm) are modelled in Chapter 6, where a theoretical model was developed to describe the observed asymmetry in the low voltage characteristics. The FNT analysis presented above only provides a rough estimate of the bulk electrostatic properties of the  $\text{Al}_2\text{O}_3$

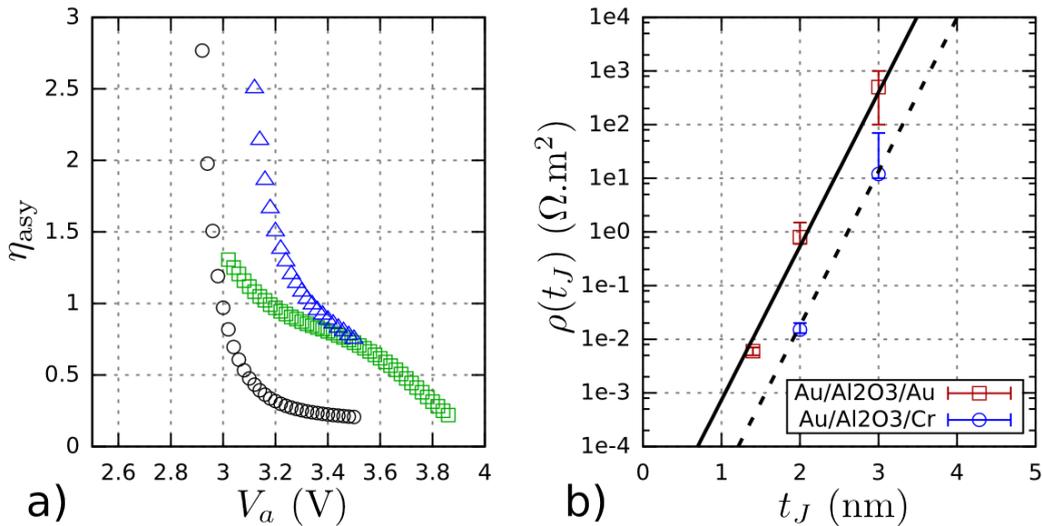


FIGURE 5.13: (a) Asymmetry in the current obtained from the 5 nm type-A devices, calculated using Equation 5.33. The colours of the symbols correspond with those in Figure 5.12. (b) Estimated low voltage surface resistivity  $\rho(t_J)$  as a function of film thickness. The error bars are estimated from the spread in the curves shown in Figures 5.12 (a) and (b). The solid and dashed line correspond to solutions of Equation 5.35.

film, and is not necessarily representative of devices with thinner films. To estimate the surface resistivity as a function of thickness, the resistance can be taken as approximately constant at low voltage. Figure 5.13 (b) shows a plot of the estimated values of the resistivity, including error bars corresponding to the observed spread in Figure 5.10. Despite the significant errors in the thicker devices, there is a clear power law governing the scaling of the resistivity with thickness that significantly exceeds this error, which is a testament to the remarkable thickness resolution and uniformity achievable with the Al<sub>2</sub>O<sub>3</sub> PEALD process. As was discussed in Chapter 2, the probability of a direct electron tunneling transition in a rectangular barrier follows a law of the form

$$P(V_a \approx 0) \propto \exp(-2\kappa t_J), \quad (5.34)$$

where  $\kappa$  represents the wave vector of electrons in the insulator potential barrier region. At a small voltage, a small energy range of electrons near the Fermi energy will result in excess transitions into the collecting electrode, leading to a finite conductance. The conductance in this small energy range is approximately proportional to the transition probability. In this approximation, it follows that the resistance, and hence the resistivity can be described as

$$\rho(t_J) = \rho_0 \exp(2\kappa t_J), \quad (5.35)$$

where  $\rho_0$  is a proportionality constant that in principle can be related to the tunnel barrier parameters. The solid and dashed lines in Figure 5.13 (b) represent the result of fitting Equation 5.35 to the estimated zero bias values of  $\rho(t_J)$  for the type-B and A devices respectively. A remarkably good fit is obtained, although the result for the type-A devices is ambiguous due to the limited number of points. Despite this, the rate of decrease of the resistivity with thickness appear to be similar as expected. The value of  $\kappa$  obtained can be used to estimate the effective rectangular barrier height using Equation 2.16 in Chapter 2

$$\phi_{\text{rect}} = \frac{\hbar^2 \kappa^2}{2m^*} \approx 1 \text{ eV}, \quad (5.36)$$

obtained with  $m^* = 0.44m_e$  as before. The value of  $\phi_{\text{rect}}$  is clearly too small in this approximation, as Schottky thermal emission becomes important for barriers below 1 eV, which was not observed. This result shows that in the available thickness range, the resistivity scales with the wavefunction decay constant  $\kappa$  of a 1 eV barrier. Similar parameters have been reported in Al-Al<sub>2</sub>O<sub>3</sub>-Al junctions using this method[254]. The error in this estimated value is quite large however, as the underlying approximation of the wavefunctions in the metals and the barrier region is crude, including the effective mass approximation of the electron energy dispersion, which is estimated from other experimental and theoretical work. To illustrate the effect of errors in the effective mass approximation, using a barrier height of 3 eV yields an effective mass around

0.14, which is just outside the reported range for ALD  $\text{Al}_2\text{O}_3$ . Using Equation 5.35 with the obtained values of  $\kappa$  and  $\rho_0$ , the thickness range required for SET systems can be estimated, where it is assumed that the material properties do not change with thickness. A value of  $\rho_0 \approx 10^{-6} \Omega\cdot\text{m}^2$  and  $\rho_0 \approx 3.3 \times 10^{-8} \Omega\cdot\text{m}^2$  was obtained for the type-B and A devices respectively, and  $\kappa \approx 3.3 \text{ nm}^{-1}$  was obtained in both devices. The values obtained for  $\rho_0$  are infeasibly large, since extrapolation of the scaling law to zero thickness with the derived parameters yields a very large value of  $\rho(0)$ , compared to the result  $\rho_0 \approx 10^{-13} \Omega\cdot\text{m}^2$  obtained by Dorneles *et al.*[254]. This suggests the barrier height and effective mass of the tunnel barrier may themselves depend on thickness, and the tunneling probability may depend on other properties in addition. The few reports of ALD fabricated SETs suggest that films on the order of 1 nm are required to produce high enough electron tunneling rates[104]. Considering the barrier parameters obtained in the FNT analysis of the 5 nm device, the value of  $\kappa$  obtained is approximately  $5.6 \text{ nm}^{-1}$ , which describes a greater rate of decrease of the resistance with decreasing thickness than that obtained from the smaller thickness devices. Considering the evidence shown in Chapter 4 for the high consistency in the  $\text{Al}_2\text{O}_3$  film thickness, and the change in the permittivity observed as thickness is decreased that is highly consistent across different batches and different device types (types A and B), the non-ideal scaling of the tunneling current as a function of thickness is likely linked to changes in the material properties as thickness is decreased, rather than significant inconsistencies in the geometry. Small inconsistencies in the geometry on the other hand could have a significant effect on the tunnel current[249, 250], which is considered in detail in Chapter 6 in terms of the evidence presented in Chapter 4 and in the present chapter.

### 5.3.2 Tunneling temperature dependence

The evidence presented in section 5.2 showed that for thick ( $t_J \geq 5 \text{ nm}$ ) deposited films, the measured material properties are consistent with high quality  $\text{Al}_2\text{O}_3$ . The I–V measurements showed that direct tunneling and FNT are likely the dominating conduction regimes, suggested by the observed material properties and the consistency of these results with various literature. In this subsection the current as a function of applied voltage and junction temperature  $I(V_a, T)$  is analysed. These measurements provide not only evidence for the direct tunneling interpretation of the current, but highlight subtle differences between the type-A and B devices. A number of devices were characterised in a temperature range between 2.6 K, using the closed loop He system, and 433 K using the hotplate system, both described in Section 3.4.1. Figures 5.14 (a) and (b) show the temperature dependence of the conductance of a type-A and B device, respectively, and each with a 2 nm thick  $\text{Al}_2\text{O}_3$  film. The conductance of type-A devices is easily calculated, as the current increases monotonically. The derivative of the

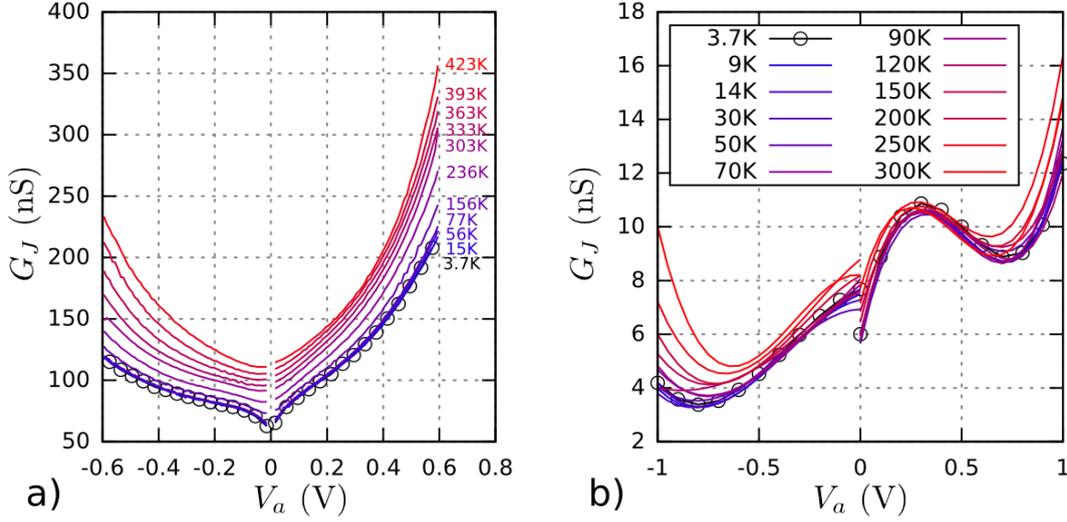


FIGURE 5.14: (a) Temperature dependence of the conductance of a type-A 2 nm device.  
 (b) Temperature dependence of the conductance of a type-B 2 nm device.

I–V curve shown in Figure 5.14 (a) was computed as

$$G_j = \frac{I_{j+1} - I_j}{V_{j+1} - V_j} \quad (5.37)$$

and using  $V_j' = 0.5 \times (V_j + V_{j+1})$  for the voltage axis. For the type-B devices, the conductance is difficult to compute using Equation 5.37 due to the excessive noise. Instead, a high order polynomial is fitted to the I–V curve, of the form

$$I(V_a) = \sum_{i=0}^n F_i V_a^i \quad (5.38)$$

up to some order  $n$ , where  $F_i$  are coefficients. It was found that  $n = 4$  was sufficient to reproduce the obtained I–V curves very well, where the goodness of fit for each curve was better than 0.999. The derivative can then be expressed analytically, which smoothens the conductance curve obtained. Figure 5.14 (b) shows the resulting smoothed conductance curve, calculated using

$$G(V_a) = \frac{dI}{dV_a} = \sum_{i=1}^n i F_i V_a^{i-1} \quad (5.39)$$

with the obtained coefficients  $F_i$ .

In the type-A device, the conductance appears temperature-independent at temperatures between 3.7 K to 77 K, after which there appears to be a small linear increase with temperature. The qualitative form of the thermal activation is consistent with the temperature dependence of Al–Al<sub>2</sub>O<sub>3</sub>–Al MIM junctions of similar thickness, formed by

plasma oxidation of the bottom Al electrode[255, 256]. In Chapter 6, the models presented are used to calculate the temperature dependence of barrier parameters in the type-A devices, as the parabolic conductance is a typical feature of thin film MIM tunnel diodes. In the type-B devices however, the obtained conductance cannot be easily modelled, due to the observed reduction in conductance, which suggests that a resonance with surface electron states exists, which have been widely observed in scanning-tunneling-spectroscopy (STS) measurements on crystalline Au surfaces[257–262]. The position of the conductance peak in the forward characteristic, approximately 0.3 V, is consistent with numerous measurements of the tunnel current emitted from different facets of crystalline Au surfaces, and reflects the density of electron states below the Fermi energy  $E_F$  in the bottom electrode. The origin of these intrinsic surface states is understood to be due to the strong spin-orbit coupling effect and strong band bending of the surface potential of an Au crystal[263, 264], and as a consequence may be linked to the surface reactivity of Au with oxygen. The band bending causes the sp band crossing to be lowered below the Fermi energy, which leads to the formation of a surface state band occupied by electrons. A detailed theoretical and experimental analysis of these surface states has previously been performed in the context of STS measurements, where it has been shown that the surface band that participates in conduction has an energy dispersion well described by[259]

$$E(k) = E_F - E_S + \frac{\hbar^2 k^2}{2m^*}, \quad (5.40)$$

where  $E_S \approx 0.4$  eV is the bottom of the surface state band, and  $m^* \approx 0.3m_e$  is its effective mass, and can be incorporated in a simple model of the current. In Davis *et al.*'s model[259], the thermal broadening of electrons in the surface band determines the position of the peak, where in the low temperature limit the peak is expected to reflect exactly the density-of-states. In the present results it is shown that at low temperature, there is no sharp cutoff, rather the peak changes little in shape but shifts slightly to a higher energy. Despite this, their results taken at room temperature are highly consistent with those shown in Figure 5.14 (b), where the peak conductance is just under a factor of two of the conductance at  $V_a \approx 0$ , and has an identical shape and position. Davis *et al.* and others[262] have also shown that the peak shape and position in STS experiments is determined by the topography of the Au surface. The peak height obtained in the present MIM devices is consistent with emission of surface electrons from flat surfaces, where the conductance peak is highest. In light of the new developments in the field of topological insulators, the surface states on Au surfaces have been described as very similar to topological states in these novel materials, originating from the bulk properties of the band structure of Au[264]. These surface states are described as temperature insensitive and robust to disorder, however, magnetic impurities are known to cause the

surface states to disappear, which could explain why no surface states were observed in the type-A devices. There is little work that treats the existence of surface states in MIM structures, however the STS experiments in the literature provide the closest analogy to the structures in this work, including the typical tip-sample distances used on the order of 2 nm.

The non trivial part of the conductance described by the odd terms  $F_1V_a + F_3V_a^3$  can be associated with the surface state, and can be empirically accounted for to reveal the underlying conduction due to tunneling from bulk states[265], as appears to be the case with the type-A devices. This was not performed, as this contribution can be clearly seen in the background of the curves shown in Figure 5.14 (b), which has a similar temperature dependence to that shown in Figure 5.14 (a). The type-A measurements presented are representative of the measurements performed on the 2 nm, 3 nm and 5 nm devices, which all showed a qualitatively identical temperature dependence. For the type-B devices, the 1.4 nm and 2 nm devices both showed similar peaks and temperature dependence, whereas in the 3 nm devices the effect of surface states could not be observed, as the low voltage current could not be accurately measured due to the limitations of the measurement equipment, but the tunneling temperature dependence is similar to that observed in the type-A devices. The difference in amplitude between the tunneling current obtained between the type-A and B devices is discussed in the next chapter. This difference in tunneling amplitude, and the presence of surface states in the type-B devices may be linked to the slight difference in the dielectric constant trends observed between the type-A and B batches. This is discussed briefly at the end of the final chapter, which presents opportunities for further work.

## 5.4 Conclusion

In the first section it was shown that a very high consistency in the capacitive properties of the junctions is obtained with the values of thickness  $t_J$  and surface area  $A_J$  used. It was shown that as the film thickness is decreased below 5 nm, the dielectric constant of the film appears to decrease significantly and consistently across all measured devices. The C–V measurements performed suggested this change is mostly intrinsic to the  $\text{Al}_2\text{O}_3$ , since a clear trend that matches the decrease in the dielectric constant was observed that deviates from the bulk characteristic. The decrease in the dielectric constant could therefore be helpful in raising the operating temperature of a metallic SET system, through the effective reduction in  $C_J$ . In the following section, the measurements and basic determination of the tunneling characteristics were described in terms of the geometry, where a clear exponential scaling of the junction resistance was obtained. The scaling observed between 1.4 nm and 3 nm thickness is thought to not be representative of the true scaling, since the properties of the film appear to be changing

with thickness, as suggested by the evidence presented in the first section. The temperature dependence of the conductance was shown to be consistent with the reported temperature dependence of  $\text{Al}_2\text{O}_3$  MIM junctions, and some important differences between type-A and B devices were highlighted, particularly the effect of surface states on the bottom electrode surface. The high consistency obtained in both the capacitive and tunneling characteristics is promising for the mass fabrication of devices, as estimated by the scaling relations, however there are limitations in the understanding of small fluctuations in the geometry, particularly the thickness of the insulating film and the morphology of the metal surfaces, which affects the material parameters determined from models based on planar surfaces. The effects of these fluctuations is discussed in the following chapter.

# System Modelling

---

## 6.1 Introduction

The fabrication and measurement of highly consistent ultra-thin film MIM devices was presented in the two previous chapters, where it was found that a number of trends occur in both the capacitive, and tunneling properties of the junction as a function of thickness. In this chapter models are developed to investigate the important trends observed in the material properties in detail in an attempt to identify the underlying physical mechanisms responsible for the observed behaviour. The first section is concerned with the determination of the barrier parameters obtained from the I–V measurements, and provides a model that can explain the asymmetry of the direct tunneling current in MIM diodes, which cannot be explained only by the metal work function difference. This work was published in Physical Review B[106]. The temperature dependence of the tunneling current is then analysed in more detail using the proposed model, and compared with modelling results of similar MIM junctions in the literature. In the final section, the decrease of the dielectric constant is modelled simply by likening this effect to dielectric screening effects at the interfaces in the MIM junction. An in depth discussion on the possible origins of this effect is then presented in terms of various models and experiments available in the literature that have treated this effect in different electronic and material systems. The evidence obtained from the impedance spectroscopy and C–V measurements are discussed together in this final section.

## 6.2 Direct tunneling field enhancement

Despite the widespread applications of ALD deposited thin films, the relation between the surface morphology and the symmetry and linearity of the I–V characteristics of MIM devices is poorly understood, in particular in devices employing ALD-deposited ultra-thin films with dielectric thickness below 5 nm. This presents significant challenges for engineering and manufacturing of MIM devices with specific symmetry requirements in their tunneling characteristics. Effects of field enhancement on the tunneling characteristics of thicker MIM devices with large surface areas are well studied in the context of FNT[250]. Alimardani *et al.* investigated the effect of roughness of the bottom electrode [266] and concluded that the rough surface of metals deposited by evaporation can cause an inversion of the forward bias configuration due to field-enhancement at

roughness peaks, and that structures designed to be symmetric show asymmetry due to this effect. More recently Lau *et al.* have reported that the roughness-smoothing effect in the  $\text{Al}_2\text{O}_3$  ALD process[204], which leads to the top electrode having a smoother interface with the oxide than the bottom electrode, enhances the asymmetry in the I–V characteristics. In the FNT regime, effects of field-enhancement are known to play an important role as the tunnel current is much more sensitive to the shape of the potential barrier[250] than in the direct tunneling regime. However, field-enhancement can still play an important role in the direct tunneling regime of MIM devices employing ALD-deposited dielectric films, which is the subject of this work.

Two type-A MIM devices featuring a 2 nm film and with surface areas  $A_{J_1}$  and  $A_{J_2}$  corresponding to  $D_{J_1} = 40 \mu\text{m}$  (device A) and  $D_{J_2} = 80 \mu\text{m}$  (device B) diameter electrodes are studied in detail in this section. Device A is shown in Figure 4.8 (a), and was used to create the TEM sample after the I–V measurements were performed. Device B is shown in Figure 6.1 (b). These two devices were a distance of about 1 cm apart on the substrate during initial fabrication steps. In the case of device B the radius of the bottom electrode is computed from the measured perimeter of the circle, the area is then simply given by  $A_{J_2} = \pi r_b^2 = 5.366 \times 10^{-9} \text{ m}^2$ . In the case of device A there is a slight misalignment that should be accounted for by computing the area of the overlapping areas as the area enclosed by the intersection of two circles. Using the procedure outlined in Ref. [267] we obtain an area  $A_{J_1} = 1.28 \times 10^{-9} \text{ m}^2$ . This area obtained corresponds to the area of a discs with a radius of  $20.02 \mu\text{m}$ . In order to understand the features of the I–V curves in detail over the whole bias range, a model was developed that takes into account the effect of field enhancement due to asymmetric roughness and the aforementioned roughness smoothing effect. The model complements Miller *et al.*'s symmetric roughness model[268] that considers thickness fluctuations but doesn't consider field-enhancement. This model successfully reproduces the experimentally observed asymmetry and is highly consistent with the measured I–V

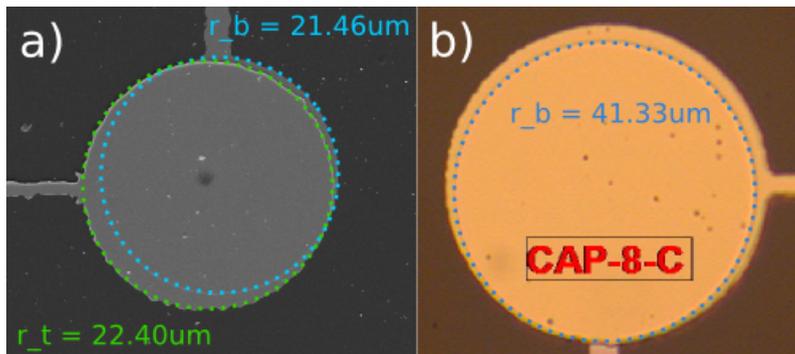


FIGURE 6.1: (a) Schematic of the cross section of a typical MIM structure. (b) Optical image of device B with a  $80 \mu\text{m}$  diameter bottom electrode, highlighting the measured radii of the bottom electrode.

characteristics over a wide bias range.

The current density as a function of the applied electric field for devices A and B is shown in Figure 6.2 in orange and blue, respectively. The current density was calculated using Equation 5.23 using  $A_{J1}$  and  $A_{J2}$ , and the applied electric field  $E_a$  was calculated using Equation 5.24 with  $t_J = 2$  nm, the measured thickness of the film. The bottom electrodes of each device are used as the ground potential. For device A shown in Figure 6.1 (a), the area was obtained by finding the offset between the circle centres and computing the area intersected by these circles. For device  $D_{J2}$  the area  $A_{J2}$  is defined by the top electrode area. The overlap of the two curves in the current density plot shows that the consistency between these two devices is very high despite using a simple parallel plate geometry. This implies that the typical length scale of the geometrical imperfections on the electrode surface that impact the I–V characteristics must be much smaller than the electrode surface area of both devices. Although the J–E curves of other devices do not overlap as well as shown in Figure 6.2, in Chapter 4 it was shown that there is a clear scaling of the junction resistivity with thickness, which implies that in general, irregular imperfections are not present in these working devices. Such imperfections are likely to be responsible for the observed open and short circuits detected in certain devices, which are extreme cases of the resistivity.

Close inspection of the TEM images presented in Figure 4.9 reveals two distinct types of surface features. Figure 6.3 (a–b) show magnified regions of the TEM image presented in Figure 4.9 (d). Figure 6.3 (a) shows a highly planar region in the MIM structure, where it can be clearly seen that the roughness on the grain surface is atomic in scale. Figure 6.3 (b) an image of a region with a protruding sharp peak outlined by the dotted yellow line and where the contrast was increased to highlight the  $\text{Al}_2\text{O}_3$  region and the morphology of the interface. The Au region overlapping the  $\text{Al}_2\text{O}_3$  is clearly a thin three-dimensional structure, as the amorphous signature and the brightness of the  $\text{Al}_2\text{O}_3$  region appears around it.

Roughness smoothing during the deposition of amorphous layers is a well-known phenomenon. This effect is clearly illustrated in a TEM image presented by Lau *et al.* (Figure 7 in Ref. [269]) of a symmetric device with a much thicker amorphous insulator than discussed here, which shows a sharp peak on the bottom electrode. Importantly, the thickness of the insulator appears to increase near the pits to the sides of the peak on the bottom electrode in their image, and thickness uniformity normal to the peak apex is not clear. In our devices, the thickness of 2 nm measured at the peak apex appears consistent with that measured in the planar regions. However due to the small thickness of our film compared to that in Lau *et al.*'s device, it is very difficult to measure accurately the thickness fluctuations. For clarity we have outlined the interfacial boundaries in Figure 6.3 (b) with dotted yellow lines.

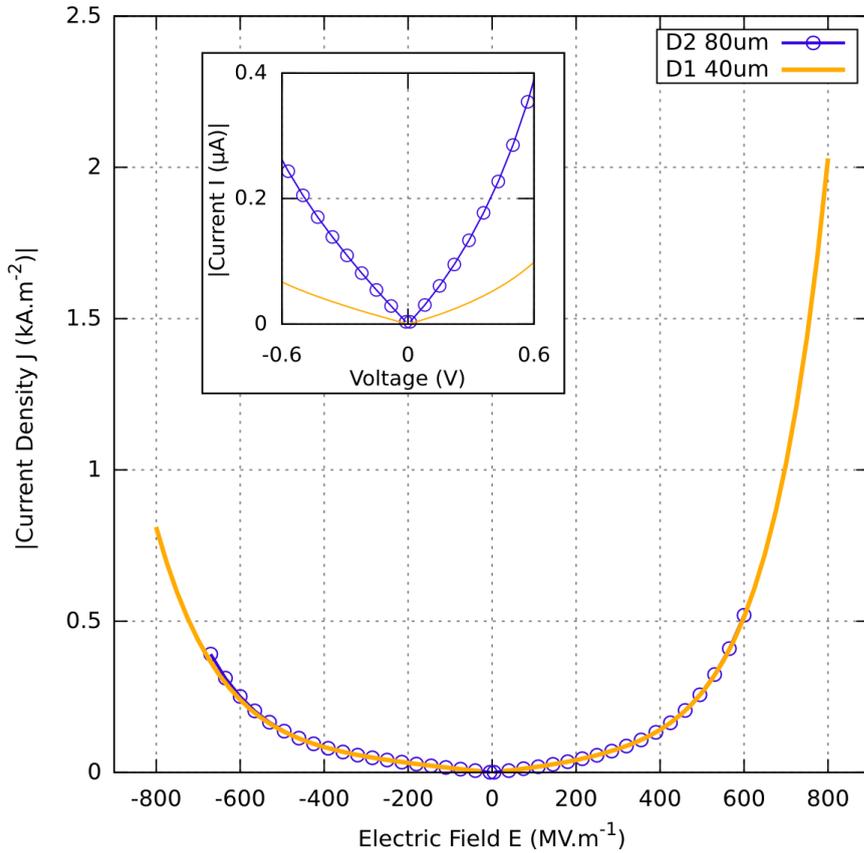


FIGURE 6.2: Current density ( $J$ ) versus applied electric field ( $E_a$ ) curves for devices A and B at room temperature. The inset shows the I-V characteristics of the same devices between  $-0.6$  V and  $0.6$  V.

The smoothing effect can be understood as a result of the finite size of the TMA precursor molecules and the steric-hindrance-limited adsorption[98] that leads to the rounding of peak and pit features of the surface at the nanoscale, leading to thicker values of the film normal to pit features. We note that an increased thickness of the dielectric near the pit regions would decrease the tunneling probability, and the electric field would be more uniform than around peaks. Therefore, it is likely that in rough regions of the electrode small thickness fluctuations due to pits lead to a smaller current than expected in MIM devices with an insulator of uniform thickness. Furthermore, since the ALD deposited film forms a ‘mould’ for the surface of the top electrode, the rounded pits cause the top electrode to have smoother, slowly varying features rather than sharp crystal plane boundaries. The consequence of this is that field-enhancement due to sharp features is stronger on average near the bottom electrode surface compared to the top electrode surface.

A finite element model was developed using COMSOL to simulate and study the effect of sharp features on the bottom electrode such as those shown in Figure 6.3 (b). We consider a peak geometry that can be constructed using a two-dimensional model

with a rotational symmetry axis as illustrated in Figure 6.4 (a), where  $r$  represents the peak apex radius,  $\theta$  the inclination angle of the surface leading to the apex, and the green dashed line indicates the axis of rotational symmetry, which is also the path along which the potential is sampled. The dielectric material is assumed to contain no free or trapped charges. For a peak radius of approximately  $r = 0.2$  nm, as for the peak highlighted in Figure 6.3 (b), the potential profiles obtained from applying a potential difference of 1 V between the top and bottom surfaces and varying the slope angle  $\theta$  are shown in Figure 6.4 (b) (solid lines). In the finite element model, the bottom electrode is chosen as the reference potential and is set to 0 at position  $x = 0$ , corresponding to the peak apex, and the top electrode potential is set to 1 V at  $x = 2$  nm. It can be seen that the potential becomes highly non-linear as  $\theta$  is increased. We note that increasing the peak radius  $r$  reduces the non-linearity only slightly as long as the radius is smaller than the film thickness, and that in this regime the non-linearity in the potential extends

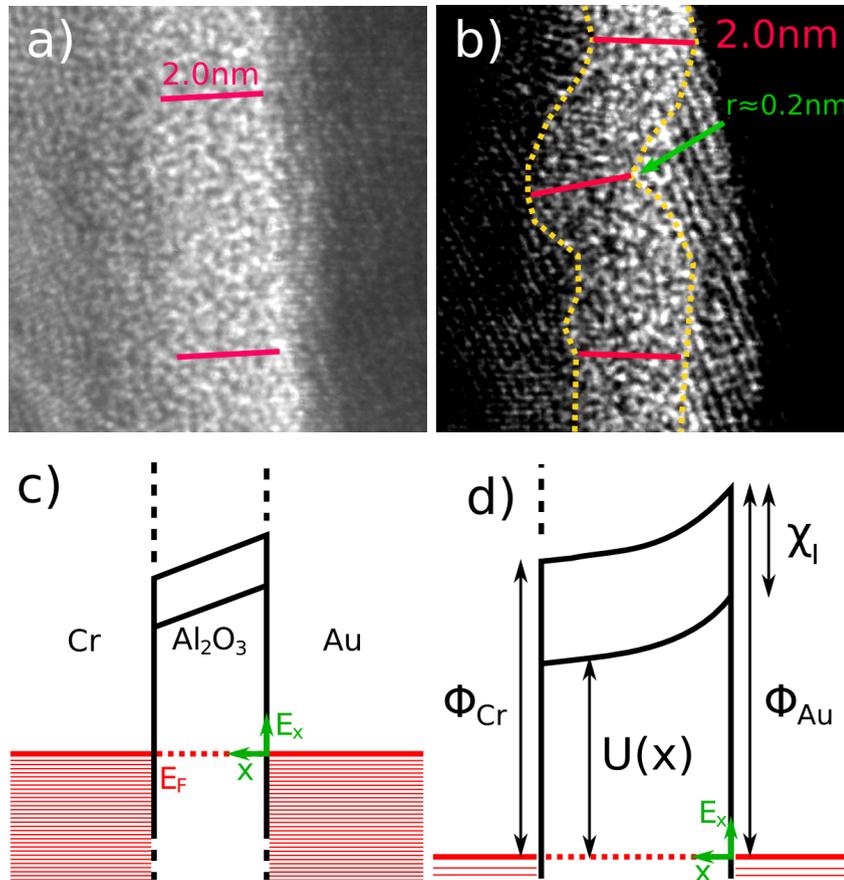


FIGURE 6.3: (a) Flat and (b) rough regions highlighted from Figure 4.9 (d). The contrast in (b) was enhanced to distinguish the  $\text{Al}_2\text{O}_3$  region shaded by the Cr more clearly. (c) and (d) Energy band diagrams corresponding to the situations shown in panels (a) and (b) with a uniform and distorted potential barrier, respectively. The metal work functions are denoted by  $\Phi_{\text{Au,Cr}}$ , the insulator electron affinity by  $\chi_I$  and  $U(x)$  is the potential energy in the insulator.

far into the dielectric layer.

The spatial dependence of the potential in this model is well described by an arbitrary function of the form

$$\varphi(x) = \Delta\varphi_T \ln \left( 1 + \frac{e-1}{t^n} x^n \right) \quad (6.1)$$

where  $t$  is the film thickness,  $n$  an arbitrary parameter,  $e$  is Euler's number, and  $\Delta\varphi_T$  is the potential difference applied in the finite element model. The approximation of the potential obtained via the finite element model by this function are represented by the open symbols in Figure 6.4 (b). Physically realistic results are obtained when the parameter  $n$  is constrained by  $n \leq 1$ . When  $n$  becomes smaller than 0.5 the potential drops very rapidly near the peak apex which further restricts the parameter range in terms of physically realistic situations. The parameter  $n$  effectively represents the sharpness of the peak for small values of  $r$  (compared to  $t$ ) and  $\theta > 10^\circ$ . For large  $\theta$ , i.e.  $\theta > 40^\circ$ , the accuracy of the approximation decreases, however, it can still be employed for qualitative considerations. In thicker tunnel barrier systems (of the order of 5 nm for  $\text{Al}_2\text{O}_3$ ) where the breakdown voltage is high enough, Fowler-Nordheim field emission dominates at voltages exceeding the barrier height, and the barrier appears triangular in shape to tunneling electrons. In this regime the precise shape of the potential barrier is very important, so field-enhancement must be considered as it significantly affects the thickness of the barrier[250] as well as its average height. The approximation given above may still be used in those cases but it may not be sufficiently accurate for a precise study of the Fowler-Nordheim characteristics at high voltages.

Equation 6.1 must be modified to account for the work function difference of the electrodes. The function is then expressed as

$$\varphi(x) = \Delta\varphi \ln \left( 1 + \frac{e-1}{t^n} x^n \right) \quad (6.2)$$

where  $\Delta\varphi$  is the internal potential difference of the junction, described by

$$\Delta\varphi = V_a + \frac{\Phi_L - \Phi_R}{e} \quad (6.3)$$

where  $V_a$  is the applied voltage, and  $\Phi_L$  and  $\Phi_R$  represent the left and right metal electrode work functions respectively. The spatial dependence of the potential energy in one dimension that includes any field-enhancement effects in the simplest case is then given by

$$U(x) = \Phi_L - \chi_I - e\varphi(x) \quad (6.4)$$

where  $\Phi_L$  is the left metal work function,  $\chi_I$  the insulator electron affinity, and  $\varphi(x)$  the electric potential at position  $x$  in the insulator. The work function values of these metals are widely reported; here we use the values  $\Phi_{\text{Cr}} = 4.5$  eV[223, 224] for the right

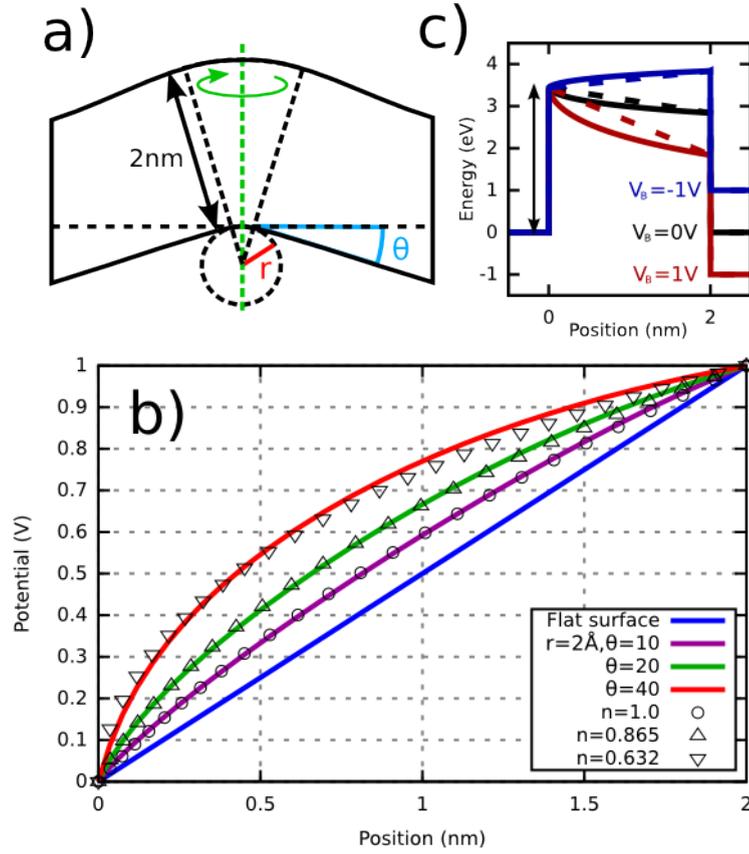


FIGURE 6.4: (a) Geometrical model used to represent a typical peak on the bottom electrode, with the apex radius  $r$  (red), the inclination angle of the surface  $\theta$  (blue), and the axis of rotational symmetry highlighted (green dashed line). (b) Potential profiles obtained for  $r = 0.2$  nm as a function of the inclination angle (solid lines). Least-square fits of Equation 6.1 to the potentials obtained from COMSOL finite element modelling are shown as open symbols. (c) Plot showing the voltage dependence of the potential barrier using Equation 6.4 and  $n = 0.63$ , which corresponds to  $r = 0.2$  nm and  $\theta = 40^\circ$ , in Equation 6.1. The dashed lines represent a uniform barrier with the same material parameters. The material parameter  $\chi_I = 1.66$  eV is taken from the fitting results (discussed below).

electrode, and  $\Phi_{Au} = 5.1$  eV[219] for the left electrode, corresponding to atomically clean surfaces in an abrupt contact with vacuum or an insulating material.

The voltage dependence of the potential energy barrier is shown in Figure 6.4 (c), which includes the potential due to the work function difference. The dependence of the potential on the position inside the insulator considering specific surface features has been studied by others in detail[171, 250, 270]. Gaillard *et al.* numerically extracted interfacial roughness profiles of the top and bottom electrodes of a 45-nm-thick dielectric MIM structure using high resolution TEM images to simulate the electric field[203]. The roughness profiles in their devices are asymmetric owing to the amorphous structure of  $Ta_2O_5$  deposited by metal-organic chemical vapour deposition (MOCVD). Their simulation results show that the electric field strength near both the interfaces can be significantly larger at protrusions but that on average the field is slightly stronger near

the bottom electrode. They concluded that the asymmetry in the leakage current is possibly correlated with the asymmetric roughness profile of the interfaces. Below, we show that the model of the potential introduced above indeed yields asymmetric I–V characteristics and that in our case the data is consistent with the interpretation of asymmetric roughness.

Figure 6.3 (c) represents the ideal trapezoidal potential barrier[271] found where regions of the barrier are highly planar such as shown in Figure 6.3 (a), and Figure 6.3 (d) represents an example of an approximated field-enhanced potential barrier expected in regions where the bottom electrode comprises sharp features such as the one shown in Figure 6.3 (b). Given the large surface area of our devices, there will be variations in the fine surface features that will influence the field structure[203], but qualitative behaviour at the sharpest peaks is very similar to that produced by Equation 6.1 where the electric field approaches a constant value far away from a peak apex. As this approximation is applied to the description of the potential barrier it is possible to use a number of different tunneling models. Here we chose to use the Wentzel-Kramers-Brillouin-Jeffreys (WKBJ) description of the transmission probability as it is widely applied to MIM structure characterisation and is the simplest model available.

Using Equation 6.4, the standard WKBJ approximation[249] applied to the transmission probability of incident electrons yields

$$P_{\pm}(E_x) = \exp \left\{ -\frac{2}{\hbar} \int_0^t \sqrt{2m_I(U(x) - E_x)} dx \right\} \quad (6.5)$$

under forward bias (+) when  $V_a > 0$ , where electrons incident on the barrier at  $x = 0$  from the left ( $x < 0$ ) are considered, and under reverse bias (–) when  $V_a < 0$ , where electrons incident on the barrier at  $x = t$  from the right ( $x > t$ ) are considered.  $m_I$  is the effective mass and  $E_x$  is the energy of incident electrons normal to the surface, measured relative to the Fermi level of the left electrode. The current can now be calculated by adding the contributions from all regions featuring a uniform dielectric layer, and thus with a trapezoidal barrier, and contributions from all regions featuring sharp peaks in the dielectric layer, and thus with a distorted barrier. The current densities for the individual regions are calculated using the one-dimensional version of Harrison’s tunnel current density equation[272],

$$J_{\pm} = \frac{2e}{h} \int_{-\infty}^{\infty} P_{\pm}(E_x) [f(E_x) - f(E_x + eV_a)] dE_x \quad (6.6)$$

where transverse wavevectors are ignored, and where  $f(E)$  is the Fermi-Dirac function, defined relative to the Fermi level of the left electrode. The sign of the current is determined by the sign of the applied voltage  $V_a$ . Assuming  $T = 0$  K, the integration in equation 6.6 can be performed from  $-eV_a$  to 0 under forward bias (when  $V_a > 0$ ), and

from 0 to  $-eV_a$  in reverse bias (when  $V_a < 0$ ). In this model, the bandstructure in the metals and the insulating region is first assumed to be isotropic, such that the effective mass is the same in all directions in reciprocal space. The usual method is then to consider the three dimensional wave vector  $\vec{k} = \vec{i}k_x + \vec{j}k_t$  of an electron in each metal in a cylindrical coordinate system, where  $k_x$  represents the electron momentum normal to the metal surface and  $k_t$  represents transverse momentum parallel to the metal surface. The total energy of free electrons in the metals is then written as[273]

$$E_T = \frac{\hbar^2}{2m_e}(k_x^2 + k_t^2) = E_x + \frac{\hbar^2 k_t^2}{2m_e}. \quad (6.7)$$

The transverse energy component therefore appears in the Fermi Dirac function in this form and can be integrated to yield a different expression for the current. In the present model, it is assumed that electrons with total momentum off the normal to the metal surface have a negligible transmission probability, since the barrier appears thicker. Then  $k_t = 0$ , which leads to  $E_T = E_x$  so that only electrons normal to the surface are considered. The three dimensional form of the current density equation was investigated, however the full numerical integration in this form was found to take much too long, and the result was sensitive to the choice of integration limits. Using Equation 6.5, the total current is now given by

$$I = A_i J_i + A_d J_d \quad (6.8)$$

where  $A_i$  and  $A_d$  are the surface areas of the uniform regions and regions with sharp features respectively, which are constrained by  $A_J = A_i + A_d$ , where  $A$  is the total area.  $J_i$  and  $J_d$  are the current densities associated with uniform and sharp regions, respectively, which are evaluated using Equation 6.6. The parameters  $\chi_I$ ,  $m_I$ ,  $A_i$ ,  $A_d$  and  $n$  can now be determined by fitting Equation 6.8 to the experimental I–V characteristics of the different MIM devices.

It is worth noting that this model can not only be used to describe the effect of field-enhancement due to geometrical features, but also for field-enhancement induced by a local change in permittivity of the film. Under the assumption that there is no free or trapped charge, geometry-induced field-enhancement produces a non-linear flux density and electric field that are related by  $D(x, V_a) = \epsilon_c \xi(x, V_a)$ , where  $D$  and  $\xi$  share the same dependence on  $x$ , and  $\epsilon_c$  is the constant permittivity of the dielectric. The equivalent effect can be achieved by using a position-dependent permittivity, that is  $D(V_a) = \epsilon(x)\xi(x, V_a)$ . The two interpretations can be linked through a well-behaved scaling function  $f(x)$  (for example, the derivative of Equation 6.1), which in the geometrical case describes  $D(x, V_a) = D(V_a)f(x)$  and  $\xi(x, V_a) = \xi(V_a)f(x)$ . The equivalent effect on the field can be achieved by a permittivity gradient  $\epsilon(x) = \epsilon/f(x)$  which shows that the two forms are equivalent when the flux density is assumed constant in the dielectric layers.

Therefore it is possible to use geometry as well as changes in permittivity to engineer the asymmetry properties of ultra-thin film MIM diodes, as has been shown extensively, for example, by sequentially depositing different dielectric materials *in-situ* by ALD in stacked layers[274, 275] or to produce alloyed materials with new properties[216].

It is also possible that an imperfect chemistry in the first few cycles of the ALD growth may cause a permittivity gradient in certain regions due to the presence of unreacted or adsorbed compounds, which are expected to lead to optical phonons in the far-IR range. In turn, such phonons would lead to inelastic scattering of electrons, however, we could not observe any signs of inelastic scattering by computing the second derivative of the I–V curve[276] taken at 3.7 K. We note that this could be due to the low probability of transmission of electrons in our system and insufficient adsorbates or trapped organic compounds. The non-linearity of the I–V curve at low voltage is a feature of a tall energetic barrier even in thinner films[275], which causes the determination of small peaks in the second derivative of the I–V curve to be very difficult.

The results of fitting Equation 6.8 to the I–V data presented in Figure 6.2 are shown

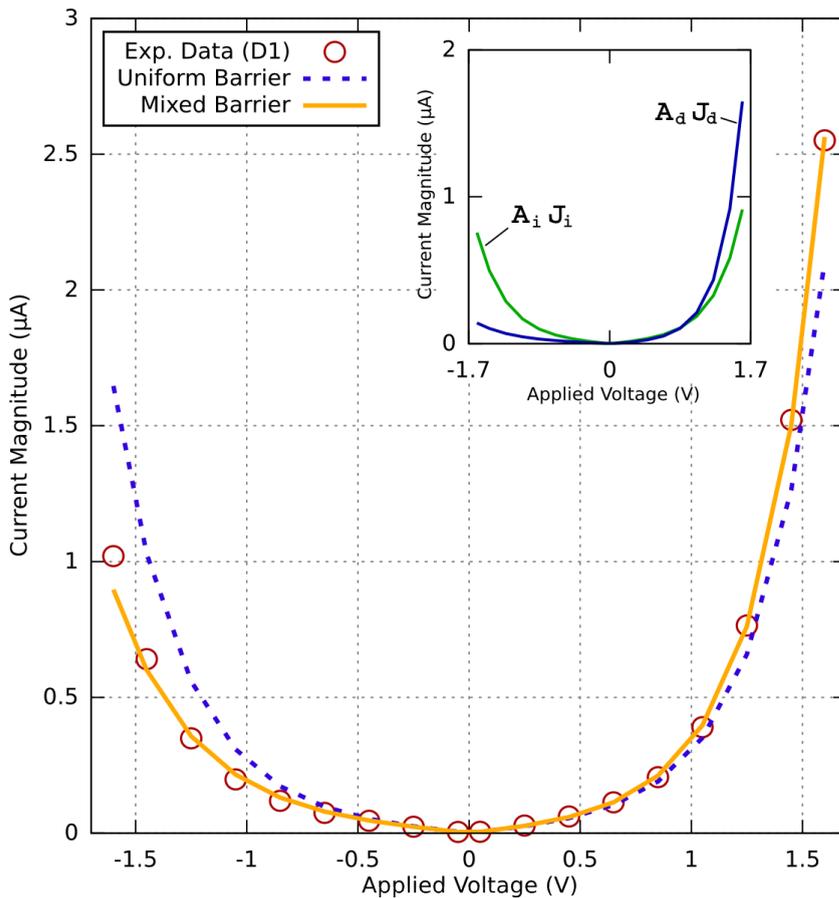


FIGURE 6.5: Fitting results of the mixed barrier model described by Equation 6.8 (solid orange line) and the uniform barrier model where  $A_d = 0$  in Equation 6.8 (dashed blue line), to the experimental data for device A (open circles). The inset shows the two contributions to the total current.

TABLE 6.1: Fitting parameter results from using Equation (6.8).

Parameter	Value	Uncertainty
$A_i$	73% of $A$	$\pm 26\%$
$A_d$	27% of $A$	$\pm 69.5\%$
$m_I$	$0.69m_e$	$\pm 3.1\%$
$\chi_I$	1.75 eV	$\pm 5.2\%$
$n$	0.83	$\pm 25.2\%$

TABLE 6.2: Estimated correlations between the parameters used in the fitting process.

Parameter	$m_I$	$n$	$A_i$
$m_I$	-	0.967	-0.941
$n$	0.967	-	-0.994
$\chi_I$	0.999	0.964	-0.935

in Figure 6.5 and the parameters obtained are summarised in Table 6.1. The dashed line shows the best fit ( $R^2 = 0.661$ ) obtained when assuming a fully uniform dielectric with no sharp features, ie  $A_d = 0$ , for which we obtain the values  $m_I = 0.75$  and  $\chi_I = 2.03$  eV. Not surprisingly, the uniform barrier model is unable to account for the asymmetry in the I–V data as the work function difference alone causes only a negligible effect in the direct tunneling regime. However, when considering the effects of sharp features on the bottom electrode, the asymmetry is enhanced by the combined effect of the work function difference and field-enhancement. This mechanism is illustrated in Figure 6.4 (c) where the shape of the barrier at  $V_a = 0$  V (black line) is distorted due to the non-zero built-in potential. The best fit ( $R^2 = 0.990$ ) to the model including both, regions where the dielectric layer is highly uniform, as well as regions where it comprises sharp features, is represented by the solid line in Figure 6.5. The inset shows the individual current components from Equation 6.8. As can be seen, the asymmetry is significantly increased in field-enhanced regions, which is of interest for the design of certain MIM devices, for example IR rectennas based on MIM diodes. The uncertainty (Table 6.1) and the correlations (Table 6.2) in the parameters were estimated from the covariance matrix[277]. To ensure that a wide range of the parameter space is probed during fitting and to ensure the covariance matrix is representative, the starting values of the parameters were chosen to be far from the resulting optimal solution.

The correlations between the parameters obtained from the fitting are shown in Table 6.2 where a correlation of 1 between two parameters means that a change in one parameter is entirely compensated for by a positive linear change in the other. A value of  $-1$  means the same except that changes in the two parameters are negatively linear to each other. The correlation is highest between the effective mass  $m_I$  and the affinity  $\chi_I$ , which is well known to cause uncertainty when fitting I–V characteristics in a limited bias voltage range[268]. Despite this, we obtained a much smaller uncertainty in these

parameters than with others due to the I–V curve spanning a wide voltage range so that there is an optimal solution that is stable. The reason for this can be explained by considering the squared integrand of Equation 6.5,  $2m_I(\Phi_L - \chi_I - e\varphi(x) - E_x)$ . The relationship between  $m_I$  and  $\chi_I$  can then be expressed as

$$m_I = \frac{C_1}{2(\Phi_L - \chi_I - e\varphi(x) - E_x)} = \frac{C_1}{2(C_2 - \chi_I)} \quad (6.9)$$

where  $C_1$  is a given value of the squared integrand and  $C_2 = \Phi_L - e\varphi(x) - E_x$ . For large values of  $C_2$  compared to  $\chi_I$ , the dependence of  $m_I$  on  $\chi_I$  is quasi-linear as  $\chi_I$  is far away from the singularity point  $\chi_I = C_2$ . As the voltage is increased, at the point  $x = t$  and for  $E_x = 0$ , the value of  $C_2$  decreases bringing the singularity point closer to the origin which decreases the linear correlation between  $m_I$  and  $\chi_I$ . With the highest voltage value used of  $V_a = 1.6$  V, we find  $C_2 = 2.9$  eV under forward bias which is a drastic shift from  $C_2 = 4.5$  eV at  $V_a = 0$  V. The other important correlation is between  $n$  and  $A_i$  (and  $A_d$ ) which is not surprising as increasing  $n$  necessarily implies a decrease in  $A_i$  (and increase in  $A_d$ ). The remaining correlations are less important in particular when  $A_d$  is small compared to  $A_i$ , as is illustrated above where the values of the effective mass and electron affinity are not significantly affected.

The values found for the effective mass are within the range of reported experimental values, i.e. between 0.23 and 0.8 for amorphous ALD deposited  $\text{Al}_2\text{O}_3$ [181, 247, 251, 266, 278], and theoretical predictions with values around 0.4[252, 253]. The values found for the electron affinity of  $\text{Al}_2\text{O}_3$  are also within the range of reported values of 1.0 eV to 2.5 eV[100, 219, 239]. Despite the very high correlation between effective mass and electron affinity, due to the I–V curve spanning a wide voltage range, there is an optimal solution that is stable for both the uniform and mixed barrier models, which is reflected in the small uncertainty in these parameters.

The value of  $n$  obtained is near the value obtained for the potential profile normal to the apex of a peak with  $\theta = 20^\circ$  and  $r = 0.2$  nm, however similar potential profiles, and thus similar values of  $n$ , are obtained with different combinations of  $\theta$  and  $r$ . Figure 6.6 (a) shows the dependence of  $n$  on the possible combinations of  $\theta$  and  $r$  of interest. The values of  $n$  were obtained by fitting Equation 6.1 to the potential profiles obtained through finite element modelling by varying the value of  $r$  from 0.1 nm to 2.0 nm in steps of 0.1 nm, and the value of  $\theta$  from  $0^\circ$  to  $45^\circ$  in steps of  $5^\circ$ . When  $n$  reaches 1, the potential is no longer well described for decreasing values of  $\theta$  by a single parameter and an additional parameter may be added to Equation 6.1 that extends its applicability into small  $\theta$  and large  $r$  values. As can be seen, if  $\theta$  is large the value of  $n$  is consistently smaller than 1 even for large  $r$  values, which represents a significant distortion in the potential. Figure 6.6 (b) shows the same data as panel (a) but here  $\theta$  is shown as a function of  $r$  for a range of selected discrete values of  $n$ . From this plot it is clear that

a certain  $n$ , for example as obtained in the fit to the I-V curve, represents an infinite number of  $(r, \theta)$  duplets.

We note that the value found for the area  $A_d$  appears to be large compared to the evidence revealed by the TEM images. This suggests that isolated sharp peaks are unlikely to dominate the field-enhancement. In contrast, sub-nanometer-scale roughness, which may extend over a large area thus yielding a higher value of  $A_d$ , may play an important role. Since the area imaged in the TEM sample is very small compared to the area of the junction,  $A_d$  cannot be easily connected to a peak surface density without a detailed knowledge of the surface-roughness over a statistically significant area. As a result, the uncertainty in the parameters  $A_i$ ,  $A_d$  and  $n$  is large. The value of  $n$  can

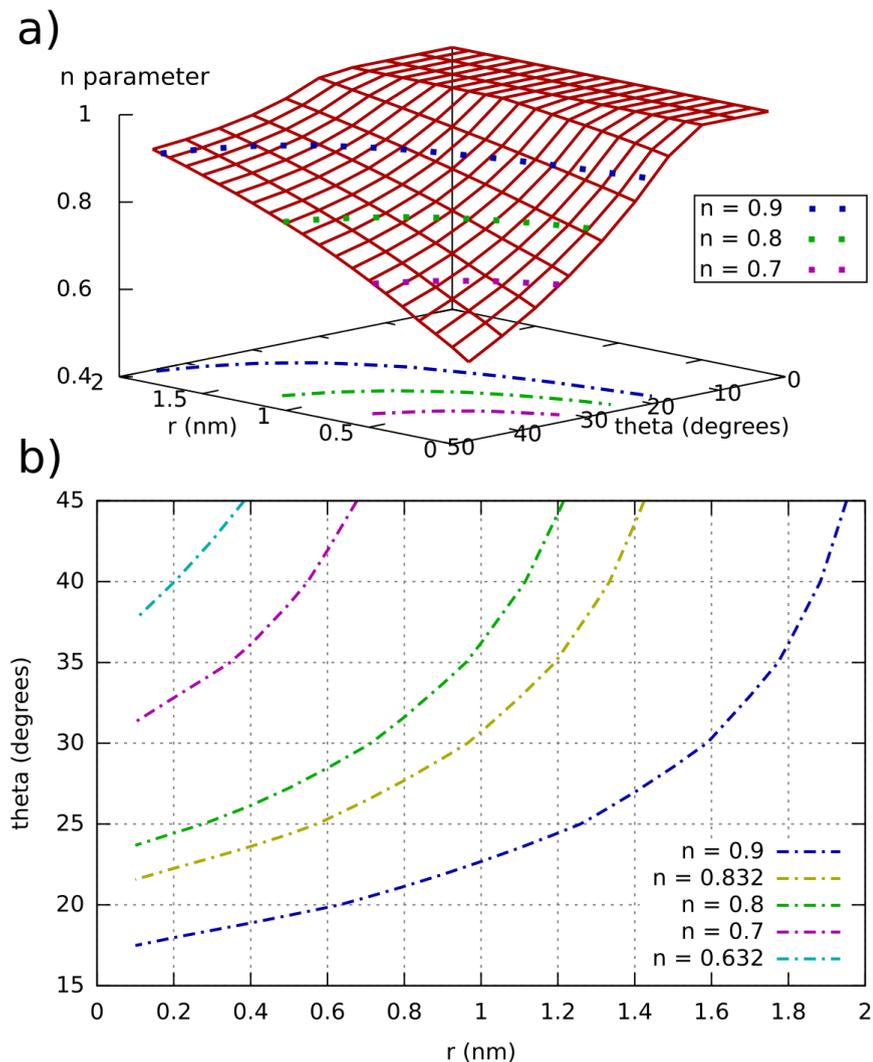


FIGURE 6.6: (a) Parameter  $n$  from Equation 6.1 as a function of  $r$  and  $\theta$  obtained by fitting Equation 6.1 to the data obtained via finite element modelling. The dot-dashed lines indicate the projections of the corresponding dotted lines onto the  $r$ - $\theta$  plane for selected values of  $n$ . (b) The same data as in panel (a), but showing  $\theta$  as a function of  $r$  for a range of fixed  $n$  values.

indeed be set to a different value and a value for  $A_i$  and  $A_d$  can be found that yields a similar solution. For example, fixing  $n = 0.63$ , corresponding to the case identified in the TEM image, causes  $A_d$  to be reduced to 14.3% of  $A$ , while both  $m_I = 0.66m_e$  and  $\chi_I = 1.66$  eV remain almost unchanged, yielding a solution with an  $R^2$  value of 0.989, hardly discernible from the optimal fit. We note that this uncertainty is due to the limited voltage range accessible in this experiment; the differences between each solution increase at very high voltages as the FNT regime ( $eV > \Phi_L - \chi_I$ ) is approached.

It is, in fact, possible to define a non-linear potential barrier over the whole area of the junction (i.e.  $A_i = 0$ ) which yields a solution with an  $R^2$  value only slightly smaller than for the optimal case. The potential obtained is close to that of a uniform barrier and corresponds to the field-enhancement resulting from sub-nanometer scale roughness. Taken together these observations suggest that a more complete ‘mosaic’ field-enhancement model of the MIM junction may be described as

$$I(A, V_a) = \sum_{j=1}^N A_j J_j(V_a, n_j) \quad (6.10)$$

where  $N$  is the number of regions with a uniquely defined potential described by  $n_j$  covering an area  $A_j$ . While this approach is unlikely to produce unambiguous distributions of different potential profiles to model the I–V characteristics of MIM diodes operating in the direct tunneling regime, it may be expected to produce valuable information near or in the FNT regime.

We now discuss the elements of the distribution of sharp features that dominate the field-enhancement to shed more light on the significance of  $n$  and  $A_d$ . We have established that a greater quantity of sharp features on the bottom electrode is due to the uniform and conformal growth of  $\text{Al}_2\text{O}_3$  by ALD, the shape of the features on the surface and the roughness smoothing effect that occurs at the nanoscale. Although roughness smoothing is not unique to ALD, it is different to a natively grown oxide, for example, where the surface features of the bottom electrode are not preserved during oxidation by oxygen atmosphere or plasma, which leads to a more gradual interface. As the bottom electrode surface in our devices is Au, being a noble metal, the sharpest features are not altered during ALD growth and it is therefore reasonable to expect a distribution that includes atomically sharp peaks and ridges, with an abrupt interface. Figure 6.7 shows surfaces with different peak geometries that cause varying degrees of net field-enhancement near the bottom electrode. When considering isolated three-dimensional peaks such as illustrated in Figure 6.7 (c), the base of the peak (highlighted in light blue) forms a ridge structure on the top electrode interface (dashed light blue) with the  $\text{Al}_2\text{O}_3$  layer, characterised by an increased radius due to the roughness smoothing. This leads to field-enhancement occurring near the top electrode in these regions, although of

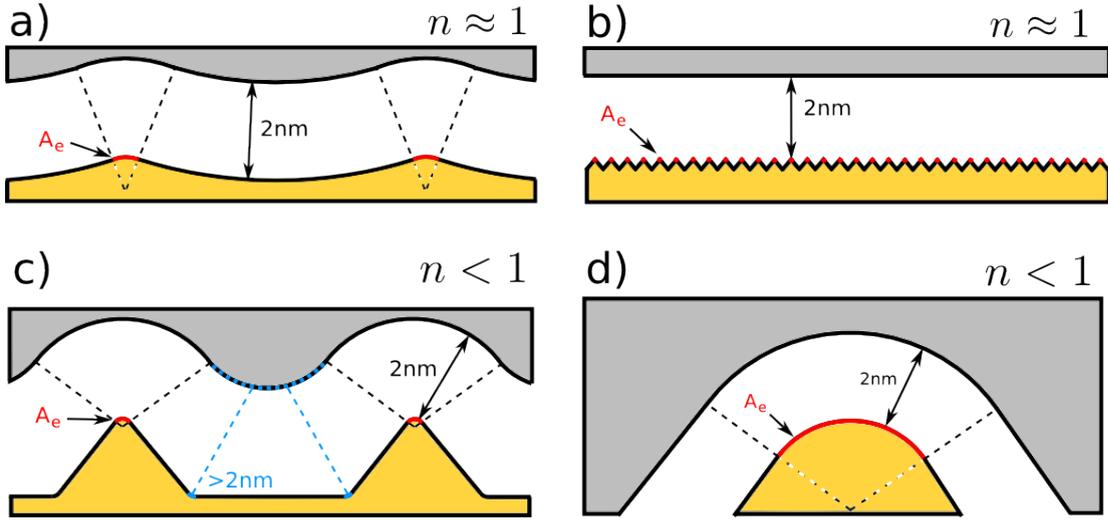


FIGURE 6.7: Schematic representations of structures that represent areas that have a higher net field-enhancement near the bottom compared to the bottom electrode. The area  $A_e$  from Equation 6.11 is highlighted in red. The uniform and conformal nature of the deposited film and the roughness smoothing effect are represented but not to scale. (a) Wave-like structure characterised by large values of  $n$  with little to no net field-enhancement near the bottom electrode. No roughness-smoothing occurs in this idealised case. (b) Sub-nanoscale asymmetric roughness over a large area that, as a result of the density of the sharp peaks, is characterised by a large value of  $n$ . This is a direct consequence of the roughness smoothing effect [203, 204]. (c) Isolated sharp peak structures characterised by small values of  $n$  and small area  $A_e$ . The roughness smoothing effect occurs at the base of sharp peaks. (d) An isolated, relatively blunt peak with a large area  $A_e$  and small value of  $n$ .

much smaller amplitude due to the increased thickness of the film and increased radius of the emitting surface. Similarly, large wave-like features characterised by small values of  $\theta$  ( $< 10^\circ$ ) and large values of  $r$  ( $> 0.2$  nm) such as shown in Figure 6.7 (a) that extend over a larger scale than the thickness of the film will cause a slight field-enhancement from both the top and bottom electrodes, which will mostly cancel out. In Figure 6.7 (b) the asymmetry in the sub-nanometer roughness is represented by a simple diagram. This causes stronger field-enhancement near the bottom electrode on average compared to near the top electrode, as shown by Gaillard *et al.* [203]. This corresponds to a situation where there are many peaks densely packed together such that the effective field-enhancement is significantly reduced compared to isolated peaks but extended over a potentially large portion of the total area.

The area  $A_d$  can be described in terms of the values of  $r$  and  $\theta$  by using the model shown in Figure 6.4 (a). The area of the spherical cap illustrated in Figure 6.7 is given by

$$A_e = 2\pi r^2(1 - \cos(\theta)) \quad (6.11)$$

and the potential normal to this area on the spherical surface does not change significantly. Therefore to a first approximation  $A_d$  can be written as the sum of all the contributing instances of  $A_e(r, \theta)$  from the bottom electrode, assuming that all the other contributions are small in comparison and hence can be neglected. We can see from Equation 6.11 that the area increases significantly with larger  $r$  values and similarly with increasing  $\theta$ . Choosing the largest value of  $A_e$  in our distribution, i.e.  $r = 1.4$  nm and  $\theta = 45^\circ$  obtained from the contour in Figure 6.6 (b), the two-dimensional density of peaks is given by  $n_p = A_d/AA_e = 7.68 \times 10^{16} \text{ m}^{-2}$  or one peak per 3.61 nm by 3.61 nm area of the surface when  $A_d = 0.27A$ . Such a high density of sharp features is not borne out of the TEM results. However, we note that it is based on a coarse approximation, which in the context of a constant thickness of the dielectric film, depends on a number of contributions that will affect the result. In order of decreasing importance the following considerations will affect the resulting density of peaks: (i) the distortion in the potential extends far beyond the area  $A_e$  for isolated sharp peaks (thus increasing  $A_e$  significantly); (ii) sub-nanometer-scale densely packed peaks that extend over a large area[203] (thus effectively increasing  $A_e$ ); (iii) sharp features on the surface of the top electrode which cancel out contributions from sharp features on the bottom electrode (increasing  $A_d$ ); and, (iv) corrections to the underestimation of  $A_J$  due to the assumption of a perfectly flat surface[279] (which affects all free parameters).

Furthermore, in the direct tunneling regime, the current density is strongly dependent on the thickness of the potential barrier, but fluctuations in the thickness of the film have not yet been considered. The effect of changes in thickness on the other parameters is illustrated if we assume that the thickness normal to a peak apex is reduced by 10%, i.e. 1.8 nm. If  $n$  is kept at 0.832, we find that  $A_d$  is reduced to 2.5% of  $A_J$ ,  $m_I = 0.73$  and  $\chi_I = 1.85$  ( $R^2 = 0.992$ ). With the value of  $A_d = 2.5\%$  of  $A_J$  and following the contour for  $n = 0.83$  as before yields a peak density of  $n_p = 7.041 \times 10^{15} \text{ m}^{-2}$  or one peak per 12 nm by 12 nm area of the surface, showing that thickness fluctuations of the film play an important role even with Å-scale fluctuations. It is very challenging to quantify such small fluctuations experimentally, in particular over a large areas. In contrast, it is possible to include thickness fluctuations in our model using Miller *et al.*'s method[268], which assumes a Gaussian distribution  $\alpha(t)$  in the film thickness, such that  $A(t) = A_J\alpha(t)$  is the portion of the total area with thickness  $t$ . However, experimentally identifying the parameters to describe the thickness distribution is non-trivial.

It is important to note that the effect of such small features in the surface geometry is unlikely to play a similarly important role in MIM structures comprised of electrode materials with a small electron density such as weakly doped semiconductors. The Fermi energy of electrons in metals is normally high, for bulk Au between 5 and 10 eV above the valence band edge[264, 280–283] and for Cr on the order of 7 eV[284, 285]. At kinetic energies above 1.5 eV for electrons, the de Broglie wavelength is smaller than 1

nm and, as the electron density is high, the probability of incident electrons to ‘see’ the fine features in the surface topography is large.

An asymmetry in the incident electron energy of the electrodes may cause a slight asymmetry in the I–V characteristics, however, the difference in Fermi energy required is of the order of at least 5 : 1 as has been pointed out by Brinkman *et al.*[286]. In our system the Fermi energies of both Au and Cr are comparable and may contribute to the asymmetry but the effect is small. However, a detailed analysis of the band structure of the metals would be required as there are multiple sub-bands in the metal valence band that each present wavefunctions with different symmetries that may change with energy[264]. The Fermi energies of s-like sub-bands should dominate the tunneling current due the spherical character of the wavefunction, however other sub-bands such as the prominent d-band in Au[287] may play a role.

Finally, it is emphasised that the high consistency obtained in these devices is due to the optimal conditions for the deposition of the bottom metal electrodes and the compatibility of these electrodes with the ALD process, which allows highly uniform and conformal growth of high quality  $\text{Al}_2\text{O}_3$ . The minimal surface roughness obtained first reduces the chances the devices break down under a small applied bias and secondly the adhesion of contaminating particles should be minimised. Additionally, the surface must have had enough reactive surface sites for highly uniform growth to occur, which is likely due to the stable adsorption of atomic oxygen onto sub-nanoscale rough Au surfaces during the oxygen plasma exposure[201].

It was shown that a high consistency in the electronic characteristics of ultra-thin MIM devices can be achieved over a large area using ALD to control precisely the thickness and uniformity of the film. A highly reliable uniform thickness of  $2 \pm 0.1$  nm was achieved for  $\text{Al}_2\text{O}_3$  using plasma-enhanced ALD, and the tunneling current density as a function of applied electric field was found to be independent of the surface area of the MIM devices. We found that the tunnel current was strongly asymmetric, beyond what would be expected as a result of the difference in work function between the two metal electrodes. Although the dielectric thin film is highly uniform, we have shown using TEM that sharp features on the bottom metal surface lead to a net field enhancement and hence an asymmetric distortion in the potential across the film. We have developed a model which takes into account such distortions and have demonstrated that field enhancement can accurately account for the observed asymmetry in the I–V characteristics of the MIM devices. These results show that deliberate and controlled introduction of sharp features in the metal film may be used to engineer an asymmetry and non-linearity into the I–V characteristics of MIM structures, which are of critical importance to a range of applications, including IR rectenna diodes, storage capacitors for DRAM, and non-volatile memory applications.

In the case of the type-B, 2 nm devices, the same was observed where the forward characteristic is more conductive than the reverse. For these devices, the current cannot be directly fitted to Equation 6.8 without including the band structure of the surface states, which has been performed by Davis *et al.*[259], as discussed in Chapter 5. The shape of the I–V curves in field enhanced regions, shown in the inset Figure 6.8, corresponds closely with the shapes observed in STM measurement on the Au surface[257–262], where field enhancement near the tunneling tip causes a much higher current when electrons are emitted from the tip, contrary to the devices presented in Chapter 5, where the field enhancement occurs more on the same surface that contains the surface states. Using the barrier parameters estimated in the 2 nm type-A devices, replacing the top electrode with Au decreases the current by a factor just under 10, and therefore the overall smaller conductance of the type-B devices cannot be explained only by the change in top electrode material.

The image force has the effect of decreasing the asymmetry in the current by rounding off the corners and decreasing the effective height of the potential barrier, where both these effects increase as thickness is decreased[227]. The conditions under which the image force should be considered are not yet clear as it is intrinsically related to the behaviour of the elastic tunneling electron inside the potential barrier[108, 288, 289] which to our knowledge remains an unsolved problem. Thus under the observation that fine distortions in the barrier are required to produce low-voltage asymmetry, it seems reasonable to expect the image force effect is small enough such that the distortion in the potential is still seen by tunneling electrons. If it were to be considered, the restorative force acting on departing electrons in the sharp regions would be expected to be more important due to the different spatial charge configuration. The effect can be included as a contribution to Equation 6.4.

It is worth noting that this model can be applied to electron tunneling devices besides SETs. For example, metal-insulator-metal (MIM) capacitors with ALD deposited high-k dielectrics are used in applications such as radio-frequency (RF) capacitors for integrated circuits[212, 216, 266, 274], storage capacitors for dynamic RAM (DRAM)[154, 156] and non-volatile memory (NVM) applications[102, 169]. Similarly, MIM diodes critically underpin high-speed electronics applications such as infra-red (IR) detectors[170, 290–292] used in energy harvesting applications and backplane selector diodes for LCD displays.

### 6.3 Temperature dependence of the barrier parameters

Using the model presented in the previous section, the I–V characteristics of a type-A, 3 nm representative device is analysed as a function of temperature. A series of I–V curves were recorded at each temperature set point between 4 and 290 K. In the

fitting routine,  $t_J$  is set to 3 nm, and the same values as in the previous section are used for the work functions of the metal electrodes. In order to extract the temperature dependence of the barrier parameters, the result of fitting the 4 K I–V curve provides values for  $A_d$  and  $A_i$ , which are subsequently kept constant for the remaining curves. The underlying assumption is that the geometry does not change with temperature, which is a reasonable approximation considering that thermal expansion effects are very small for isolated sharp features. Figure 6.8 shows a plot of the I–V curves taken at the lowest and highest temperatures, where the mixed barrier result is represented with the solid lines, and the uniform barrier by the thin dashed lines. The natural weighting of each point in the fitting routine results in a good fit for the current at higher voltages. As with the 2 nm devices, the asymmetry and qualitative shape of the I–V curves is better described by the mixed barrier model that includes regions where field enhancement occurs near the bottom electrode. The low temperature affinity obtained was  $\chi_I \approx 1.4$  eV, which is smaller than that obtained for the 2 nm devices, and the effective mass for electrons in the insulator conduction band obtained was  $m^* \approx 0.43m_e$ , which is also smaller than obtained for the 2 nm devices. Although these are significant differences, decreasing  $\chi_I$  decreases the current density, but decreasing  $m^*$  increases the current density, and thus the parameters obtained as a function of thickness reflect possible errors in the values of  $A_J$  and  $t_J$ , the underlying approximations made in the model and changes in the material properties with thickness. The value obtained for  $A_d$  is 18% of  $A_J$ , which again is smaller than obtained in the 2 nm devices, but does not represent a difference as large as with the electronic barrier parameters and is well within the uncertainty obtained in the 2 nm devices for this parameter. The uncertainty in the obtained parameters for the 3 nm devices have similar ratios to those shown in Table 6.1, but are generally smaller at all temperatures.

Fitting of the I–V curves was performed in two ways. In the first pass, represented in Figure 6.8, the curves were fitted leaving only  $\chi_I$  as a free parameter at all temperatures. The effective mass is kept at the value obtained from the 4 K I–V curve. Figure 6.9 (a) shows the dependence of the obtained values of  $\chi_I$  as function of temperature. As can be seen, generally the barrier height

$$\phi_B(T) = \Phi_{L,R} - \chi_I(T) \quad (6.12)$$

decreases with increasing temperature, and changes little at low temperature. In the second pass, the curves were fitted using both  $\chi_I$  and  $m^*$  as free parameters at all temperatures, which resulted in an identical dependence of  $\chi_I$  against temperature, and a slight increase of 10% in the effective mass was obtained over the whole temperature range. The  $R^2$  values obtained were almost identical in both cases, which suggests there is little statistical significance to the derived variation of the effective mass. Figure 6.9

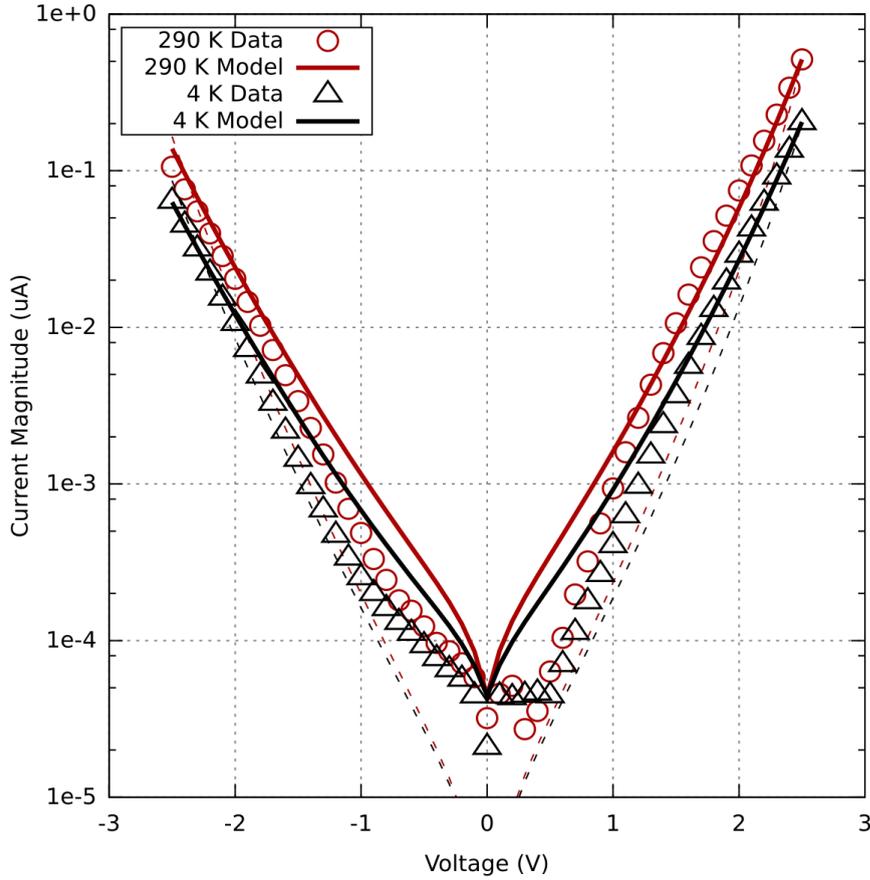


FIGURE 6.8: Fitting results of the mixed barrier model described by Equation 6.8 (solid lines) and the uniform barrier model (dashed lines) where  $A_d = 0$  in Equation 6.8, to the experimental data for the 3 nm thick device at 4 and 290 K.

(b) shows the goodness of fit obtained for the first pass, which shows that the fit is better at lower temperatures, which is expected since the I–V curves change little and the values of  $A_d$  and  $A_i$  were kept constant. The trend in  $\chi_I$  is qualitatively identical in both cases, although the change in barrier height appears greater when the effective mass is allowed to change. Physically, the change in the barrier height reflects an underlying change in the bandstructure of the  $\text{Al}_2\text{O}_3$  layer, which implies both the band gap  $E_G$  and  $\chi_I$  can be temperature dependent.

The temperature dependence of the band gap has a physical explanation and has previously been analysed in a 2 nm  $\text{Al}_2\text{O}_3$  MIM device, where the  $\text{Al}_2\text{O}_3$  layer was formed by oxidation in atmosphere for 24 hours[256]. They present qualitatively identical results to those presented in Figure 6.8 (a), and interpret the temperature dependence of the current as due entirely to the change in band gap with temperature. It has been shown that in many semiconductors, the band gap decreases with increasing temperature due to the electron-phonon interaction[293]. Phonons are quasi-particles that represent collective vibrations of the atomic lattice and generally have energies corresponding to the far-infra red region of the electromagnetic spectrum. The phonon population

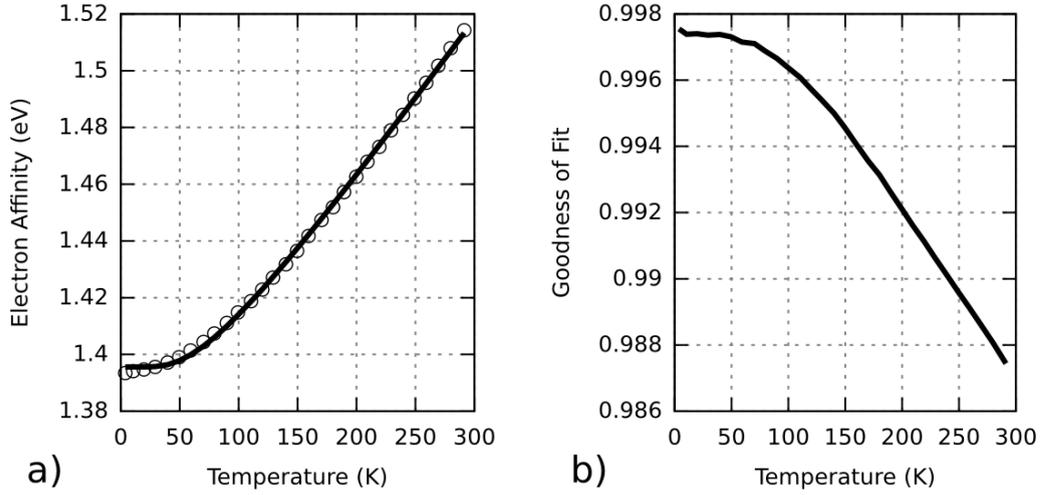


FIGURE 6.9: (a) Temperature dependence of the electron affinity, represented by the open circles, which reflects the change in the  $\text{Al}_2\text{O}_3$  barrier height through Equation 6.12. The solid line represents the fitting result to Equation 6.13. (b) Goodness of fit ( $R^2$ ) of the mixed barrier model as a function of temperature.

is described by Bose-Einstein statistics, and due to their small energies which reach maximum values in the meV range, their population is highly sensitive to temperature. At low temperatures, lattice vibrations are strongly suppressed and thus the effect of the electron-phonon interaction is reduced. In a wide band gap insulator such as  $\text{Al}_2\text{O}_3$ , phonons can interact with valence electrons and in turn modulate the band gap. The modulation of the band gap occurs due to the deformation of the atomic potential under lattice vibrations[294]. As the phonon spectrum of a material is highly complex, efforts have been made to describe the modulation of the band gap as due to an average phonon mode  $\langle \hbar\omega \rangle$  representative of the distribution of phonon frequencies that depends on the material[295]. A simple model that expresses the dependence of the band gap on the temperature was developed based on the temperature dependence of the phonon population by O'Donnell *et al.*, which reads[293]

$$E_G(T) = E_G(0) - S_{pe} \langle \hbar\omega \rangle \left[ \coth \left( \frac{\langle \hbar\omega \rangle}{2k_B T} \right) - 1 \right] \quad (6.13)$$

where  $S_{pe}$  is interpreted as a coupling constant representative of the strength of the electron-phonon interactions.

In the present work the value of  $E_G$  was not measured, although it is estimated to be greater than 6 eV according to reports on ALD deposited  $\text{Al}_2\text{O}_3$ [239, 240]. Following the procedure used by Patiño *et al.*[256],  $\chi_I$  is assumed to be constant against temperature, such that band gap can be written as

$$E_G(T) = \gamma_a \phi_B(T) \quad (6.14)$$

using Equation 6.12 for the barrier height  $\phi_B$  and where  $\gamma_a$  depends on the band alignment at the  $\text{Al}_2\text{O}_3$  and metal layer interfaces, which should take a value near 2, according to the barrier heights on the order of 3 eV obtained in the previous and present section. Equation 6.13 shows that the zero temperature value of the band gap  $E_G(0)$  can be arbitrarily chosen as it plays no role in determining the slope of the temperature dependent part. The value of  $\langle \hbar\omega \rangle$  controls the temperature at which  $E_G(T)$  becomes relatively temperature independent, and  $S_{pe}$  controls the rate at which the band gap decreases with temperature when the thermal energy is greater than  $\langle \hbar\omega \rangle$ . Due to the arbitrariness of the choice of  $E_G(0)$ ,  $\chi_I(T)$  can be fitted to Equation 6.13 by only inverting the sign of the temperature dependent part, and is shown in Figure 6.9 (a). The goodness of fit obtained was  $R^2 > 0.999$  and the parameters obtained were  $S_{pe} = 3.35$  and  $\langle \omega \rangle = 26$  THz. The value of  $\langle \omega \rangle$  obtained is in close agreement with Patiño *et al.*'s result of 20.3 THz, however the value of  $S_{pe}$  obtained here is twice the size of theirs, which indicates the band gap in our devices appears to be decreasing faster, within the interpretation of the model. The result is also very close to derived values of the average phonon frequency  $\langle \omega \rangle = 22.4$  THz obtained from speed of sound measurements in amorphous  $\text{Al}_2\text{O}_3$  thin films[296], which Patiño *et al.* also compared against. The discrepancy in the value of  $S_{pe}$  could be due to the different experimental conditions, where in this work thicker films and a much broader voltage range was used. There are few reports of this nature on the temperature dependence of electron tunneling in  $\text{Al}_2\text{O}_3$  MIM devices, however despite this, a reasonable agreement is obtained with these parameters. Values of  $S_{pe}$  on the order of 3 have been reported for Ga based semiconductor materials[293]. Other devices were measured but not to such a high temperature resolution, however the same trend occurs where at low temperatures less than 70 K, the thermal activation becomes very small, which yields very similar values of  $\langle \omega \rangle$ , and similarly for the slope.

## 6.4 Dielectric constant

Due to the significant challenges facing the down-scaling of CMOS devices as critical device dimensions on the molecular scale are approached[297–300], there is considerable interest in novel two dimensional materials, which due to the significant confinement in one dimension, show novel properties and present new opportunities for the mass fabrication of devices. Interest is focused on materials that can be used to provide basic circuit elements, formed of conductors, semiconductors, and insulators. For instance, graphene shows great promise for use as metallic contacts[214, 298], and possibly as a semiconductor[301–303]. Newer, more exotic monolayer materials are being investigated for use as semiconductors, such as  $\text{MoS}_2$ [304–306] and  $\text{SnS}_2$ [307]. ALD is a highly versatile and reproducible technique, that can be used to fabricate very thin films, the composition of which can be selected from a vast array of materials[98, 308]. The

sequential chemical growth allows the controlled deposition of many of these materials to sub-nm precision, over large surface areas relevant to industrial scale manufacturing. The deposition of  $\text{Al}_2\text{O}_3$  by ALD has by now been widely used in research and in industry, and continues to be used today as a highly reliable insulating material in novel applications in nanoelectronics, such as spintronics[309], magnetic tunnel junctions[310], solar cells[311], SETs[102, 103] and MIM capacitor and tunnel diode structures for RF[171, 312] and IR[290, 313] applications. Less interest has focused on the properties of ultra-thin ALD deposited  $\text{Al}_2\text{O}_3$  based MIM structures however, as more effort has focused on ALD deposited  $\text{Al}_2\text{O}_3$  employed as a gate dielectric in nanoscale FET devices, which was used in the Intel 45 nm CMOS technology node.

Size effects on the dielectric constant have already been demonstrated in experimental and theoretical reports. Groner *et al.* created  $\text{Al}_2\text{O}_3$  MIM structures formed of n-doped Si and Mo electrodes, and found a significant decrease in the dielectric constant below 60 nm  $\text{Al}_2\text{O}_3$  thickness, and the smallest devices they created featured a 3 nm insulator thickness, which gave a dielectric constant on the order of half the bulk value they obtained[199]. In their MIM structure, they used materials that readily oxidise when exposed to the oxygen based precursors used, and thus their result is well explained by the inclusion of a low permittivity oxide on either material used as the seed layer in the deposition, the thickness of which must be determined independently, as shown by others. In contrast, the permittivity in confined materials can be considerably reduced compared to their bulk equivalent. An atomistic model of an Si slab has shown that both the ionic and electronic contributions to the dielectric constant decrease with thickness below a critical thickness[314], which shows that confinement effects in semiconductors play a role not only on the electronic transport properties, but on the dynamical properties represented by the dielectric constant. They also show that at the molecular scale, fluctuations of the local permittivity near a contact with a material can drastically affect the measured dielectric constant. Similar thickness dependent properties have been observed and predicted in nanoscale materials such as  $\text{MoS}_2$ [306, 315].

The qualitative shape of the decrease in the dielectric constant as thickness decreases can be reproduced in a simple one dimensional classical macroscopic model. The decrease in the dielectric constant can be interpreted as a dielectric screening effect, where regions of small permittivity screen the regions of higher permittivity, so that the effective bulk dielectric constant is reduced. Consider the model of the permittivity in the planar MIM structure shown in Figure 6.10 (a), where two interfacial regions of thickness  $t_{i1}$  and  $t_{i2}$  are introduced, which each have a smaller permittivity  $\epsilon_{i1}$  and  $\epsilon_{i2}$  than the bulk dielectric constant  $\epsilon_b$ . The equivalent circuit of the capacitance is then given by

$$\frac{1}{C_J} = \frac{1}{C_{i1}} + \frac{1}{C_b} + \frac{1}{C_{i2}}, \quad (6.15)$$

where each capacitor has the same surface area  $A_J$ . Since there is no knowledge that can be used to predetermine the properties that define each of  $C_{i1}$  and  $C_{i2}$ , for the purposes of analysis they can be lumped together into a single interfacial capacitance  $C_I = \epsilon_0 \epsilon_I A_J / t_I$ , so that  $t_J = t_b + t_I$  and  $t_I = t_{i1} + t_{i2}$ . The measured permittivity  $\epsilon_J$  can be expressed in terms of each region as

$$\epsilon_J = t_J \left( \frac{t_J - t_I}{\epsilon_b} + \frac{t_I}{\epsilon_I} \right)^{-1}. \quad (6.16)$$

This can then be used to fit the values of  $\epsilon_J$  obtained from the capacitance as a function of thickness  $t_J$ , where it is assumed that  $t_I$ ,  $\epsilon_b$  and  $\epsilon_I$  do not depend on  $t_J$ . The result obtained for both the type-A and B devices is shown in Figure 6.10 (b), for which  $R^2 > 0.95$  was obtained in both cases. The solved parameters are not unique, and a range of solutions exist for both  $t_I$  and  $\epsilon_I$ , constrained by the values of  $t_J$  and  $\epsilon_J$ . The most immediately relevant result is to ascribe the interface capacitance to a vacuum layer, of permittivity  $\epsilon_I$ , so that the smallest possible value of  $t_I$  can be determined. Using this method and  $\epsilon_b = 8.8$ ,  $t_I = 2 \text{ \AA}$  was obtained for the type-B devices, and  $t_I = 1.935 \text{ \AA}$  was obtained for the type-A devices, which represents a distance comparable to the atomic spacing in solid materials. Identical solutions are obtained when the permittivity of the interfacial layer is increased, and the corresponding thickness can be determined as

$$t_I = g \left( \frac{1}{\epsilon_b} - \frac{1}{\epsilon_I} \right)^{-1}, \quad (6.17)$$

where  $g$  is a negative variable that depends on  $\epsilon_J$ ,  $\epsilon_b$ ,  $t_J$  and can be determined using Equation 6.16.

Such a model of the interfacial capacitance has commonly been used to describe MIM devices with ferroelectric dielectric materials[317–321], and has been called the dead layer effect, which in such devices occurs at much greater thicknesses on the order of  $1 \mu\text{m}$ . Although the quantitative and qualitative results of this model are very acceptable, the underlying classical description and interpretation of interfacial capacitance presents a number of issues. In this model, the dielectric layers are treated as being in a perfect chemical contact, so that there are no metallic electrodes between the layers. This then requires the charge to be conserved across all the dielectric layers, rather than in each capacitor separately as for the multi-island metallic SET systems. In the planar structure, the flux density is constant across all the layers, and thus the charge on each capacitor is the same, i.e.  $Q_J = Q_{i1} = Q_b = Q_{i2}$ . From this and Equation 6.15 it follows that the voltage across each dielectric layer is related to the potential difference between the electrodes

$$V_k = \Delta\varphi \frac{C_J}{C_k}, \quad (6.18)$$

where  $k$  can be each of  $i1$ , and  $i2$  and  $b$ . This shows that depending on the choice

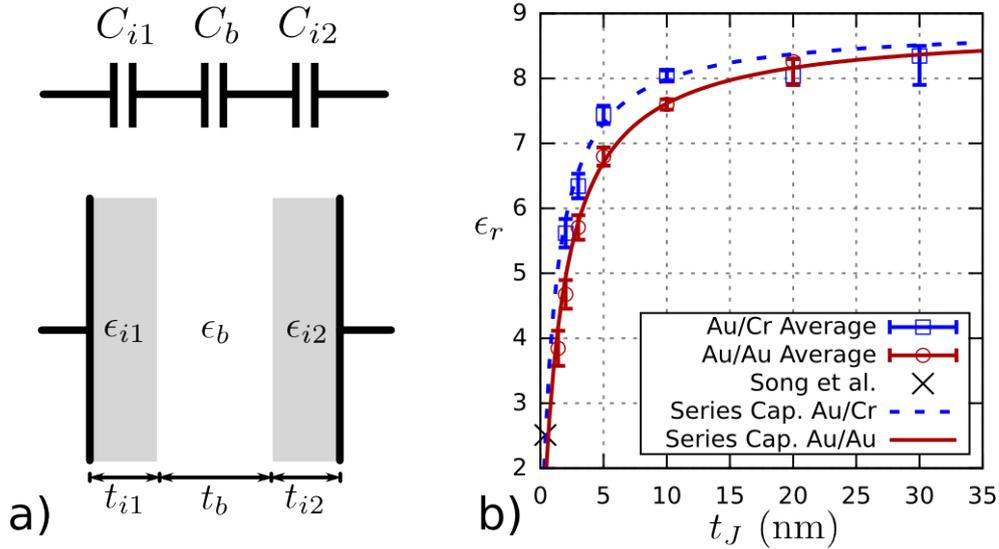


FIGURE 6.10: (a) Equivalent circuit and electrostatic model of the classical dielectric screening effect, represented by the interfacial capacitances  $C_{i1}$  and  $C_{i2}$ . (b) Result of using Equation 6.16 to fit the thickness dependence of the permittivity. Song *et al.*'s theoretical result for an  $\text{Al}_2\text{O}_3$  monolayer is included[316].

of parameters, the shape of potential barrier is affected, which implies the tunneling characteristics should be affected. As discussed in Section 6.2, this can be effectively accounted for in the tunneling model. This analysis is usually performed for multi-layer dielectrics in MIM devices, which are used to engineer the Fowler-Nordheim characteristics[274]. If the layer is interpreted as a vacuum layer, then the tunneling electrons see a very thin but tall potential barrier before the  $\text{Al}_2\text{O}_3$  regions, the dimensions of which significantly affect the tunnel current and asymmetry, particularly for large values of  $t_I$ . For the value of  $t_I$  obtained here, the vacuum region is virtually transparent to electrons and mainly has the effect of decreasing the amplitude of the current rather than affecting its behaviour under bias. As can be seen from Equation 6.18, as  $t_I$  is made very small,  $C_I$  becomes very large compared to  $C_J$  and thus most of the voltage is across the bulk layer rather than the interfacial layers.

At the length scale of the values of  $t_I$  obtained, the interpretation of a vacuum layer as presented is questionable first because it is atomic in scale, but also because the macroscopic concept of the permittivity as used so far is not applicable at this scale. Clearly a microscopic model of the permittivity as a function of space is required to describe the observed trend, that provides a continuous, but position and thickness dependent permittivity. A classical model that considers the microscopic dielectric constant as a function of position and thickness has shown that a reduction of the intrinsic local permittivity occurs at the interfaces of a thin dielectric slab[322]. In this method, a lattice of electric dipoles is defined, and the local electric field due to each dipole is

then due to the superposition of all the dipole electric fields. The atomic polarisability  $\alpha_a$  of each dipole is empirically determined from the bulk dielectric constant using the Clausius-Mossotti relation

$$\frac{\epsilon_J - 1}{\epsilon_J + 2} = \frac{N\alpha_a}{3\epsilon_0}, \quad (6.19)$$

where  $N$  is the density of dipoles in space. Using a cubic lattice, it was found that the local permittivity is smaller in only two layers of dipoles from each interface, and the outer-most layers show a reduction of about 20% from the bulk value. These generalised dipoles cannot easily be related to spatial dimensions in  $\text{Al}_2\text{O}_3$ , due to its amorphous structure, and the ionic nature of the Al–O bonds. Starting from a phenomenological description in one dimension based on Maxwell's equations, the displacement current density can be written as

$$J(\omega) = \frac{\partial D}{\partial t} \quad (6.20)$$

where  $D$  is the frequency dependent electric flux density in one dimension, which is expressed as

$$D(\omega) = \epsilon_0\xi(\omega) + P(\omega), \quad (6.21)$$

where  $P$  is the polarisation density in one dimension, which represents the response from the dielectric material, and therefore contains all the relevant dynamical electronic properties of the material. In general,  $P$  is due to all dipoles that align with the direction of the applied alternating electric field  $\xi(\omega)$ , whether intrinsic or not to the dielectric material, where each contribution can be expressed as a susceptibility  $\chi_i(\omega)$  in the frequency domain as

$$P(\omega) = \epsilon_0\xi(\omega) \sum_i^N \chi_i(\omega), \quad (6.22)$$

where it is assumed the polarisation is a linear function of the electric field, and that there is no static polarisation.

In diatomic, ionic materials such as  $\text{Al}_2\text{O}_3$ , two intrinsic contributions arise, first the polarisation of the electronic clouds responsible for the optical dielectric properties can be expressed as  $\chi_{\text{ele}}$ . The second contribution  $\chi_{\text{ion}}$  is due to the induced dipoles formed by relative displacements of the oppositely charged ionic cores, which for  $\text{Al}_2\text{O}_3$  are  $\text{Al}^{3+}$  and  $\text{O}^{2-}$ . Extrinsic contributions to  $P$ , such as the dipoles created by hopping of trapped electrons, tend to respond very slowly to the time dependent electric field in the range of frequencies used in this work (1 Hz - 200 kHz)[207, 323, 324], however no significant loss tangent was measured in the whole frequency range as shown in Section 5.2. Therefore the flux density can be expressed solely in terms of the intrinsic contributions

$$D(\omega) = \epsilon_0\xi(\omega) [1 + \chi_{\text{ele}}(\omega) + \chi_{\text{ion}}(\omega)] = \epsilon_0\epsilon_J(\omega)\xi(\omega), \quad (6.23)$$

which shows the phenomenological description of the measured relative permittivity  $\epsilon_r(\omega)$ . The response of each susceptibility component to a change in the electric field against time can be approximated as a first order Debye process for the purpose of this discussion. Under a constant electric field, an induced ionic dipole is in thermal equilibrium. If at  $\tau = 0$  the field is set to zero instantaneously, there is a delay for the induced dipole to reach equilibrium, and the polarisation can be written as  $P_{\text{ion}}(\tau) = \epsilon_0 \chi_{\text{ion}}(0) \exp(-\tau/\tau_{\text{ion}})$ . Taking the (positive) Fourier transform of this expression yields the frequency dependent susceptibility

$$\chi_{\text{ion}}(\omega) = \chi_{\text{ion}}(0) \frac{\omega_{\text{ion}}}{\omega_{\text{ion}} + j\omega}, \quad (6.24)$$

where  $\omega_{\text{ion}}$  is the frequency at which the amplitude of the contribution is decreased by half. A similar expression can be assumed for the electronic contribution, which is characterised by a resonant frequency  $\omega_{\text{ele}}$  of much greater amplitude than  $\omega_{\text{ion}}$ . The real part of the susceptibility then contributes to the (real) capacitance, and the imaginary part presents a frequency dependent real impedance known as the dielectric loss, which consequently affects the phase of the measured impedance. It was established experimentally in Section 5.2 that there is no discernable dielectric loss in the range of electric field frequencies used and at all thicknesses, which shows that  $\omega_{\text{ele}} \gg \omega_{\text{ion}} \gg \omega$ , as expected, since  $\omega_{\text{ion}}$  is expected to be found in the THz range[325]. Therefore the measured values of  $\epsilon_J$  as a function of thickness contains both  $\chi_{\text{ion}}(0)$  and  $\chi_{\text{ele}}(0)$ .  $\chi_{\text{ele}}(0)$  can then be obtained through knowledge of the refractive index  $n_i$ , which gives  $\chi_{\text{ele}}(0) = n_i^2 - 1$ . Values of  $n_i$  have been reported to vary between 1.5 and 1.8, and thus  $\chi_{\text{ele}}(0)$  is between 2.25 and 3.25. This contribution is usually called the optical dielectric constant, and is written as

$$\epsilon_{\infty} = 1 + \chi_{\text{ele}}(0) = n_i^2, \quad (6.25)$$

so that the static dielectric constant obtained from the impedance spectroscopy measurements can be written as

$$\epsilon_J = \epsilon_{\infty} + \chi_{\text{ion}}(0). \quad (6.26)$$

In high-k materials,  $\epsilon_{\infty}$  is usually much smaller than  $\chi_{\text{ion}}(0)$ , which shows that the high permittivities in such materials are linked to the motion of ions in the dielectric material[209].

First principles calculations of a structurally stable  $\text{Al}_2\text{O}_3$  monolayer have shown that the band-gap and the static permittivity both decrease by a significant fraction[316]. It was found that both  $\epsilon_{\infty}$  (1.27) and the ionic part of the dielectric constant (1.24) decrease perpendicular to the plane of the monolayer, giving a value of approximately 2.5 for the total dielectric constant, which is much smaller than in the bulk value. Their result is shown in Figure 6.10 (b), where a thickness of 4 Å was used, which is no longer a

well defined quantity, and could be larger or smaller when in contact with a material by up to a few Å, depending on the interactions of the layer with the top and bottom electrodes. Even within these uncertainties, their result aligns quite well with the result from the dead layer capacitance model. The results found here suggest that the dielectric constant of Al<sub>2</sub>O<sub>3</sub> in MIM devices transitions towards this monolayer limit, starting at a critical thickness, below which some progressive and fundamental change occurs in the dynamical material properties. The decrease in the contributions to the dielectric constant are caused by a change in the bonding structure, where Al–O bonds become smaller and more covalent, such that electrons are shared more between ions. This effect is attributed to a reduction in the coordination environment, i.e. the number of bonds that each Al and O atom. As discussed previously, it has been extensively shown that ALD deposited Al<sub>2</sub>O<sub>3</sub> has an amorphous structure. Two dimensional nuclear-magnetic-resonance (NMR) studies on ALD deposited Al<sub>2</sub>O<sub>3</sub> have shown that the coordination environment of each atom is already reduced due to the disorder, so that Al atoms with 4 bound O atoms are much more common[241], rather than Al atoms with 6 bound O atoms, as expected in crystalline Al<sub>2</sub>O<sub>3</sub>. At the interfaces, there is inevitably a change in coordination environment compared to the bulk that likely affects the local dielectric constant, depending on the interface material, and in a bulk system would not be apparent in capacitance measurements. However, as the amorphous material is made thinner approaching atomic dimensions, the effect of the interfaces on the remaining bulk of the material must become more important, due to the already reduced coordination environment. Such subtle changes in the bonding properties could be responsible for both the observed trends in the dielectric constant and a change in the tunnel barrier parameters with thickness.

The behaviour of the permittivity under an applied electric field and as a function of temperature has been experimentally and theoretically investigated by Bécu *et al.*[225], who's C–V results of Al<sub>2</sub>O<sub>3</sub> MIM structures were found to be consistent with those presented here. Due to the wide band gap of Al<sub>2</sub>O<sub>3</sub> and the apparent mid-gap alignment of the Fermi level (according to the barrier height results obtained from the tunneling model), valence electrons are tightly bound to the atomic cores, and thus they are displaced little by the local electric field and thermal fluctuations. In contrast, the motion of the ionic cores is much more flexible, since they are trapped in a comparatively weaker (inter-atomic) potential. They consider an oxygen anion in a tetrahedral cell formed of Al cations, which corresponds to the crystalline case of Al<sub>2</sub>O<sub>3</sub>, and compute the interaction potential of the ions in the cell using the Mie-Grüneisen (MG) potential with a bond energy  $E_{MG} = 1.5$  eV estimated from the elastic modulus of Al<sub>2</sub>O<sub>3</sub>

$$\phi_{MG}(r) = E_{MG} \frac{n_r n_a}{n_r - n_a} \left[ \frac{1}{n_r} \left( \frac{r_0}{r} \right)^{n_r} - \frac{1}{n_a} \left( \frac{r_0}{r} \right)^{n_a} \right] \quad (6.27)$$

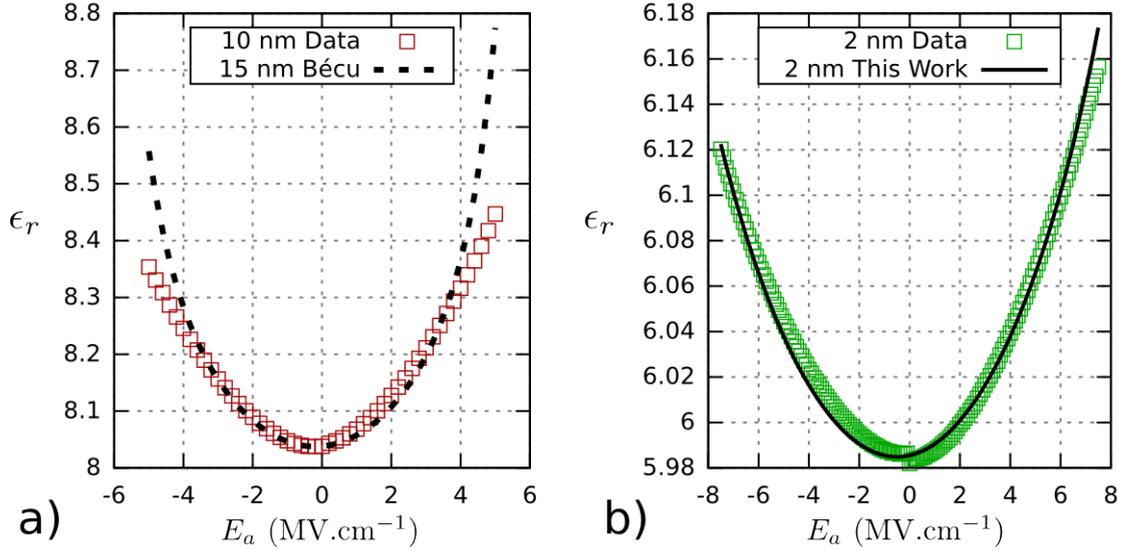


FIGURE 6.11: (a) Plot of the permittivity against applied electric field data obtained for a 10 nm type-A device. The green dashed line shows Bécú *et al.*'s result for a 15 nm  $\text{Al}_2\text{O}_3$  MIM device. (b) Fitting of the permittivity against applied electric field data obtained for a 2 nm device using the same model.

where  $n_a = 1$  controls the attractive part of the potential,  $n_r = 9$  controls the repulsive part of the potential and  $r_0 = 1.9 \text{ \AA}$  is the Al–O bond length. The MG potential then allows the free energy of an ensemble of cells to be calculated from thermodynamic principles. They then express the ionic susceptibility as

$$\chi_{\text{ion}}(V_a) = n_M k_B T \frac{1}{2} \frac{\partial^2 U}{\partial \xi_{\text{loc}}^2}, \quad (6.28)$$

where  $U$  is the free energy of the system (the authors use the notation  $F$  in their paper),  $n_M = 3.2 \times 10^{22} \text{ cm}^{-3}$  is the number density of Al cations per unit volume and  $\xi_{\text{loc}}$  is the local electric field, which can be defined as

$$\xi_{\text{loc}} = \xi \frac{3\epsilon_J}{2\epsilon_J + 1} = \xi \lambda_{\text{loc}} \quad (6.29)$$

when treating the dipoles in the classical Onsager model. This model was implemented as prescribed, and fitting of the data was performed by evaluating Equation 6.28 numerically and using the definition of  $U$ , the interaction potential, and the parameter values reported in Ref. [225], then  $\lambda_{\text{loc}}$  and  $\epsilon_\infty$  are used as free parameters. The  $\epsilon_J$  data is then minimised using Equation 6.26. Using the bulk value of  $\epsilon_J = 8.2$ , a value of  $\lambda_{\text{loc}} \approx 1.414$  for the local field factor is expected. In their experiments, performed at 100 kHz, Bécú *et al.* found a value of 1.71 for  $\lambda_{\text{loc}}$ , and they used  $\epsilon_\infty = 1.77^2$ . Figure 6.11 (a) shows their result of using this model, superimposed on our measurements of a 10 nm type-A device. Due to the greater range of electric fields applied in our devices,

a greater deviation of the model at high electric fields can be seen, where the ionic susceptibility increases faster than a second order polynomial equation in the high electric field regions according to this model. A value of  $\epsilon_\infty \approx 2.7$  had to be used to adjust Bécu *et al.*'s result to ours, indicating they obtained a dielectric constant of about 8.6 in their device, which was only expressed in terms of the capacitance density in their paper. Figure 6.11 (b) shows the result of applying this model to the measurements obtained from a 2 nm type-B device. This result was obtained using  $\lambda_{\text{loc}}$  and  $\epsilon_\infty$  as free parameters, where it is assumed the interaction potential is identical to the bulk case. Since  $\lambda_{\text{loc}}$  controls the amplitude of the cell distortion, which in turn affects the rate of increase of the ionic susceptibility as a function of applied electric field, we found a considerably smaller value of  $\lambda_{\text{loc}} = 1.1$ . A value of  $\epsilon_\infty = 0.666$  was found which is unrealistically small, since the minimum value should be 1 in the case of a vacuum. This indicates that the amplitude of the  $\chi_{\text{ion}}$  produced by the model is too large. The amplitude is controlled by the bond energy, where a larger bond energy decreases both the cell distortion amplitude and the resulting susceptibility. Under the interpretation of this model, this suggests the decrease in the permittivity with thickness could be explained by a change in the interaction potential and the internal electric field, due to chemical effects occurring in the  $\text{Al}_2\text{O}_3$  layer as the monolayer limit is approached.

The qualitative features of the C–V curves produced and much of the underlying physics in this microscopic model corresponds closely to the real system, however the application of this model to the thin film devices presents a number of issues in terms of the interpretation and the quantities produced, although some of them can be addressed. The structure of the  $\text{Al}_2\text{O}_3$  layer is amorphous and as discussed O and Al atoms have a lowered coordination environment as a result. Therefore no particular cell can be defined to a sufficiently small scale to capture the disorder of the structure, and large cells make the definition of the MG potential and the ensuing calculations much more complicated. A fundamental assumption made by Bécu *et al.* to simplify the calculations is that the cells do not interact with each other, which in a bulk system is not necessarily unreasonable since the desired properties are reproduced in those cases, notably independence of the result on the size of the  $\text{Al}_2\text{O}_3$  layer. Furthermore, as Equation 6.28 is determined from thermodynamic principles, it is assumed the ensemble of cells is very large, which appears to be applicable in  $\text{Al}_2\text{O}_3$  MIM devices down to about 5 nm thickness, according to the capacitance measurements shown here and in Chapter 5. These assumptions certainly break down in the ultra-thin film limit, since in the direction perpendicular to the interfaces there are a countable number of cells, for which the interactions with the interfaces and each other begin to influence the electronic properties of the layer.

Some insight into this apparent size effect of the dielectric constant can be gained from first principles studies using density-functional-theory (DFT). DFT is a very effective numerical method for solving the many body Hamiltonian of a crystalline system of

atoms, whereby the symmetry of such systems is exploited to considerably simplify the mathematical problem, and there exist numerous approaches to different physical problems[326]. Momida *et al.*[200] have performed a DFT study of the bulk dielectric properties in amorphous  $\text{Al}_2\text{O}_3$ , using a large cell composed of 120 atoms to capture the random character of the structure. They found a close match to the experimental reported values of both  $\epsilon_\infty \approx 2.7$  and  $\chi_{\text{ion}}$  between 5 and 8 depending on the cell, and concluded that the obtained result is very sensitive to the particular atomic arrangements in each cell. With regards to the ionic susceptibility, they use the Berry phase description of electric polarisation[327–329], which is explicitly only defined for infinite systems, and thus the theory does not express long range electric field effects in a completely accessible way. They show using this method that the Born effective charges[329, 330] of the Al and O sites in the bulk system are close to ionic values of  $Z_{\text{Al}}^* = +3e$  and  $Z_{\text{O}}^* = -2e$  respectively, which shows that the bonds are highly ionic in nature in this model, i.e. the charges are highly localised to each respective ion. This is in contrast to the monolayer system investigated by Song *et al.*[316], who found the electrons are delocalised from the ionic cores, indicating the Born effective charges of the Al and O sites should be smaller in this case, which in turn decreases the ionic contribution to the dielectric constant, although it is still larger than the electronic contribution in their work. Importantly, Momida *et al.* find that the lattice vibrational properties of the bulk  $\text{Al}_2\text{O}_3$  are significantly affected by subtle variations in the atomic configuration, which in turn also influence the ionic contribution.

Long wavelength phonons in polar materials give rise to electric fields that extend over many unit cells typically used in perturbative DFT (DFPT), and thus they are treated separately as the resulting atomic potential modulated by this field is no longer periodic with the lattice[326]. In an ionic diatomic material, two types of lattice vibration component exist due to the ionic interactions, which are called the transverse and longitudinal modes. These modes exist in all materials, but are prominent in ionic materials due to the significant Born effective charge. If the three dimensional wave vector of the phonon is  $\mathbf{q}$ , then transverse modes represent ionic motion perpendicular to  $\mathbf{q}$ , and longitudinal modes represent ionic motion parallel to  $\mathbf{q}$ . There exist two general types of phonons, the acoustic type where ions move in phase with the wave, which must be of greater wavelength than a number of atomic periods, and the optical type where ions move out of phase with each other. Longitudinal optical (LO) phonons contribute to the dielectric constant as they describe the response of a diatomic lattice to an external oscillating electric field[330]. In the case of cubic crystal, the longitudinal optical mode  $\omega_{\text{LO}}$  is simply expressed in terms of the transverse mode  $\omega_{\text{TO}}$

$$\omega_{\text{LO}} = \sqrt{\omega_{\text{TO}}^2 + \frac{4\pi Z^*}{\Omega\epsilon_\infty}}, \quad (6.30)$$

where  $Z^*$  is the Born effective charge of ions in the cell and  $\Omega$  is the cell volume. In a completely non-ionic material, the longitudinal and transverse modes become degenerate as the effect of the lattice displacement along any high symmetry axis for any  $\mathbf{q}$  looks the same. It was shown by Lyddane, Sachs and Teller (LST) that the static dielectric constant due to the LO phonons in this cubic system can be expressed in terms of the phonon modes in a remarkably simple equation[330]

$$\frac{\epsilon_J}{\epsilon_\infty} = \frac{\omega_{\text{LO}}^2}{\omega_{\text{TO}}^2}, \quad (6.31)$$

which has since been generalised to three dimensions and used in many systems, including in studies on  $\text{Al}_2\text{O}_3$  phonon properties[331, 332], where it can be used to extract the permittivity. This equation shows that the LO/TO splitting due to the effective charge is responsible for the ionic contribution, through the product of the ratio of each set of modes squared. Therefore, high energy modes tend to contribute less as the ratio between LO and TO modes is smaller. The three dimensional form of the LST relation is expressed as[333]

$$\frac{\epsilon_J}{\epsilon_\infty} = \prod_{j=1}^{N_c} \frac{\omega_{\text{LO}j}^2}{\omega_{\text{TO}j}^2} \quad (6.32)$$

where the number of LO and TO mode pairs  $N_c$  is determined by the symmetry of the crystal and the composition of its unit cell. Table 6.3 summarises the optical modes identified in various phases of  $\text{Al}_2\text{O}_3$ , collected from various authors experimental work using IR ellipsometry techniques, and the resulting LST ratios  $\epsilon_J/\epsilon_\infty$  calculated from their data. In the crystalline phases of  $\text{Al}_2\text{O}_3$ , there are at least six pairs of LO/TO modes due to the structure of the unit cell[252, 334, 335], although not all of them appear in experimental work as their intensity often cannot be detected. The static permittivity values obtained using Equation 6.32 with  $\epsilon_\infty$  on the order of 3 yields values very close to the experimentally reported values for each phase, i.e. between 8 and 10, where the higher values correspond to the crystalline phases, which have a slightly higher density. In general one or two sets of modes (usually the higher energy ones) are found in amorphous  $\text{Al}_2\text{O}_3$  thin films by IR spectroscopy (see [331] and the references therein). The LST relation has also been generalised to amorphous systems, where the ratio can be found from the average TO and LO phonon modes[336, 337].

From Equation 6.30 and Equation 6.31, the probable causes for the permittivity decrease with thickness are a change in the Born effective charge, and a related change in the LO/TO splitting, as well as a possible change in vibrational frequency or the introduction of new modes due to the coupling of the materials at the interfaces of the MIM structure[332]. Subtle changes in the structure both intrinsic to a confined insulator and due to the influence of the interfaces can therefore cause a considerable change at a critical slab thickness. The results of Song *et al.* with the  $\text{Al}_2\text{O}_3$  monolayer[316]

Al <sub>2</sub> O <sub>3</sub> Phase	$\alpha$ [332, 338]		$\gamma$ [331]		Amorphous [331]	
Optical Mode (cm <sup>-1</sup> )	TO	LO	TO	LO	TO	LO
1	385	387.6	357	403	422	537
2	439.1	481.7	536	669	721	959
3	569	629.5	744	783	-	-
4	633.6	906.6	807	917	-	-
<b>LST ratio</b> ( $\perp$ -modes)	3.057		2.839		2.865	

TABLE 6.3: Results of the phonon spectrum measurements of different phases of bulk Al<sub>2</sub>O<sub>3</sub> summarised from various sources.

and those of others[304, 307, 314] in different systems show that both the ionic and electronic susceptibilities decrease with thickness due to changes in the electronic and atomic structure. It has been shown that in ferroelectric SrTiO<sub>2</sub>, the very large LO/TO splitting of a special soft mode, so called due to its near zero value in the ferroelectric state, increases with decreasing thickness due to subtle structural changes that occur over a long range[318], which explains the considerable dielectric screening observed in those systems. In their study the shift in phonon frequencies was observed as a function of thickness of the material using high resolution IR spectral ellipsometry. The results on the MIM structures described in this work suggest that an effect of this kind occurs, such that the measured out-of-plane static dielectric constant can be written as

$$\epsilon_J = 1 + \chi_{\text{ele}}(t_J) + \chi_{\text{ion}}(t_J), \quad (6.33)$$

where each susceptibility component decreases with  $t_J$  below a critical thickness, that may be different for each contribution. The qualitative features of the permittivity curves as a function of thickness are almost identical to those produced by the dead layer capacitance model[314], therefore the parameters obtained from this model could easily be related to a suitable theoretical result.

## 6.5 Conclusion

In this chapter the electronic measurements were analysed in detail to shed light on the origin of the observed trends as a function of the geometry. For the tunneling measurements, a model was developed that accounts for field enhancement effects near the bottom electrode. Using this model, the Al<sub>2</sub>O<sub>3</sub> potential barrier parameters  $\chi_I$  and  $m^*$  were extracted, and were found to vary with thickness, which is possibly due to the incompleteness of the model, the high correlation between  $\chi_I$  and  $m^*$  in the model, a real change in these material parameters or a combination of these factors. The asymmetry in the I–V curves as a function of voltage polarity is well explained by field enhancement effects. This model shows that such field enhancement effects are an important consideration in the design of SET tunnel junction of very small surface area, where a

device at 10 nm length scales containing a single sharp peak could be adversely affected, particularly when the thickness of the film near the peak is smallest. As long as the thickness fluctuations are small, and the geometry sufficiently planar, the asymmetry should be similar in very small devices, which in this case is negligible. However these subtle surface features will inevitably cause subtle variations in the single electron tunneling rates between islands in an SET system. The clear scaling and minimal spread of the resistance in the ultra-thin film devices discussed in Chapter 5 shows that ALD can be relied on to produce highly uniform thickness thin films. The thickness dependence of the permittivity was expressed in terms of a simple phenomenological model based on the dielectric screening effect. It was found that this model produces excellent qualitative results, however the interpretation of the underlying physical mechanisms is not straight forward. The unphysical dimensions and permittivities of the interface capacitances obtained suggest the change in permittivity with thickness is likely due to an intrinsic effect. Using a model of the bias voltage dependent permittivity, it was shown that the deviation from the bulk C–V characteristic in the thin devices could be intrinsic to the dielectric, measured as a change in the local electric field within the dielectric layer. These observations are then discussed in terms of a recently developed theoretical framework of the lattice vibrational properties of  $\text{Al}_2\text{O}_3$ . The permittivity and C–V measurements in general show that PEALD deposited  $\text{Al}_2\text{O}_3$  provides a reliable dielectric constant value that scales very well with  $t_J$  and  $A_J$ . In general, it is expected that scaling down the geometrical parameters to the dimensions required for SETs would produce more repeatable values of  $C_J$  rather than  $R_J$ , as the tunnel current has been shown to be more sensitive to geometrical fluctuations.

### Conclusion

---

In Chapter 2 the orthodox theory of Coulomb blockade was used to detail the challenges associated with employing metallic, lithographically-defined SETs as highly sensitive label-free biosensors. It was shown using results from the literature that sub-electron charge resolution can be achieved in SET systems in general and could potentially provide single-molecule resolution. The challenges associated with exploiting this charge sensitivity are twofold. Firstly, the need for room temperature operable SETs is the most critical requirement that follows from the broad aim of this project, since the biosensors are targeted towards widespread commercial use, similarly to CMOS devices for modern computer hardware. Secondly, the details of the coupling between the charge from a bound molecule and the SET islands or floating gate are expected to play a significant role in the charge sensitivity. The use of linear arrays of tunnel junctions was described as one of the promising schemes that could address the first challenge, by balancing the number of junctions and the ratio of the gate capacitance to the junction capacitance  $\alpha_J$  (Equation 2.52), the charging energy of electrons near the middle of the chain can significantly exceed the thermal energy, and the size of the system as a whole can be restricted allowing for denser arrays of devices. This method can be exploited provided the properties of the junctions, particularly the electron tunneling properties, are homogeneous, which is not easily achieved using conventional SET fabrication methods. Therefore ALD was identified as a potential candidate to deposit uniform thin-films suitable for producing MIM junctions with homogeneous dielectric and electron tunneling properties. As there is little literature that treats the application of ALD deposited  $\text{Al}_2\text{O}_3$  thin-films over large areas, and initial studies that discuss SETs fabricated by ALD do not discuss the material properties of the deposited film, the properties of the  $\text{Al}_2\text{O}_3$  film needed to be determined. In the context of mass-fabrication and engineering, the uniformity of the film properties over large areas and their scaling relationship with the junction geometry are crucial properties. Therefore, large-area MIM capacitors were identified as a good method of determining the scaling behaviour of the electronic properties of the  $\text{Al}_2\text{O}_3$  film, notably the dielectric constant and the tunneling impedance, which in turn determine the charging energy and the tunneling rates between the islands.

The fabrication of the large-area MIM devices was detailed in Chapter 4, and it was shown that the deposition of  $\text{Al}_2\text{O}_3$  by ALD is compatible with photolithography-based metal lift-off processes, and a good tolerance of the geometrical parameters could be achieved. The top and bottom electrode dimensions could be formed to a high enough

resolution that the material parameters could be extracted to a good accuracy using the surface area defined by the electrode diameter. The thickness of the  $\text{Al}_2\text{O}_3$  film was determined to approximately 0.1 nm resolution using both AFM and TEM techniques, the latter of which allowed identification of the atomic structure of the materials, which revealed a highly uniform and high quality junction. It is concluded that the  $\text{Al}_2\text{O}_3$  growth is linear, since the thickness follows a linear relationship with the number of cycles down to the molecular scale, where significant non-linearities of the growth would be evident. The uniformity of the thickness of the film observed in the TEM images of the 2 nm thick film shows that growth does occur uniformly across the entire surface area within the first few ALD cycles. Therefore it is reasonable to expect that, provided the bottom electrode surface is sufficiently reactive as was found in all the fabricated devices, thinner films can be produced using the linear GPC value obtained, which was also shown to be very close to literature values for the same process reported in the literature (Section 3.2.3). The STEM and EDX imaging showed that Al and O atoms are completely contained within the boundaries of the film region, indicating the boundaries between the materials in the MIM structure are sharp and thus well defined at the molecular scale. It was shown by AFM image analysis that the thin metallic films used for the bottom electrodes have a minimal roughness, and thus the error in the real surface area compared to that determined from the electrode diameter is not excessive. Similarly, it is not unreasonable to expect the bottom and top electrodes can be made much thinner, however due to the nature of the deposition in the EBE process, a minimum thickness on the order of 5 nm is required to obtain a dense continuous film, which is sufficient for creating lateral or vertical metallic SET devices.

The electrical characterisation of a series of MIM device batches was discussed in detail in Chapter 5, where it was shown that in general there is a good consistency between the properties of all devices with similar thickness  $\text{Al}_2\text{O}_3$  films. The impedance spectroscopy measurements and determination of the capacitance of each device as a function of varying surface area and thickness revealed that the static dielectric constant of the film decreases with thickness. This indicated that interfacial effects and changes in the dielectric properties of the  $\text{Al}_2\text{O}_3$  layer likely occur as the molecular limit of the film thickness is approached, which was consistent in all devices of varying surface area, indicating the effect is not due to significant inhomogeneties in the geometry, which are shown to be minimal in the TEM images and the AFM measurements of the film thickness. Measurement of the tunneling current revealed that a good consistency is also obtained against varying surface area, and that a clear exponential dependence of the junction resistance at low bias can be observed as a function of thickness. Measurements of the tunnel current as a function of temperature showed that the dominant conduction mechanism in films less than 5 nm thick was indeed direct tunneling, evidenced by the small thermal activation over a broad temperature range. The odd-symmetry component

in the  $G$ - $V$  curves of the type-B devices was attributed to the existence of surface states on the bottom electrode Au surface. The conductance enhancement produced at low voltages could be a useful property that allows a slightly thicker film to be used for a smaller capacitance. However in the present stacked structure of the MIM devices, the effect occurs only on electron emission from the bottom electrode, and thus in an SET formed of an array of such junctions, the resulting asymmetries in the tunnel rates between islands will play an important role in determining the SET conductance. Asymmetry in the  $I$ - $V$  curves of the type-A devices could also be observed, which could not be completely explained by basic models of uniform tunnel barriers. The scaling of the tunnel resistance as a function of thickness was determined to have the slope expected of direct tunneling at low voltage within the range of thicknesses that could be measured. However, as the thickness range is narrow, and the tunnel barrier properties also appear to be changing within the range of thicknesses investigated, the obtained scaling relationship cannot reliably be used to predict the resistance of the devices at smaller thicknesses. The best indicator of the consistency of the geometry of the system over each batch is the obtained consistency in the permittivity of the films, the scaling relationship of which can reliably be used in this case.

In Chapter 6 the effect of small fluctuation in the geometry on the tunnel current was analysed using a simple model of the observed surface features. It was shown that the asymmetry in the  $I$ - $V$  curves can be attributed to the effect of electric-field enhancement occurring near the bottom electrode, which arise due to sharper features being found on the bottom electrode surface on average. Although a quantitative estimate of the distribution of the roughness could in principle be obtained from this model, the choice of the distribution is rather arbitrary, and furthermore its relation to a real surface formed of atoms is not clear. The model therefore is phenomenological in nature, and is useful to produce the necessary qualitative features expected when field-enhancement effects are present, and therefore could serve as a starting point for more sophisticated models of the tunneling process. The barrier parameters obtained for the 2 and 3 nm films using this model were found to be consistent with extensive literature treating ALD deposited  $\text{Al}_2\text{O}_3$  thin films. The barrier parameters at each thickness were not mutually consistent, as varying values of mostly the effective mass were obtained using the model presented, which suggests that additional unidentified effects should be considered to provide a physically complete description of the MIM junction. These considerations are not expected to alter the asymmetry of the curves produced by the model much, rather the amplitude of the tunnel current produced is expected to change, which would then affect the barrier parameters obtained from fitting the experimental data. The permittivity as a function of thickness was described in terms of a vacuum interface layer, which gave an excellent fit to the experimental data averaged over both types of devices produced in this work. It is expected that in an empirical interpretation of this

vacuum layer model, the permittivity of the film can be accurately predicted outside the range of thicknesses investigated. A phenomenological description of the permittivity of the film was given to elucidate the contributions to the polarisation, and the link to the VCC of the permittivity. A review of related size dependent effects of the permittivity reported experimentally and theoretically also point to the occurrence of subtle changes in the properties of the films as thickness is decreased, indicating that quantum size effects begin to play a significant role in films less than 5 nm thick. Theoretical results on the dielectric properties of an  $\text{Al}_2\text{O}_3$  monolayer in the literature were compared with our results and appear to be consistent, which suggests the dielectric properties undergo a transition towards this monolayer limit, i.e. a transition from bulk behaviour to monolayer or two-dimensional behaviour. In the present MIM devices, the thickness at which this transition begins to occur appears to be just under 10 nm, which is consistent with the properties of MIM devices investigated in the literature, where thick films are typically used.

In conclusion, the experimental and theoretical characterisation of ALD-grown MIM devices presented in this thesis shows some promising results for application in room temperature operable SETs. It was also shown that some significant challenges can be expected, particularly with regards to the homogeneity of the system. The primary advantages identified in this work are (i) the exceptional reproducibility of the deposited  $\text{Al}_2\text{O}_3$  thin films in terms of both the tunneling properties and the permittivity, (ii) the compatibility of the MIM system with state-of-the-art mass fabrication techniques, and (iii) the robust operation of the MIM devices under high electric fields and at high temperatures. The primary challenges with regards to the fabrication of the junction array are expected to be the small inhomogeneities in the electrode surfaces that will become important in devices with nanoscale electrode areas. In Chapter 6 it is shown that the electrode surface roughness can impact the tunneling properties much more than the dielectric properties, which may lead to asymmetries in the response of the SET junction arrays. Small inconsistencies in the junction capacitances cause the charging energy profile to become disordered, which can influence the operating temperature of the device. Small inconsistencies in the gate capacitances are also important to consider, particularly at low temperatures, since the charging energy levels may no longer be able to align, and may lead to devices that cannot be turned on due to the absence of available states for electron tunneling between the islands, no matter what the gate voltage is.

## 7.1 Future work

As this project is in its infancy, a great deal of work could be performed to follow up on the work presented in this thesis, therefore only direct follow-up work is discussed, and more speculative discussion is reserved for the final section in this chapter. A number

of questions follow naturally from the conclusions presented, notably, are the dielectric and tunneling properties of SET scale junctions consistent with those observed in larger area devices? In order to answer this question, a large number of single nanoscale junctions could be fabricated in a mostly identical process but scaled down using EBL methods. If the junctions are small enough for Coulomb blockade to occur at liquid helium temperatures, the dielectric and tunneling properties of the junction can be determined from the amplitude of the observed Coulomb blockade. The method of using large area MIM devices at films thinner than those used in this work presents significant difficulties, since the junctions will become significantly more conductive. In this case the dielectric properties can no longer be discerned in the same frequency range, requiring a higher frequency range, which may be inaccessible anyway due to the series resistance in the connecting leads causing significant uncertainty in such measurements. Therefore the area must be scaled down in any case for these types of measurements to be possible using standard electronic characterisation equipment as in this work. Using nanoscale junctions, the growth in the ALD  $\text{Al}_2\text{O}_3$  process could also be characterised further from a materials science point of view. Depending on the outcome of this work, the number of nanoscale junctions used could be gradually increased to test experimentally the relationship between the number of junctions and the single electron properties, notably the temperature dependence. This could be performed to unprecedented accuracy if it is found that the junction properties are sufficiently homogeneous. A great advantage of ALD is ability to tune the chemistry of the process, therefore in the event that a solution to a problem can be identified, the recipe for  $\text{Al}_2\text{O}_3$  can be altered, precursors can be replaced with different ones, and additional precursors can be used to alter the chemical composition with new elements. The use of this technique may therefore provides highly versatile solutions to issues related to the material properties of the junctions. For instance, using the readily available  $\text{HfO}_2$  precursors, an alloy can be formed with alumina, the stoichiometry of which can be finely controlled, which results in fine control of the barrier height and the dielectric properties[216].

Another pressing issue is with the origin of the permittivity, although the effect is likely beneficial to Coulomb blockade, the origin of the thickness dependence of the static permittivity in  $\text{Al}_2\text{O}_3$  is not clear. Using the existing MIM devices, the dielectric properties could be studied as a function of temperature as well as thickness, which can yield information on the energetic properties of the induced dipoles. Using a suitable model, the temperature dependent behaviour can probably be related to the lattice-vibrational properties, i.e. the phonon properties. Subsequently, this information could also serve to better understand the temperature dependence of the  $\text{Al}_2\text{O}_3$  bandgap, which is understood to be related to the phonon properties of the material.

The tunneling model used is considerably simplified with regard to the behaviour of electrons in metals. Some amount of the discrepancy observed in the scaling relationship

of the barrier parameters with thickness could be resolved by improving the tunneling model, for example by including a model of the bandstructure of the metals rather than free electrons. This could lead to a more quantitative estimation of the barrier parameters, and therefore a better understanding of conduction in a SET system composed of such junctions. Particularly with regard to the observed surface states in the type-B devices, an improved model is required to imitate the temperature dependence of the observed conductance. The observed deviation from Davis *et al.*'s model[259] at low temperatures suggests that the electric field at the interface containing the surface states might have a significant effect, since the surface band will be out of equilibrium. The existence of the surface states in only the type-B devices also presents a number of questions, however they are more of interest to physics in general than to room temperature SETs at this stage. Experiments involving the testing of different top electrode materials, and perhaps using magnetic fields could shed light on this topic.

## 7.2 Outlook

From a more general perspective, it is noted that there is little literature treating the detailed scaling behaviour of the direct tunneling properties in  $\text{Al}_2\text{O}_3$  MIM junctions with the geometry, despite the significant amount of work that has been reported on these systems. This is understandable however, since efforts have mainly focused on RF applications for such systems, where films of minimum 5 nm thickness are sufficient, and the minimum thickness is generally limited by the requirements of the VCC. Emerging applications where these properties are of interest besides SET research is in IR rectenna applications, where sub-5 nm MIM junctions could be used as rectifiers for energy harvesting and physics applications.

This thesis has focused on the application of ALD- $\text{Al}_2\text{O}_3$  films deposited on metals as the material system. This was primarily since the recipe used is widely available throughout industry and research. However, using novel ALD chemistries and surface treatments it is possible to avoid the oxidation of the underlying material, such that atomically sharp interfaces can be formed from a compositional point of view, as observed in the present MIM structures. With such capabilities, the electrode materials could be replaced with more suitable materials if required. As the orthodox theory generally holds true in well behaved SET systems, and since the extension to include confinement effects is straight forward to a first approximation, new avenues could be used for the electrode materials while retaining ALD as the technique of choice for forming the tunnel barriers. In this case confinement effects could be used to enhance the Coulomb blockade. In addition to the SET material systems discussed in Chapter 1, a significant number of emerging material systems have been reported, including using CMOS technology[339, 340]. Self-assembly[341] fabrication methods are becoming more popular, whereby a

bottom-up approach is used, using for example Au nano-particles[342, 343], InGaAs nano-particles[344], carbon nanotubes[345], graphene[346], 2D materials[347, 348], and exotic organic materials[349]. For many of these systems, it is still possible to use ALD to form tunnel-junctions, and thus a significant range of possibilities exist for implementing room temperature SETs that use an ALD-based fabrication process.

In addition to the application of new materials in SET research, a new field has emerged that could solve the issues associated with the thermal suppression of Coulomb blockade due to the broadened distribution of energetic electrons. Electron cooling[350–358] involves the use of energy filters, such as resonant tunnel barriers, to suppress the tunneling of hot electrons into the island. This technique has been used to achieve room temperature Coulomb blockade in a single CdSe-QD SET[354], using a remarkably simple system where Cr was oxidised in atmosphere, and the resulting band bending creates a quantum well with a well defined energy level, which sets up a two-step tunneling process into the QD occurring most favourably for electrons with energy aligned with the quantum well level. This cooling technique has also been demonstrated in 2DEG systems[350], which have also been used to create electrostatically formed SETs. This avenue raises interesting questions for the electrostatics and thermodynamics of the electrons and the lattice, but could potentially be used to renew interest in SET technology, for which the benefits have been long awaited[56].

From the perspective of biology, the bio-SET platform could possibly be used in research provide new insight into the charge-dynamical properties of proteins, provided single-molecule detection can be reliably achieved. SET research in the context of read-out circuits for quantum computers show that the system is well suited to the detection of charge fluctuations[359]. Thus in principle the charge fluctuations that exist in single biomolecules, depending on their environment, could be detected and spatially resolved, with knowledge of the reaction sites, using an advanced form of the proposed technology presented in this thesis.



---

## Bibliography

---

- [1] A. Jemal, F. Bray, and J. Ferlay. “Global Cancer Statistics: 2011”. In: *CA Cancer J Clin* 61 (2011), pp. 69–90.
- [2] E. K. Vadar et al. “Serum biomarker profiles as diagnostic tools in lung cancer”. In: *Cancer Biomark.* 10.1 (2015), pp. 37–54.
- [3] D. W. Yang et al. “Role of a serum-based biomarker panel in the early diagnosis of lung cancer for a cohort of high-risk patients”. In: *Cancer* 121.S17 (2015), pp. 3113–3121.
- [4] V. T. DeVita and S. A. Rosenberg. “Two Hundred Years of Cancer Research”. In: *New England Journal of Medicine* 366 (2012), pp. 2207–2214.
- [5] R. Etzioni et al. “The case for early detection.” In: *Nature reviews. Cancer* 3.4 (Apr. 2003), pp. 243–52.
- [6] M. Laudicella et al. “Cost of care for cancer patients in England: evidence from population-based patient-level data.” In: *British journal of cancer* 114.11 (2016), pp. 1286–1292.
- [7] A. N. Bhatt et al. “Cancer biomarkers - Current perspectives”. In: *Indian J Med Res* 132 (2010), pp. 129–149.
- [8] A. J. Vargas and C. C. Harris. “Biomarker development in the precision medicine era: lung cancer as a case study”. In: *Nature Reviews Cancer* 16.8 (2016), pp. 525–537.
- [9] E. F. Patz et al. “Panel of serum biomarkers for the diagnosis of lung cancer”. In: *Journal of Clinical Oncology* 25.35 (2007), pp. 5578–5583.
- [10] S. Choi et al. “Microfluidic-based biosensors toward point-of-care detection of nucleic acids and proteins”. In: *Microfluidics and Nanofluidics* 10.2 (June 2010), pp. 231–247.
- [11] S. Prakash, M. Pinti, and B. Bhushan. “Theory, fabrication and applications of microfluidic and nanofluidic biosensors.” In: *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences* 370.1967 (May 2012), pp. 2269–303.
- [12] E. Yong. “Cancer biomarkers: Written in blood.” In: *Nature* 511.7511 (2014), pp. 524–6.

- [13] G. Zheng and C. M. Lieber. “Nanowire Biosensors for Label-Free, Real-Time, Ultrasensitive Protein Detection”. In: *Methods Mol Biol.* 790 (2011), pp. 223–237.
- [14] F. Zhang. “Novel serum biomarkers for lung cancer early diagnosis and clinical outcome”. PhD thesis. The University of Texas, 2014.
- [15] G. Zheng et al. “Multiplexed electrical detection of cancer markers with nanowire sensor arrays”. In: *Nature Biotechnology* 23.10 (2005), pp. 1294–1301.
- [16] S. Ma et al. “Multiplexed Serum Biomarkers for the Detection of Lung Cancer”. In: *EBioMedicine* 11 (2016), pp. 210–218.
- [17] N. Lopez-Barbosa, J. D. Gamarra, and J. F. Osma. “The future point-of-care detection of disease and its data capture and handling”. In: *Analytical and Bioanalytical Chemistry* (2016), pp. 1–11.
- [18] J. F. Rusling et al. “Measurement of biomarker proteins for point-of-care early detection and monitoring of cancer”. In: *Analyst.* 135.10 (2010), pp. 2496–2511.
- [19] R. A. Smith et al. “Cancer Screening in the United States , 2015 : A Review of Current American Cancer Society Guidelines and Current Issues in Cancer Screening”. In: *CA: A Cancer Journal for Clinicians* 65.6 (2015), pp. 95–114.
- [20] C. Parra-Cabrera, J. Samitier, and A. Homs-Corbera. “Multiple biomarkers biosensor with just-in-time functionalization: Application to prostate cancer detection”. In: *Biosensors and Bioelectronics* 77 (2016), pp. 1192–1200.
- [21] N. Pashayan et al. “Mean sojourn time, overdiagnosis, and reduction in advanced stage prostate cancer due to screening with PSA: Implications of sojourn time on screening”. In: *British Journal of Cancer* 100.7 (2009), pp. 1198–1204.
- [22] K. Ueda et al. “Plasma low-molecular-weight proteome profiling identified neuropeptide-y as a prostate cancer biomarker polypeptide”. In: *Journal of Proteome Research* 12.10 (2013), pp. 4497–4506.
- [23] P. M. Kosaka et al. “Detection of cancer biomarkers in serum using a hybrid mechanical and optoplasmonic nanosensor.” In: *Nature nanotechnology* 9.12 (2014), pp. 1047–53.
- [24] L. A. Liotta, M. Ferrari, and E. Petricoin. “Clinical proteomics: written in blood.” In: *Nature* 425.6961 (2003), p. 905.
- [25] K. Merrell et al. “Analysis of low-abundance, low-molecular-weight serum proteins using mass spectrometry.” In: *Journal of biomolecular techniques : JBT* 15.4 (2004), pp. 238–248.

- [26] M. Muthu et al. "Tracing the voyage of SELDI-TOF MS in cancer biomarker discovery and its current depreciation trend - need for resurrection?" In: *TrAC - Trends in Analytical Chemistry* 76 (2016), pp. 95–101.
- [27] J. Yang et al. "Identification of novel low molecular weight serum peptidome biomarkers for non-small cell lung cancer (NSCLC)." In: *Journal of clinical laboratory analysis* 26.3 (2012), pp. 148–54.
- [28] S. S. Hori and S. S. Gambhir. "Mathematical Model Identifies Blood Biomarker-Based Early Cancer Detection Strategies and Limitations". In: *Sci Transl Med* 3.109 (2012).
- [29] N. Teuscher. *Ligand Binding Assays*. 2014. URL: <https://www.certara.com/2014/07/22/ligand-binding-assays/> (visited on 07/22/2014).
- [30] S. K. Vashist et al. "One-step antibody immobilization-based rapid and highly-sensitive sandwich ELISA procedure for potential in vitro diagnostics." In: *Scientific reports* 4 (2014), p. 4407.
- [31] P. Fan et al. "Enhanced sensitivity for detection of HIV-1 p24 antigen by a novel nuclease-linked fluorescence oligonucleotide assay". In: *PLoS ONE* 10.4 (2015), pp. 1–17.
- [32] B. V. Chikkaveeraiah et al. "Microfluidic electrochemical immunoarray for ultra-sensitive detection of two cancer biomarker proteins in serum". In: *Biosensors and Bioelectronics* 26.11 (2011), pp. 4477–4483.
- [33] *ELISA technical guide and protocols*. Thermo Scientific. 2010.
- [34] Y. Uludag and I. E. Tothill. "Cancer biomarker detection in serum samples using surface plasmon resonance and quartz crystal microbalance sensors with nanoparticle signal amplification". In: *Analytical Chemistry* 84.14 (2012), pp. 5898–5904.
- [35] J. D. Wulfkühle, L. a. Liotta, and E. F. Petricoin. "Proteomic applications for the early detection of cancer." In: *Nature reviews. Cancer* 3.4 (2003), pp. 267–275.
- [36] I. Lundström. "Real-time biospecific interaction analysis". In: *Biosensors and Bioelectronics* 9.9-10 (1994), pp. 725–736.
- [37] H. H. iep Nguyen et al. "Surface plasmon resonance: a versatile technique for biosensor applications". In: *Sensors (Basel, Switzerland)* 15.5 (2015), pp. 10481–10510.
- [38] Y. Liu et al. "Surface Plasmon Resonance Biosensor Based on Smart Phone Platforms". In: *Scientific Reports* 5.1 (2015), p. 12864.
- [39] E. Buitrago et al. "The top-down fabrication of a 3D-integrated, fully CMOS-compatible FET biosensor based on vertically stacked SiNWs and FinFETs". In: *Sensors and Actuators B: Chemical* 193 (Mar. 2014), pp. 400–412.

- [40] M. Curreli et al. “Real-Time , Label-Free Detection of Biological Entities Using Nanowire-Based FETs”. In: *IEEE Transactions On Nanotechnology* 7.6 (2008), pp. 651–667.
- [41] Y. Huang and A. J. Mason. “Lab-on-CMOS integration of microfluidics and electrochemical sensors.” In: *Lab on a chip* 13.19 (Oct. 2013), pp. 3929–34.
- [42] T. Kudo et al. “Fabrication of Si Nanowire Field-Effect Transistor for Highly Sensitive, Label-Free Biosensing”. In: *Japanese Journal of Applied Physics* 48.6 (June 2009), 06FJ04.
- [43] Y. Chen et al. “Silicon-based nanoelectronic field-effect pH sensor with local gate control”. In: *Applied Physics Letters* 89.22 (2006), p. 223512.
- [44] K. Taniguchi et al. “The Si-Tag for Immobilizing Proteins on a Silica Surface”. In: *Biotechnology and Bioengineering* 96.6 (2007), pp. 1023–1029.
- [45] B. He, T. J. Morrow, and C. D. Keating. “Nanowire sensors for multiplexed detection of biomolecules”. In: *Current opinion in chemical biology* 12.5 (2009), pp. 522–528.
- [46] J.-H. Ahn et al. “Double-gate nanowire field effect transistor for a biosensor”. In: *Nano letters* 10.8 (Aug. 2010), pp. 2934–8.
- [47] P. R. Nair, S. Member, and M. A. Alam. “Design Considerations of Silicon Nanowire Biosensors”. In: *IEEE Transactions on Electron Devices* 54.12 (2007), pp. 3400–3408.
- [48] G. Shalev et al. “Specific and label-free femtomolar biomarker detection with an electrostatically formed nanowire biosensor”. In: *NPG Asia Materials* 5.3 (Mar. 2013), e41.
- [49] A. Nakajima, T. Kudo, and S. Furuse. “Biomolecule detection based on Si single-electron transistors for practical use”. In: *Applied Physics Letters* 103.4 (2013), p. 043702.
- [50] A. N. Korotkov. “Intrinsic noise of the single-electron transistor”. In: *Physical Review B* 49.15 (1994), pp. 10381–10392.
- [51] C. W. J. Beenakker. “Theory of Coulomb-blockade oscillations in the conductance of a quantum dot”. In: *Physical Review B* 44.4 (1991), pp. 1646–1656.
- [52] C. A. Neugebauer and M. B. Webb. “Electrical Conduction Mechanism in Ultrathin, Evaporated Metal Films”. In: *Journal of Applied Physics* 33.1 (1962), p. 74.
- [53] R. Landauer. “Fluctuations in Bistable Tunnel Diode Circuits”. In: *Journal of Applied Physics* 33.7 (1962), pp. 2209–2216.

- [54] K. Likharev. “Correlated discrete transfer of single electrons in ultrasmall tunnel junctions”. In: *IBM J. RES. DEVELOP.* 32.1 (1988), pp. 144–158.
- [55] D. V. Averin and K. K. Likharev. “Coulomb blockade of single-electron tunneling, and coherent oscillations in small tunnel junctions”. In: *Journal of Low Temperature Physics* 62.3-4 (Feb. 1986), pp. 345–373.
- [56] K. K. Likharev. “Single-electron devices and their applications”. In: *Proceedings of the IEEE* 87.4 (Apr. 1999), pp. 606–632.
- [57] Y. X. Zhou et al. “Simple Fabrication of Molecular Circuits by Shadow Mask Evaporation”. In: *Nano Letters* 3.10 (Oct. 2003), pp. 1371–1374.
- [58] Y. Nakamura, C. Chen, and J.-S. Tsai. “100-K Operation of Al-Based Single-Electron Transistors”. In: *Japanese Journal of Applied Physics* 11A.2 (1996), pp. 1465–1467.
- [59] T. A. Fulton and G. J. Dolan. “Observation of single-electron charging effects in small tunnel junctions”. In: *Physical Review Letters* 59.1 (1987), pp. 109–112.
- [60] H. Wolf et al. “Investigation of the offset charge noise in single electron tunneling devices”. In: *IEEE Transactions on Instrumentation and Measurement* 46.2 (Apr. 1997), pp. 303–306.
- [61] Y. Y. Wei et al. “Single-electron transistor as an electrometer measuring chemical potential variations”. In: *Applied Physics Letters* 71.17 (1997), p. 2514.
- [62] J. Kim et al. “Effect of cross capacitance in three-junction single-electron transistors”. In: *Physical Review B* 59.3 (Jan. 1999), pp. 1629–1632.
- [63] K. K. Yadavalli et al. “Single electron memory devices utilizing Al<sub>2</sub>O<sub>3</sub> tunnel oxide barriers”. In: *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures* 22.6 (2004), p. 3119.
- [64] T. F. Li et al. “Low-frequency charge noise in suspended aluminum single-electron transistors”. In: *Applied Physics Letters* 91 (2007), p. 033107.
- [65] T. Matsui et al. “Evaluation of Single-Electron Transistor as Nanoscale Thermometer for a Cryogenic Radiation Detector”. In: *Journal of Nuclear Science and Technology* 45.6 (2008), pp. 73–77.
- [66] V. Tuboltsev et al. “Silicon-based Coulomb blockade thermometer with Schottky barriers”. In: *Applied Physics Letters* 104.16 (Apr. 2014), p. 163507.
- [67] H. T. A. Brenning et al. “A single electron transistor on an atomic force microscope probe”. In: *Nano Letters* 6.5 (2006), pp. 937–941.
- [68] T. Ferrus et al. “Detection of charge motion in a non-metallic silicon isolated double quantum dot”. In: *New Journal of Physics* 13 (2011), p. 103012.

- [69] Y. Ono et al. “Si complementary single-electron inverter with voltage gain”. In: *Applied Physics Letters* 76.21 (2000), p. 3121.
- [70] A. Aassime et al. “Radio-frequency single-electron transistor as readout device for qubits: Charge sensitivity and backaction”. In: *Physical Review Letters* 86.15 (2001), pp. 3376–3379.
- [71] H. Scherer et al. “Characterization of all-chromium tunnel junctions and single-electron tunneling devices fabricated by direct-writing multilayer technique”. In: *Journal of Applied Physics* 86.12 (1999), p. 6956.
- [72] X. Luo, A. O. Orlov, and G. L. Snider. “Origin of Coulomb blockade oscillations in single-electron transistors fabricated with granulated Cr/Cr<sub>2</sub>O<sub>3</sub> resistive microstrips”. In: *Microelectronics Journal* 36.3-6 (Mar. 2005), pp. 308–312.
- [73] M. V. Feigel’Man and A. S. Ioselevich. “Coulomb blockade for tunneling through a long island”. In: *Physical Review B* 90.11 (Sept. 2014), p. 115108.
- [74] L. P. Kouwenhoven, D. G. Austing, and S. Tarucha. “Few-electron quantum dots”. In: *Reports on Progress in Physics* 64.6 (June 2001), pp. 701–736.
- [75] F. A. Zwanenburg et al. “Silicon quantum electronics”. In: *Reviews of Modern Physics* 85.3 (July 2013), pp. 961–1019.
- [76] C. Joachim, J. K. Gimzewski, and a. Aviram. “Electronics using hybrid-molecular and mono-molecular devices.” In: *Nature* 408.6812 (Nov. 2000), pp. 541–8.
- [77] J. Park et al. “Coulomb blockade and the Kondo effect in single-atom transistors.” In: *Nature* 417.6890 (June 2002), pp. 722–5.
- [78] K. Miyaji and T. Hiramoto. “Control of full width at half maximum of Coulomb oscillation in silicon single-hole transistors at room temperature”. In: *Applied Physics Letters* 91.5 (2007), p. 053509.
- [79] V. N. Golovach et al. “Single-dopant resonance in a single-electron transistor”. In: *Physical Review B* 83.7 (2011), p. 075401.
- [80] M. Fuechsle et al. “A single-atom transistor”. In: *Nature Nanotechnology* 7.4 (2012), pp. 242–246.
- [81] S. J. Ray and R. Chowdhury. “Double gated single molecular transistor for charge detection”. In: *Journal of Applied Physics* 116.3 (July 2014), p. 034307.
- [82] R. M. Metzger. “Unimolecular Electronics”. In: *Chemical Reviews* 115.11 (2015), pp. 5056–5115.
- [83] M. Yoshihira et al. “Single electron transistors with ultra-thin Au nanowires as a single Coulomb island”. In: *Applied Physics Letters* 102.20 (2013), p. 203117.

- [84] J. Fang et al. “Room-temperature characterization of gold self-assembled single electron tunneling devices”. In: *Microelectronic Engineering* 108 (Aug. 2013), pp. 1–4.
- [85] S. I. Khondaker, K. Luo, and Z. Yao. “The fabrication of single-electron transistors using dielectrophoretic trapping of individual gold nanoparticles.” In: *Nanotechnology* 21.9 (Mar. 2010), p. 095204.
- [86] R. Hanson et al. “Spins in few-electron quantum dots”. In: *Reviews of Modern Physics* 79.4 (Oct. 2007), pp. 1217–1265.
- [87] S. J. Angus et al. “Gate-defined quantum dots in intrinsic silicon.” In: *Nano letters* 7.7 (July 2007), pp. 2051–5.
- [88] G. J. Podd et al. “Charge sensing in intrinsic silicon quantum dots”. In: *Applied Physics Letters* 96.8 (2010), p. 082104.
- [89] K. Ohkura, T. Kitade, and A. Nakajima. “Periodic Coulomb oscillations in Si single-electron transistor based on multiple islands”. In: *Journal of Applied Physics* 98.12 (2005), p. 124503.
- [90] S. J. Shin et al. “Si-based ultrasmall multiswitching single-electron transistor operating at room-temperature”. In: *Applied Physics Letters* 97.10 (2010), p. 103101.
- [91] S. J. Shin et al. “Room-temperature charge stability modulated by quantum effects in a nanoscale silicon island.” In: *Nano letters* 11.4 (Apr. 2011), pp. 1591–7.
- [92] M. A. Rafiq et al. “Conduction bottleneck in silicon nanochain single electron transistors operating at room temperature”. In: *Japanese Journal of Applied Physics* 51.2 PART 1 (Feb. 2012), p. 025202.
- [93] V. Deshpande et al. “Scaling of trigate nanowire (NW) MOSFETs to sub-7 nm width: 300 K transition to Single Electron Transistor”. In: *Solid-State Electronics* 84.c (2013), pp. 179–184.
- [94] T. Altebaeumer and H. Ahmed. “Tunnel Barrier Formation in Silicon Nanowires”. In: *Japanese Journal of Applied Physics* 42.Part 1, No. 2A (Feb. 2003), pp. 414–417.
- [95] T. Kudo and A. Nakajima. “Highly sensitive ion detection using Si single-electron transistors”. In: *Applied Physics Letters* 98.12 (2011), p. 123705.
- [96] T. Kudo and A. Nakajima. “Biomolecule detection based on Si single-electron transistors for highly sensitive integrated sensors on a single chip”. In: *Applied Physics Letters* 100.2 (2012), p. 023704.
- [97] A. Nakajima. “Application of Single-Electron Transistor to Biomolecule and Ion Sensors”. In: *Applied Sciences* 6.4 (2016), p. 94.

- [98] R. L. Puurunen. “Surface chemistry of atomic layer deposition: A case study for the trimethylaluminum/water process”. In: *Journal of Applied Physics* 97.12 (2005), p. 121301.
- [99] S. Bécu, S. Crémer, and J. L. Autran. “Capacitance non-linearity study in Al<sub>2</sub>O<sub>3</sub> MIM capacitors using an ionic polarization model”. In: *Microelectronic Engineering* 83.11-12 (2006), pp. 2422–2426.
- [100] N. Alimardani and J. F. Conley. “Enhancing metal-insulator-insulator-metal tunnel diodes via defect enhanced direct tunneling”. In: *Applied Physics Letters* 105.8 (2014), p. 082902.
- [101] A. Rossi and D. G. Hasko. “Microwave-assisted transport via localized states in degenerately doped Si single electron transistors”. In: *Journal of Applied Physics* 108.3 (2010), p. 034509.
- [102] H. C. George et al. “Platinum single-electron transistors with tunnel barriers made by atomic layer deposition”. In: *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures* 28.6 (2010), pp. 6–8.
- [103] M. S. McConnell et al. “Atomic layer deposition of Al<sub>2</sub>O<sub>3</sub> for single electron transistors utilizing Pt oxidation and reduction”. In: *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* 34.1 (2016), 01A139.
- [104] G. Karbasian et al. “Metal-Insulator-Metal Single Electron Transistors with Tunnel Barriers Prepared by Atomic Layer Deposition”. In: *Applied Sciences* 7.3 (2017), p. 246.
- [105] H. Birey. “Thickness dependence of the dielectric constant and resistance of Al<sub>2</sub>O<sub>3</sub> films”. In: *Journal of Applied Physics* 48.12 (1977), pp. 5209–5212.
- [106] L. Fry-Bouriaux et al. “Field-enhanced direct tunneling in ultrathin atomic-layer-deposition-grown Au-Al<sub>2</sub>O<sub>3</sub>-Cr metal-insulator-metal structures”. In: *Physical Review B* 96.11 (2017), pp. 1–11.
- [107] J. W. Haus et al. “Quantum conductivity for metal-insulator-metal nanostructures”. In: *Journal of the Optical Society of America B* 31.2 (Jan. 2014), p. 259.
- [108] R. Landauer and T. Martin. “Barrier interaction time in tunneling”. In: *Reviews of Modern Physics* 66.1 (1994), pp. 217–228.
- [109] A. S. Landsman et al. “Ultrafast resolution of tunneling delay time”. In: *Optica* 1.5 (2014), p. 343.
- [110] L. Torlina et al. “Interpreting attoclock measurements of tunnelling times”. In: *Nature Physics* 11.6 (2015), pp. 503–508.
- [111] D. V. Averin and Y. Nazarov. “Virtual electron diffusion during quantum tunneling of the electric charge”. In: *Physical Review Letters* 65.19 (1990), p. 2446.

- [112] J. P. Pekola et al. “Single-electron current sources: Toward a refined definition of the ampere”. In: *Reviews of Modern Physics* 85.4 (Oct. 2013), pp. 1421–1472.
- [113] P. Delsing et al. “Observation of single-electron-tunneling oscillations”. In: *Physical Review B* 42.12 (1990), pp. 7439–7449.
- [114] G. Y. Hu and R. F. O’Connell. “Exact solution for the charge soliton in a one-dimensional array of small tunnel junctions”. In: *Physical Review B* 49.23 (1994), pp. 773–776.
- [115] A. N. Korotkov et al. “Single-Electron Transistors as Ultrasensitive Electrometers”. In: *Single-Electron Tunneling and Mesoscopic Devices*. Springer, 1992, pp. 45–59.
- [116] A. N. Korotkov and M. A. Paalanen. “Charge sensitivity of radio frequency single-electron transistor”. In: *Applied Physics Letters* 74.26 (1999), p. 4052.
- [117] N. Zimmerman, J. Cobb, and A. Clark. “Modulation of the charge of a single-electron transistor by distant defects”. In: *Physical Review B* 56.12 (Sept. 1997), pp. 7675–7678.
- [118] A. N. Tavkhelidze and J. Mygind. “Low-frequency noise in single electron tunneling transistor”. In: *Journal of Applied Physics* 83.1 (1998), pp. 310–317.
- [119] Y. Sagi et al. “Spectrum of two-level systems with discrete frequency fluctuations”. In: *Physical Review Letters* 104.25 (2010), pp. 1–4.
- [120] E. Paladino et al. “ $1/f$  noise: Implications for solid-state quantum information”. In: *Reviews of Modern Physics* 86.2 (2014), pp. 361–418.
- [121] C. Müller et al. “Interacting two-level defects as sources of fluctuating high-frequency noise in superconducting circuits”. In: *Physical Review B* 92 (2015), p. 035442.
- [122] P. W. Anderson, B. I. Halperin, and C. M. Varma. “Anomalous low-temperature thermal properties of glasses and spin glasses”. In: *Philosophical Magazine* 25.1 (1972), pp. 1–9.
- [123] T. Songjaroen et al. “Novel, simple and low-cost alternative method for fabrication of paper-based microfluidics by wax dipping”. In: *Talanta* 85.5 (2011), pp. 2587–2593.
- [124] T. Songjaroen et al. “Blood separation on microfluidic paper-based analytical devices”. In: *Lab on a Chip* 12 (2012), pp. 3392–3398.
- [125] E. Yildiz-Ozturk and O. Yesil-Celiktas. “Diffusion phenomena of cells and biomolecules in microfluidic devices”. In: *Biomicrofluidics* 9.5 (2015), p. 052606.
- [126] J. C. Love et al. “Self-assembled monolayers of thiolates on metals as a form of nanotechnology”. In: *Chemical Reviews* 105 (2005), pp. 1103–1169.

- [127] N. K. Chaki and K. Vijayamohan. “Self-assembled monolayers as a tunable platform for biosensor applications”. In: *Biosensors and Bioelectronics* 17.1-2 (2002), pp. 1–12.
- [128] M. Wang et al. “Self-assembled silane monolayers: fabrication with nanoscale uniformity.” In: *Langmuir* 21.5 (Mar. 2005), pp. 1848–57.
- [129] Z. Matharu et al. “Self-assembled monolayer for low density lipoprotein detection.” In: *Journal of molecular recognition : JMR* 21.6 (2008), pp. 419–24.
- [130] J. S. Daniels and N. Pourmand. “Label-free impedance biosensors: Opportunities and Challenges”. In: *Electroanalysis* 19.12 (2007), pp. 1239–1257.
- [131] E. Pensa et al. “The chemistry of the sulfur-gold interface: In search of a unified model”. In: *Accounts of Chemical Research* 45.8 (2012), pp. 1183–1192.
- [132] C. A. Canaria et al. “Formation and removal of alkylthiolate self-assembled monolayers on gold in aqueous solutions”. In: *Lab on a Chip* 6.2 (2006), p. 289.
- [133] D. Evans et al. “Electrical protein detection in cell lysates using high-density peptide-aptamer microarrays.” In: *Journal of biology* 7.1 (Jan. 2008), p. 3.
- [134] B. Chen, A. Parashar, and S. Pandey. “Folded floating-gate CMOS biosensor for the detection of charged biochemical molecules”. In: *IEEE Sensors Journal* 11.11 (2011), pp. 2906–2910.
- [135] R. Y. Lai et al. “Differential labeling of closely spaced biosensor electrodes via electrochemical lithography”. In: *Langmuir* 22.4 (2006), pp. 1932–1936.
- [136] A.-L. Haag et al. “Selective in situ potential-assisted SAM formation on multi electrode arrays”. In: *Nanotechnology* 27.45 (2016), p. 455501.
- [137] J. P. Labukas and G. S. Ferguson. “Direct route to well-defined, chemically diverse electrode arrays”. In: *Langmuir* 27.7 (2011), pp. 3219–3223.
- [138] C. Wälti et al. “Direct selective functionalization of nanometer-separated gold electrodes with DNA oligonucleotides”. In: *Langmuir* 19.4 (2003), pp. 981–984.
- [139] M. Szymonik, A. G. Davies, and C. Walti. “DNA self-assembly-driven positioning of molecular components on nanopatterned surfaces”. In: *Nanotechnology* 27.39 (2016), p. 395301.
- [140] J. C. McDonald et al. “Fabrication of microfluidic systems in poly(dimethylsiloxane).” In: *Electrophoresis* 21.1 (2000), pp. 27–40.
- [141] S. Ferrari et al. “Atomic layer deposited Al<sub>2</sub>O<sub>3</sub> as a capping layer for polymer based transistors”. In: *Organic Electronics: physics, materials, applications* 8.4 (2007), pp. 407–414.
- [142] N. F. Mott and E. a. Davis. “Conduction in non-crystalline systems”. In: *Philosophical Magazine* 17.150 (1968), pp. 1269–1284.

- [143] P. Thissen, M. Valtiner, and G. Grundmeier. “Stability of phosphonic acid self-assembled monolayers on amorphous and single-crystalline aluminum oxide surfaces in aqueous solution.” In: *Langmuir : the ACS journal of surfaces and colloids* 26.1 (Jan. 2010), pp. 156–64.
- [144] N. S. Bakhvalov et al. “Single-electron solitons in one-dimensional tunnel structures”. In: *Sov. Phys. JETP* 68.3 (1989), pp. 581–587.
- [145] F. Ruggeri et al. “Single-molecule electrometry”. In: *Nature Nanotechnology* 12.5 (2017), pp. 488–495.
- [146] H. Ban, B. Lin, and Z. Song. “Effect of electrical double layer on electric conductivity and pressure drop in a pressure-driven microchannel flow”. In: *Biomicrofluidics* 4.1 (2010), pp. 1–13.
- [147] G. S. Hartley. “The application of the Debye-Huckel theory to colloidal electrolytes.” In: *Transactions of the Faraday Society* 31 (1935), pp. 31–50.
- [148] S. Hideshima et al. “Theoretical Optimization Method of Buffer Ionic Concentration for Protein Detection Using Field Effect Transistors”. In: *Journal of The Electrochemical Society* 157.12 (2010), J410.
- [149] A. M. Smith, A. A. Lee, and S. Perkin. “The Electrostatic Screening Length in Concentrated Electrolytes Increases with Concentration”. In: *Journal of Physical Chemistry Letters* 7.12 (2016), pp. 2157–2163.
- [150] A. A. Lee et al. “Underscreening in concentrated electrolytes”. In: *Faraday Discussions* (2017), pp. 1–21.
- [151] L. S. Kuzmin et al. “Single-electron charging effects in one-dimensional arrays of ultrasmall tunnel junctions”. In: *Physical Review Letters* 62.21 (1989), pp. 2539–2542.
- [152] C. Auth et al. “45nm High-k+Metal Gate Strain-Enhanced Transistors”. In: *Intel Journal of Technology* 12.45 (2008), pp. 77–85.
- [153] W. Yang et al. “The Integration of Sub-10 nm Gate Oxide on MoS2 with Ultra Low Leakage and Enhanced Mobility”. In: *Scientific Reports* 5 (2015), p. 11921.
- [154] *International Technology Roadmap for Semiconductors 2.0 Interconnect*. Accessed: 2017-06-14. 2015. URL: [https://www.semiconductors.org/main/2015\\_itrs\\_itwgs/](https://www.semiconductors.org/main/2015_itrs_itwgs/).
- [155] *Intel Manufacturing Day: Nodes must die, but Moore’s Law lives!* Accessed: 2017-06-14. 2017. URL: <https://www.semiwiki.com/forum/content/6698-intel-manufacturing-day-nodes-must-die-but-moores-law-lives.html>.

- [156] *International Technology Roadmap for Semiconductors 2.0 Lithography*. Accessed: 2017-06-14. 2015. URL: [https://www.semiconductors.org/main/2015\\_itrs\\_itwgs/](https://www.semiconductors.org/main/2015_itrs_itwgs/).
- [157] N. Mojarad, J. Gobrecht, and Y. Ekinici. “Beyond EUV lithography: a comparative study of efficient photoresists’ performance”. In: *Scientific Reports* 5.1 (2015), p. 9235.
- [158] *Suss MJB3 3 Mask Aligner*. Karl Suss. 2017. URL: <http://engineering.dartmouth.edu/microeng/equipment/maskaligner/Suss%20MJB-3%20operator%20Manual.pdf>.
- [159] L. F. Thompson and M. J. Bowden. “The Lithographic Process : The Physics”. In: *Introduction to Microlithography*. American Chemical Society, 1983, pp. 15–85.
- [160] *Microposit® S1813® Series Photo Resists*. Shipley. 2017. URL: <http://engineering.dartmouth.edu/microeng/processing/lithography/S1800.pdf>.
- [161] H. J. Levinson. *Principles of Lithography*. Third Edition. SPIE, 2010.
- [162] *LOR and PMGI Resists*. Micro-Chem. 2017. URL: <http://microchem.com/pdf/PMGI-Resists-data-sheetV-rhcredit-102206.pdf>.
- [163] *Nano™ PMMA and Copolymer*. Micro-Chem. 2017. URL: [http://microchem.com/pdf/PMMA\\_Data\\_Sheet.pdf](http://microchem.com/pdf/PMMA_Data_Sheet.pdf).
- [164] R. Clang. “Vacuum Evaporation”. In: *Handbook of Thin Film Technology*. McGraw Hill, 1970.
- [165] C. A. Neugebauer. “Condensation, Nucleation, and Growth of Thin Films”. In: *Handbook of Thin Film Technology*. McGraw Hill, 1970.
- [166] K. Bordo and H.-G. Rubahn. “Effect of Deposition Rate on Structure and Surface Morphology of Thin Evaporated Al Films on Dielectrics and Semiconductors”. In: *Materials Science* 18.4 (2012), pp. 313–317.
- [167] J. Zhang et al. “Gold Nanohole Array with Sub-1 nm Roughness by Annealing for Sensitivity Enhancement of Extraordinary Optical Transmission Biosensor”. In: *Nanoscale Research Letters* 10.1 (2015), p. 238.
- [168] V. Miikkulainen et al. “Crystallinity of inorganic films grown by atomic layer deposition: Overview and general trends”. In: *Journal of Applied Physics* 113.2 (2013), p. 021301.
- [169] K. K. Yadavalli et al. “Aluminum oxide tunnel barriers for single electron memory devices”. In: *Microelectronics Journal* 36.3 (2005), pp. 272–276.
- [170] M. Bareiß et al. “Printed array of thin-dielectric metal-oxide-metal (MOM) tunneling diodes”. In: *Journal of Applied Physics* 110.4 (2011), p. 044316.

- [171] K. Choi et al. "A Focused Asymmetric Metal-Insulator-Metal Tunneling Diode: Fabrication, DC Characteristics and RF Rectification Analysis". In: *IEEE Transactions on Electron Devices* 58.10 (Oct. 2011), pp. 3519–3528.
- [172] A. Sanchez et al. "The MOM tunneling diode: Theoretical estimate of its performance at microwave and infrared frequencies". In: *Journal of Applied Physics* 49.10 (1978), p. 5270.
- [173] J. A. Bean, A. Weeks, and G. D. Boreman. "Performance optimization of antenna-coupled Al/AlO<sub>x</sub>/Pt tunnel diode infrared detectors". In: *IEEE Journal of Quantum Electronics* 47.1 (2011), pp. 126–135.
- [174] N. Alimardani and J. F. Conley. "Step tunneling enhanced asymmetry in asymmetric electrode metal-insulator-insulator-metal tunnel diodes". In: *Applied Physics Letters* 102.14 (2013), p. 143501.
- [175] T. Yamaguchi et al. "Tunnel spin injection into graphene using Al<sub>2</sub>O<sub>3</sub> barrier grown by atomic layer deposition on functionalized graphene surface". In: *Journal of Magnetism and Magnetic Materials* 324.5 (Mar. 2012), pp. 849–852.
- [176] S. W. Lee et al. "Influences of metal, non-metal precursors, and substrates on atomic layer deposition processes for the growth of selected functional electronic materials". In: *Coordination Chemistry Reviews* 257.23-24 (Dec. 2013), pp. 3154–3176.
- [177] H. B. Profijt et al. "Plasma-Assisted Atomic Layer Deposition: Basics, Opportunities, and Challenges". In: *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* 29.5 (2011), p. 050801.
- [178] S. B. S. Heil et al. "Reaction mechanisms during plasma-assisted atomic layer deposition of metal oxides: A case study for Al<sub>2</sub>O<sub>3</sub>". In: *Journal of Applied Physics* 103.10 (2008), p. 103302.
- [179] E. Langereis et al. "Surface chemistry of plasma-assisted atomic layer deposition of Al<sub>2</sub>O<sub>3</sub> studied by infrared spectroscopy". In: *Applied Physics Letters* 92.23 (2008), p. 231904.
- [180] H.-Y. Li et al. "Method for Aluminum Oxide Thin Films Prepared through Low Temperature Atomic Layer Deposition for Encapsulating Organic Electroluminescent Devices". In: *Materials* 8 (2015), pp. 600–610.
- [181] K. B. Jinesh et al. "Dielectric Properties of Thermal and Plasma-Assisted Atomic Layer Deposited Al<sub>2</sub>O<sub>3</sub> Thin Films". In: *Journal of The Electrochemical Society* 158.2 (2011), G21.

- [182] C. M. Jackson et al. “Interface trap characterization of atomic layer deposition Al<sub>2</sub>O<sub>3</sub>/GaN metal-insulator-semiconductor capacitors using optically and thermally based deep level spectroscopies”. In: *Journal of Applied Physics* 113.20 (2013), p. 204505.
- [183] H. Spahr et al. “Impact of morphological defects on the electrical breakdown of ultra thin atomic layer deposition processed Al<sub>2</sub>O<sub>3</sub> layers”. In: *Thin Solid Films* 534 (May 2013), pp. 172–176.
- [184] J. L. van Hemmen et al. “Plasma and Thermal ALD of Al<sub>2</sub>O<sub>3</sub> in a Commercial 200mm ALD Reactor”. In: *Journal of The Electrochemical Society* 154.7 (2007), G165.
- [185] J. Haeberle et al. “Ellipsometry and XPS comparative studies of thermal and plasma enhanced atomic layer deposited Al<sub>2</sub>O<sub>3</sub>-films.” In: *Beilstein journal of nanotechnology* 4 (Jan. 2013), pp. 732–42.
- [186] S. Bals et al. “Annular dark field imaging in a TEM”. In: *Solid State Communications* 130.10 (2004), pp. 675–680.
- [187] R. Li et al. “An analytic model for accurate spring constant calibration of rectangular atomic force microscope cantilevers”. In: *Scientific Reports* 5.June (2015), p. 15828.
- [188] *EELS Atlas*. <http://www.eels.info/atlas>. Accessed: 2016-12-01. 2016.
- [189] V. Shevelko. *Atoms and Their Spectroscopic Properties*. Vol. 18. Springer Series on Atomic, Optical, and Plasma Physics. Springer, 1997.
- [190] H. Mullejans and J. Bruley. “Electron energy-loss spectroscopy (EELS); comparison with x-ray analysis”. In: *Journal De Physique* 3.C7 (1993), p. 2083.
- [191] *Model 7225 DSP Lock-in Amplifier*. Signal Recovery. 2017. URL: [http://www.ameteksi.com/-/media/ameteksi/download\\_links/documentations/7225/190491-a-mnl-b.pdf](http://www.ameteksi.com/-/media/ameteksi/download_links/documentations/7225/190491-a-mnl-b.pdf).
- [192] *Programmable LCR Bridge HM8118*. Rohde & Schwarz. 2015. URL: [https://cdn.rohde-schwarz.com/pws/dl\\_downloads/dl\\_common\\_library/dl\\_manuels/gb\\_1/h/hm8118\\_1/HM8118\\_UserManual\\_de\\_en\\_06.pdf](https://cdn.rohde-schwarz.com/pws/dl_downloads/dl_common_library/dl_manuels/gb_1/h/hm8118_1/HM8118_UserManual_de_en_06.pdf).
- [193] J. G. Simmons. “Conduction in thin dielectric films”. In: *Journal of Physics D: Applied Physics* 4.5 (1971), p. 202.
- [194] J. Oh et al. “Etch Behavior of ALD Al<sub>2</sub>O<sub>3</sub> on HfSiO and HfSiON Stacks in Acidic and Basic Etchants”. In: *Journal of The Electrochemical Society* 158.4 (2011), p. D217.

- [195] Z. H. Ni et al. “Reflection and Contrast Spectroscopy Graphene Thickness Determination Using Reflection and Contrast Spectroscopy”. In: *Nano letters* 7.9 (2007), pp. 2758–2763.
- [196] I. Jung et al. “Colors of graphene and graphene-oxide multilayers on various substrates”. In: *Nanotechnology* 23.2 (2012), p. 025708.
- [197] K. Grigoras, S. Franssila, and V.-M. Airaksinen. “Investigation of sub-nm ALD aluminum oxide films by plasma assisted etch-through”. In: *Thin Solid Films* 516.16 (June 2008), pp. 5551–5556.
- [198] A. J. Elliot et al. “Probing the nucleation of Al<sub>2</sub>O<sub>3</sub> in atomic layer deposition on aluminum for ultrathin tunneling barriers in Josephson junctions”. In: *IEEE Transactions on Applied Superconductivity* 23.3 (2013), pp. 1–5.
- [199] M. D. Groner et al. “Electrical characterization of thin Al<sub>2</sub>O<sub>3</sub> films grown by atomic layer deposition on silicon and various metal substrates”. In: *Thin Solid Films* 413.1–2 (2002), pp. 186–197.
- [200] H. Momida et al. “Theoretical study on dielectric response of amorphous alumina”. In: *Physical Review B* 73.5 (Feb. 2006), p. 054108.
- [201] T. A. Baker, C. M. Friend, and E. Kaxiras. “Atomic Oxygen Adsorption on Au(111) Surfaces with Defects”. In: *The Journal of Physical Chemistry C* 113.8 (2009), pp. 3232–3238.
- [202] K. Okazaki-Maeda and M. Kohyama. “Atomic oxygen adsorption on Au(100) and Au(111): Effects of coverage”. In: *Chemical Physics Letters* 492.4-6 (2010), pp. 266–271.
- [203] N. Gaillard et al. “In situ electric field simulation in metal/insulator/metal capacitors”. In: *Applied Physics Letters* 89.13 (2006), pp. 23–26.
- [204] W. S. Lau et al. “Surface smoothing effect of an amorphous thin film deposited by atomic layer deposition on a surface with nano-sized roughness”. In: *AIP Advances* 4.2 (2014), pp. 2012–2017.
- [205] O. M. E. Ylivaara et al. “Aluminum oxide from trimethylaluminum and water by atomic layer deposition: The temperature dependence of residual stress, elastic modulus, hardness and adhesion”. In: *Thin Solid Films* 552 (2014), pp. 124–135.
- [206] G. V. Hansson and S. A. Flodström. “Photoemission study of the bulk and surface electronic structure of single crystals of gold”. In: *Physical Review B* 18.4 (1978), pp. 1572–1585.
- [207] A. K. Jonscher. “Dielectric relaxation in solids”. In: *Journal of Physics D: Applied Physics* 32.14 (1999), R57–R70.

- [208] D. S. Campbell and A. R. Morley. “Electrical conduction in thin metallic, dielectric and metallic-dielectric films”. In: *Reports on Progress in Physics* 34.1 (1971), pp. 283–368.
- [209] J. Robertson. “High dielectric constant gate oxides for metal oxide Si transistors”. In: *Reports on Progress in Physics* 69.2 (Feb. 2006), pp. 327–396.
- [210] S. J. Kim et al. “Improvement of voltage linearity in high-k MIM capacitors using HfO<sub>2</sub>-SiO<sub>2</sub> stacked dielectric”. In: *IEEE Electron Device Letters* 25.8 (2004), pp. 538–540.
- [211] S. K. Lee et al. “Characterizing voltage linearity and leakage current of high density Al<sub>2</sub>O<sub>3</sub>/HfO<sub>2</sub>/Al<sub>2</sub>O<sub>3</sub> MIM capacitors”. In: *IEEE Electron Device Letters* 32.3 (2011), pp. 384–386.
- [212] Q.-X. Zhang et al. “Full ALD Al<sub>2</sub>O<sub>3</sub>/ZrO<sub>2</sub>/SiO<sub>2</sub>/ZrO<sub>2</sub>/Al<sub>2</sub>O<sub>3</sub> Stacks for High-Performance MIM Capacitors”. In: *IEEE Elec* 35.11 (2014), pp. 1121–1123.
- [213] J. A. Babcock et al. “Analog Characteristics of Metal – Insulator – Metal Capacitors Using PECVD Nitride Dielectrics”. In: *IEEE Electron Device Letters* 22.5 (2001), pp. 230–232.
- [214] A. Bag et al. “Graphene oxide-based flexible metal – insulator – metal capacitors”. In: *Semiconductor Science and Technology* 28 (2013), p. 055002.
- [215] E. Hourdakis, A. Travlos, and A. G. Nassiopoulou. “High-performance MIM capacitors with nanomodulated electrode surface”. In: *IEEE Transactions on Electron Devices* 62.5 (2015), pp. 1568–1573.
- [216] H. Hu et al. “MIM Capacitors Using Atomic-Layer-Deposited High-k (HfO<sub>2</sub>)<sub>1-x</sub>(Al<sub>2</sub>O<sub>3</sub>)<sub>x</sub> Dielectrics”. In: *IEEE Electron Device Letters* 24.2 (2003), pp. 60–62.
- [217] A. Kahn. “Fermi level, work function and vacuum level”. In: *Mater. Horiz.* 3 (2016), pp. 7–10.
- [218] D. Yamashita and A. Ishizaki. “In situ measurements of change in work function of Pt, Pd and Au surfaces during desorption of oxygen by using photoemission yield spectrometer in air”. In: *Applied Surface Science* 363 (2016), pp. 240–244.
- [219] Y. C. Yeo, T. J. King, and C. Hu. “Metal-dielectric band alignment and its implications for metal gate complementary metal-oxide-semiconductor technology”. In: *Journal of Applied Physics* 92.12 (2002), pp. 7266–7271.
- [220] N. C. Lindquist et al. “Engineering metallic nanostructures for plasmonics and nanophotonics”. In: *Reports on Progress in Physics* 75.3 (2012), p. 036501.
- [221] X. Chen et al. “Atomic layer lithography of wafer-scale nanogap arrays for extreme confinement of electromagnetic waves”. In: *Nature Communications* 4.May (Jan. 2013), p. 2361.

- [222] C. Ciracì et al. “Film-coupled nanoparticles by atomic layer deposition: Comparison with organic spacing layers”. In: *Applied Physics Letters* 104.2 (Jan. 2014), p. 023109.
- [223] J. J. Attema et al. “Work function anisotropy and surface stability of half-metallic Cr O<sub>2</sub>”. In: *Physical Review B* 77.16 (2008), pp. 1–9.
- [224] A. J. Hong et al. “Cr metal thin film memory”. In: *Journal of Applied Physics* 110.5 (2011), pp. 2–6.
- [225] S. Bécu, S. Crémer, and J. L. Autran. “Microscopic model for dielectric constant in metal-insulator-metal capacitors with high- permittivity metallic oxides”. In: *Applied Physics Letters* 88.052902 (2006), pp. 1–4.
- [226] W. Park et al. “Sub-10 nm feature chromium photomasks for contact lithography patterning of square metal ring arrays”. In: *Scientific Reports* 6.March (2016), p. 23823.
- [227] J. G. Simmons. “Generalized Formula for the Electric Tunnel Effect between Similar Electrodes Separated by a Thin Insulating Film”. In: *Journal of Applied Physics* 34.6 (1963), p. 1793.
- [228] B. Hu, J. Yao, and B. J. Hinds. “Nanogap electrodes formed at the exposed edge of Au/self-assembled monolayer/Al<sub>2</sub>O<sub>3</sub>/Au tunnel structures grown by atomic layer deposition”. In: *Applied Physics Letters* 97.20 (2010), pp. 2010–2012.
- [229] M. Manheller et al. “Reliable fabrication of 3 nm gaps between nanoelectrodes by electron-beam lithography.” In: *Nanotechnology* 23.12 (Mar. 2012), p. 125302.
- [230] R. T. Tung. “The physics and chemistry of the Schottky barrier height”. In: *Applied Physics Reviews* 1.1 (Mar. 2014), p. 011304.
- [231] S. Ganichev et al. “Distinction between the Poole-Frenkel and tunneling models of electric-field-stimulated carrier emission from deep levels in semiconductors”. In: *Physical Review B* 61.15 (Apr. 2000), pp. 10361–10365.
- [232] Y. Qi et al. “Modified Schottky emission to explain thickness dependence and slow depolarization in BaTiO<sub>3</sub> nanowires”. In: *Physical Review B* 91.24 (2015), p. 245431.
- [233] J. G. Simmons. “Transition from electrode-limited to bulk-limited conduction processes in metal-insulator-metal systems”. In: *Physical Review* 166.3 (1968), pp. 912–920.
- [234] J. Frenkel. “On Pre-Breakdown Phenomena in Insulators and Electronic Semiconductors”. In: *Physical Review Letters* 54 (June 1938), pp. 647–648.
- [235] C.-C. Yeh et al. “Frenkel-Poole trap energy extraction of atomic layer deposited Al<sub>2</sub>O<sub>3</sub> and HfxAlyO thin films”. In: *Applied Physics Letters* 91.11 (2007), p. 113521.

- [236] J. Kolodzey et al. “Electrical conduction and dielectric breakdown in aluminum oxide insulators on silicon”. In: *IEEE Transactions on Electron Devices* 47.1 (2000), pp. 121–128.
- [237] D. S. Jeong and C. S. Hwang. “Tunneling-assisted Poole-Frenkel conduction mechanism in HfO<sub>2</sub> thin films”. In: *Journal of Applied Physics* 98.11 (2005), p. 113701.
- [238] D. S. Jeong, H. B. Park, and C. S. Hwang. “Reasons for obtaining an optical dielectric constant from the Poole-Frenkel conduction behavior of atomic-layer-deposited HfO<sub>2</sub> films”. In: *Applied Physics Letters* 86.7 (2005), pp. 1–3.
- [239] M. L. Huang et al. “Energy-band parameters of atomic-layer-deposition Al<sub>2</sub>O<sub>3</sub>/InGaAs heterostructure”. In: *Applied Physics Letters* 89.1 (2006), p. 012903.
- [240] N. Alimardani et al. “Investigation of the impact of insulator material on the performance of dissimilar electrode metal-insulator-metal diodes”. In: *Journal of Applied Physics* 116.2 (July 2014), p. 024508.
- [241] S. K. Lee et al. “Structure and disorder in amorphous alumina thin films: Insights from high-resolution solid-state NMR”. In: *Journal of Physical Chemistry C* 114.32 (2010), pp. 13890–13894.
- [242] J. G. Simmons. “Theory of Metallic Contacts on High Resistivity Solids (I) Shallow Traps”. In: *Journal of Physics and Chemistry of Solids* 32 (1971), pp. 1987–1999.
- [243] J. G. Simmons. “Theory of Metallic Contacts on High Resistivity Solids (II) Deep Traps”. In: *Journal of Physics and Chemistry of Solids* 32 (1971), pp. 2581–2591.
- [244] M. Specht et al. “Transport mechanisms in atomic-layer-deposited Al<sub>2</sub>O<sub>3</sub> dielectrics”. In: *Applied Physics Letters* 84.16 (2004), p. 3076.
- [245] J. Robertson and B. Falabretti. “Band offsets of high K gate oxides on III-V semiconductors”. In: *Journal of Applied Physics* 100.1 (2006), p. 014111.
- [246] T. W. Hickmott. “Polarization and Fowler–Nordheim tunneling in anodized Al–Al<sub>2</sub>O<sub>3</sub>–Au diodes”. In: *Journal of Applied Physics* 87.11 (2000), pp. 7903–7912.
- [247] K. El Hajjam et al. “Highly transparent low capacitance plasma enhanced atomic layer deposition Al<sub>2</sub>O<sub>3</sub>–HfO<sub>2</sub> tunnel junction engineering”. In: *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* 32.1 (2014), 01A132.
- [248] Y. Bruynseraede et al. “Fundamentals of Giaever and Josephson Tunneling”. In: *The New Superconducting Electronics*. Dordrecht: Springer Netherlands, 1993, pp. 1–28.
- [249] R. G. Forbes. “On the need for a tunneling pre-factor in Fowler–Nordheim tunneling theory”. In: *Journal of Applied Physics* 103.11 (2008), p. 114911.

- [250] R. G. Forbes. “Extraction of emission parameters for large-area field emitters, using a technically complete Fowler–Nordheim-type equation”. In: *Nanotechnology* 23.9 (2012), p. 095706.
- [251] E. W. Cowell et al. “Barrier height estimation of asymmetric metal-insulator-metal tunneling diodes”. In: *Journal of Applied Physics* 114.21 (2013), p. 213703.
- [252] Y. Yourdshahyan et al. “First-principles calculations on the atomic and electronic structure of  $\kappa$ -Al<sub>2</sub>O<sub>3</sub>”. In: *Physical Review B* 56.14 (1997), pp. 8553–8558.
- [253] T. V. Perevalov, V. A. Gritsenko, and V. V. Kaichev. “Electronic structure of aluminum oxide: ab initio simulations of  $\alpha$  and  $\gamma$  phases and comparison with experiment for amorphous films”. In: *The European Physical Journal Applied Physics* 52.3 (2010), p. 30501.
- [254] L. S. Dorneles et al. “The use of Simmons’ equation to quantify the insulating barrier parameters in Al/AlO<sub>x</sub>/Al tunnel junctions”. In: *Applied Physics Letters* 82.17 (2003), pp. 2832–2834.
- [255] T. Hartman and J. Chivian. “Electron Tunneling Through Thin Aluminum Oxide Films”. In: *Physical Review* 134.4A (1964), A1094–A1101.
- [256] E. J. Patiño and N. G. Kelkar. “Experimental determination of tunneling characteristics and dwell times from temperature dependence of Al/Al<sub>2</sub>O<sub>3</sub>/Al junctions”. In: *Applied Physics Letters* 107.25 (2015), pp. 10–14.
- [257] J. Lambe. “Observation of surface states on gold films by electron tunneling spectroscopy”. In: *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* 2.2 (1984), p. 792.
- [258] D. P. Woodruff, W. A. Royer, and N. V. Smith. “Empty surface states, images states and band edge on Au(111)”. In: *Physical Review B* 34.2 (1986), pp. 764–767.
- [259] L. C. Davis et al. “Theory of the local density of surface states on a metal: Comparison with scanning tunneling spectroscopy of a Au(111) surface”. In: *Physical Review B* 43.5 (1991), pp. 3821–3830.
- [260] W. Chen, H. Ahmed, and K. Nakazoto. “Coulomb blockade at 77 K in nanoscale metallic islands in a lateral nanostructure”. In: *Applied Physics Letters* 66.24 (1995), p. 3383.
- [261] P. Kowalczyk et al. “STS investigations of temperature dependence of Au(111) surface state energy position”. In: *Surface Science* 600.8 (2006), pp. 1604–1607.
- [262] R. Rösch and R. Schuster. “Tunneling spectroscopy of clean and adsorbate-covered gold surfaces in humid air, measured with fast bias voltage ramps”. In: *Surface Science* 631 (2015), pp. 105–111.

- [263] J. Seo et al. “Transmission of topological surface states through surface barriers.” In: *Nature* 466.7304 (2010), pp. 343–346.
- [264] B. Yan et al. “Topological States on the Gold Surface”. In: *Nature communications* 6 (2015), p. 10167.
- [265] H. Albrecht, G. Keller, and F. Thieme. “A new approach to the elastic tunneling current in metal-insulator-metal diodes”. In: *Surface Science* 69.2 (1977), pp. 677–688.
- [266] N. Alimardani et al. “Impact of electrode roughness on metal-insulator-metal tunnel diodes with atomic layer deposited Al<sub>2</sub>O<sub>3</sub> tunnel barriers”. In: *Journal of Vacuum Science & Technology A: Vacuum, Surfaces, and Films* 30.2012 (2012), 01A113.
- [267] E. W. Weisstein. *Circle-Circle Intersection*. <http://mathworld.wolfram.com/Circle-CircleIntersection.html>. From *MathWorld*—A Wolfram Web Resource. Accessed: 2016-12-01.
- [268] C. W. Miller et al. “Impact of interfacial roughness on tunneling conductance and extracted barrier parameters”. In: *Applied Physics Letters* 90.4 (2007), p. 043513.
- [269] W. S. Lau et al. “Mechanism of IV asymmetry of MIM capacitors based on high-*k* dielectric”. In: *Semiconductor Technology International Conference (CSTIC), 2015 China*. Mar. 2015, pp. 1–3.
- [270] M. V. Krylov and R. A. Suris. “Electron Tunneling Through Layers with Statistically Rough Surfaces”. In: *Soviet Physics Jetp* 61.6 (1985), pp. 1303–1307.
- [271] A. Modinos, G. C. Aers, and B. V. Paranjape. “Quantum size effect in normal-metal tunneling”. In: *Physical Review B* 19.8 (1979), pp. 3996–4011.
- [272] W. A. Harrison. “Tunneling from an independent-particle point of view”. In: *Physical Review* 123.1 (1961), pp. 85–89.
- [273] R. Tsu and L. Esaki. “Tunneling in a finite superlattice”. In: *Applied Physics Letters* 22.11 (1973), p. 562.
- [274] I. E. Hashem, N. H. Rafat, and E. A. Soliman. “Theoretical Study of Metal-Insulator-Metal Tunneling Diode Figures of Merit”. In: *IEEE Journal of Quantum Electronics* 49.1 (2013), pp. 72–79.
- [275] M. Alhazmi et al. “NSTOA-13-RA-108 Comparison of the Effects of Varying of Metal Electrode in Metal-Insulator-Metal Diodes with multi-dielectric layers”. In: *Austin Journal of Nanomedicine & Nanotechnology* 2.2 (2014), pp. 1–4.
- [276] J. Lambe and R. C. Jaklevic. “Molecular Vibration Spectra by Inelastic Electron Tunneling”. In: *Physical Review* 165.3 (1968), pp. 821–832.

- [277] M. Newville et al. *LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python*. 2014. URL: <https://doi.org/10.5281/zenodo.11813>.
- [278] S. Y. No et al. “Property Changes of Aluminum Oxide Thin Films Deposited by Atomic Layer Deposition under Photon Radiation”. In: *Journal of The Electrochemical Society* 153.6 (2006), F87.
- [279] L. Lai and E. A. Irene. “Area evaluation of microscopically rough surfaces”. In: *J. Vac. Sci. Technol. B* 17.1 (1999), pp. 33–39.
- [280] G. M. Ramchadani. “Energy band structure of gold”. In: *J. Phys. C Metal Phys. Suppl* 3.1 (1970), S1–9.
- [281] C. Fadley and D. Shirley. “Electronic densities of states from x-ray photoelectron spectroscopy”. In: *Journal of Research of the National Bureau of Standards Section A: Physics and Chemistry* 74A.4 (1970), p. 543.
- [282] K. Liang and W. Salaneck. “X-ray photoemission studies of thin gold films”. In: *Solid State Communications* 19 (1976), pp. 1–5.
- [283] T. Rangel et al. “Band structure of gold from many-body perturbation theory”. In: *Physical Review B* 86.April 2016 (2012), p. 125125.
- [284] J. L. Fry et al. “Hartree band structure, Fermi surface, and nesting wave vector for paramagnetic chromium”. In: *Physical Review B* 21.2 (1980), pp. 384–390.
- [285] P. E. S. Persson and L. I. Johansson. “Bulk band structure of chromium”. In: *Physical Review B* 34.4 (1986), pp. 2284–2292.
- [286] W. F. Brinkman. “Tunneling Conductance of Asymmetrical Barriers”. In: *Journal of Applied Physics* 41.5 (1970), p. 1915.
- [287] A. Sekiyama et al. “The prominent 5d-orbital contribution to the conduction electrons in gold”. In: *New Journal of Physics* 12 (2010), pp. 1–8.
- [288] J. M. Pitarke, P. M. Echenique, and F. Flores. “Apparent barrier height for tunneling electrons in STM”. In: *Surface Science* 217.1-2 (1989), pp. 267–275.
- [289] A. Schenk and G. Heiser. “Modeling and simulation of tunneling through ultra-thin gate dielectrics”. In: *Journal of Applied Physics* 81.12 (1997), pp. 7900–7908.
- [290] S. Grover et al. “Applicability of Metal / Insulator / Metal ( MIM ) Diodes to Solar Rectennas”. In: *IEEE Journal of Photovoltaics* 1.1 (2011), pp. 78–83.
- [291] A. Mayer et al. “Analysis of the efficiency with which geometrically asymmetric metal–vacuum–metal junctions can be used for the rectification of infrared and optical radiations”. In: *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures* 30.3 (2012), p. 031802.

- [292] P. Periasamy et al. “Metal-insulator-metal diodes: Role of the insulator layer on the rectification performance”. In: *Advanced Materials* 25.9 (2013), pp. 1301–1308.
- [293] K. P. O’Donnell and X. Chen. “Temperature dependence of semiconductor band gaps”. In: *Applied Physics Letters* 58.25 (1991), pp. 2924–2926.
- [294] P. B. Allen and M. Cardona. “Temperature dependence of the direct gap of Si and Ge”. In: *Physical Review B* 27.8 (1983), pp. 4760–4769.
- [295] A. Collins et al. “Indirect energy gap of <sup>13</sup>C diamond”. In: *Physical Review Letters* 65.7 (1990), pp. 891–894.
- [296] C. Rossignol et al. “Elastic properties of ultrathin permalloy/alumina multilayer films using picosecond ultrasonics and Brillouin light scattering”. In: *Physical Review B* 70.9 (2004), pp. 1–12.
- [297] G. Fiori et al. “Electronics based on two-dimensional materials”. In: *Nature Nanotechnology* 9.10 (2014), pp. 768–779.
- [298] F. Schwierz, R. Granzner, and J. Pezoldt. “Two-dimensional materials and their prospects in transistor electronics”. In: *Nanoscale* 7 (2015), pp. 8261–8283.
- [299] G. R. Bhimanapati et al. “Recent Advances in Two-Dimensional Materials beyond Graphene”. In: *ACS nano* 9.12 (2015), pp. 11509–11539.
- [300] K. S. Novoselov et al. “2D materials and van der Waals heterostructures”. In: *Science* 353.6298 (2016), aac9439.
- [301] T. Ihn et al. “Graphene single-electron transistors”. In: *Materials Today* 13.3 (Mar. 2010), pp. 44–50.
- [302] H. J. Song et al. “Large scale metal-free synthesis of graphene on sapphire and transfer-free device fabrication”. In: *Nanoscale* 4.10 (May 2012), p. 3050.
- [303] Y. Banadaki and A. Srivastava. “Effect of Edge Roughness on Static Characteristics of Graphene Nanoribbon Field Effect Transistor”. In: *Electronics* 5.1 (2016), p. 11.
- [304] T. Cheiwchanchamnangij and W. R. L. Lambrecht. “Quasiparticle band structure calculation of monolayer, bilayer, and bulk MoS<sub>2</sub>”. In: *Physical Review B* 85.20 (2012), pp. 1–4.
- [305] W. Yang et al. “The Integration of Sub-10 nm Gate Oxide on MoS<sub>2</sub> with ultra low Leakage and Enhanced Mobility”. In: *Scientific Reports* 5 (2015), p. 11921.
- [306] X. Gu, B. Li, and R. Yang. “Layer thickness-dependent phonon properties and thermal conductivity of MoS<sub>2</sub>”. In: *Journal of Applied Physics* 119.8 (2016), pp. 1–27.

- [307] J. M. Gonzalez and I. I. Oleynik. “Layer-dependent properties of SnS<sub>2</sub> and SnSe<sub>2</sub> two-dimensional materials”. In: *Physical Review B* 94.12 (2016), pp. 1–10.
- [308] N. P. Dasgupta et al. “Recent Advances in Atomic Layer Deposition”. In: *Chemistry of Materials* 28.7 (2016), pp. 1943–1947.
- [309] R. Rammula et al. “Atomic layer deposition of aluminum oxide films on graphene”. In: *IOP Conference Series: Materials Science and Engineering* 49.1 (2013), p. 012014.
- [310] X. Liu and J. Shi. “Magnetic tunnel junctions with Al<sub>2</sub>O<sub>3</sub> tunnel barriers prepared by atomic layer deposition”. In: *Applied Physics Letters* 102.20 (2013), p. 202401.
- [311] J. Schmidt et al. “Atomic-layer-deposited aluminum oxide for the surface passivation of high-efficiency silicon solar cells”. In: *2008 33rd IEEE Photovoltaic Specialists Conference* (May 2008), pp. 1–5.
- [312] M. K. Hota, C. K. Sarkar, and C. K. Maiti. “Frequency-dependent dielectric response of HfTaO<sub>x</sub>-based metal-insulator-metal capacitors”. In: *Semiconductor Science and Technology* 27.8 (2012), p. 085002.
- [313] P. C. D. Hobbs et al. “Efficient waveguide-integrated tunnel junction detectors at 1.6  $\mu\text{m}$ .” In: *Optics express* 15.25 (Dec. 2007), pp. 16376–89.
- [314] F. Giustino and A. Pasquarello. “Theory of atomic-scale dielectric permittivity at insulator interfaces”. In: *Physical Review B* 71.14 (2005), pp. 1–13.
- [315] F. Hüser, T. Olsen, and K. S. Thygesen. “How dielectric screening in two-dimensional crystals affects the convergence of excited-state calculations: Monolayer MoS<sub>2</sub>”. In: *Physical Review B* 88.24 (2013), pp. 1–9.
- [316] T. T. Song et al. “The stability of aluminium oxide monolayer and its interface with two-dimensional materials.” In: *Scientific Reports* 6.March (2016), p. 29221.
- [317] C. Zhou and D. M. Newns. “Intrinsic dead layer effect and the performance of ferroelectric thin film capacitors”. In: *Journal of Engineering and Applied Science* 82.6 (1997), pp. 3081–3088.
- [318] A. A. Sirenko et al. “Soft-mode hardening in SrTiO<sub>3</sub> thin films”. In: *Nature* 404.6776 (2000), pp. 373–6.
- [319] M. Stengel and N. a. Spaldin. “Origin of the dielectric dead layer in nanoscale capacitors.” In: *Nature* 443.7112 (2006), pp. 679–682.
- [320] J. Liu et al. “Theoretical methods of domain structures in ultrathin ferroelectric films: A review”. In: *Materials* 6.9 (2014), pp. 6502–6568.
- [321] H. J. Kim, M. Osada, and T. Sasaki. “Advanced capacitor technology based on two-dimensional nanosheets”. In: *Japanese Journal of Applied Physics* 55.11 (2016), 1102A3.

- [322] K. Natori, D. Otani, and N. Sano. “Thickness dependence of the effective dielectric constant in a thin film capacitor”. In: *Applied Physics Letters* 73.5 (1998), pp. 632–634.
- [323] A. K. Jonscher. “Physical basis of dielectric loss”. In: *Nature* 253.5494 (1975), pp. 717–719.
- [324] R. M. Hill and A. K. Jonscher. “DC and AC conductivity in hopping electronic systems”. In: *Journal of Non-Crystalline Solids* 32.1-3 (1979), pp. 53–69.
- [325] G. D. Wilk, R. M. Wallace, and J. M. Anthony. “High- $\kappa$  gate dielectrics: Current status and materials properties considerations”. In: *Journal of Applied Physics* 89.10 (2001), p. 5243.
- [326] S. Baroni et al. “Phonons and related properties of extended systems from density-functional perturbation theory”. In: *Reviews of Modern Physics* 73.2 (2000), p. 52.
- [327] R. Resta. “Macroscopic polarization in crystalline dielectrics: The geometric phase approach”. In: *Reviews of Modern Physics* 66.3 (1994), pp. 899–915.
- [328] R. Resta and D. Vanderbilt. “Theory of polarization: A modern approach”. In: *Topics in Applied Physics* 105.Cm (2007), pp. 31–68.
- [329] N. A. Spaldin. “A beginners guide to the modern theory of polarization”. In: *Journal of Solid State Chemistry* 195 (2012), pp. 2–10.
- [330] R. H. Lyddane, R. G. Sachs, and E. Teller. “On the polar vibrations of alkali halides”. In: *Physical Review* 59.8 (1941), pp. 673–676.
- [331] Y. T. Chu et al. “Optical dielectric functions for amorphous Al<sub>2</sub>O<sub>3</sub> and gamma-Al<sub>2</sub>O<sub>3</sub>”. In: *Journal of Applied Physics* 64.7 (1988), pp. 3727–3730.
- [332] B. G. Frederick, G. Apai, and T. N. Rhodin. “Surface phonons in thin aluminum oxide films: Thickness, beam-energy, and symmetry-mixing effects”. In: *Physical Review B* 44.4 (1991), pp. 1880–1890.
- [333] A. S. Barker. “Long-wavelength soft modes, central peaks, and the Lyddane-Sachs-Teller relation”. In: *Physical Review B* 12.10 (1975), pp. 4071–4084.
- [334] S.-H. Cai et al. “Phase transformation mechanism between  $\gamma$ - and  $\theta$ -alumina”. In: *Physical Review B* 67.June 2016 (2003), pp. 1–10.
- [335] C. M. Fang and R. A. de Groot. “The nature of electron states in AlN and  $\alpha$ -Al<sub>2</sub>O<sub>3</sub>”. In: *Journal of Physics: Condensed Matter* 19.38 (2007), p. 386223.
- [336] T. W. Noh and A. J. Sievers. “Generalization of the Lyddane-Sachs-Teller Relation to Disordered Dielectrics”. In: *Physical Review Letters* 63.17 (1989), pp. 1800–1802.

- [337] A. J. Sievers and J. B. Page. “Generalized Lyddane-Sachs-Teller relation and disordered solids”. In: *Physical Review B* 41.6 (1990), pp. 3455–3459.
- [338] W. W. Peng et al. “Experimental and ab initio study of vibrational modes of stressed alumina films formed by oxidation of aluminium alloys under different atmospheres”. In: *Acta Materialia* 59.7 (2011), pp. 2723–2730.
- [339] X. Jehl, Y.-M. Niquet, and M. Sanquer. “Single donor electronics and quantum functionalities with advanced CMOS technology”. In: *Journal of Physics: Condensed Matter* 28.10 (2016), p. 103001.
- [340] K. Weis, S. Wirths, and A. Winden. “Excited states and quantum confinement in room temperature few nanometre scale silicon single electron transistors”. In: *Nanotechnology* 28 (2017), p. 125208.
- [341] D. L. Feldheim and C. D. Keating. “Self-assembly of single electron transistors and related devices”. In: *Chemical Society Reviews* 27.1 (1998), p. 1.
- [342] K. Tapio et al. “Toward Single Electron Nanoelectronics Using Self-Assembled DNA Structure”. In: *Nano Letters* 16.11 (2016), pp. 6780–6786.
- [343] P. U. Vivitasari et al. “Coulomb blockade and Coulomb staircase behavior observed at room temperature”. In: *Materials Research Express* 4.2 (2017), p. 024004.
- [344] K. D. Osborn, M. W. Keller, and R. P. Mirin. “Single-electron transistor spectroscopy of InGaAs self-assembled quantum dots”. In: *Physica E* 21.2-4 (2004), pp. 501–505.
- [345] L. Marty et al. “Self-assembly of carbon-nanotube-based single-electron memories”. In: *Small* 2.1 (2006), pp. 110–115.
- [346] G. Luo et al. “Coupling graphene nanomechanical motion to a single-electron transistor”. In: *Nanoscale* 9 (2017), pp. 5608–5614.
- [347] K. Kalantar-Zadeh and J. Z. Ou. “Biosensors Based on Two-Dimensional MoS<sub>2</sub>”. In: *ACS Sensors* 1.1 (2016), pp. 5–16.
- [348] B. Ryu et al. “Cyclewise Operation of Printed MoS<sub>2</sub> Transistor Biosensors for Rapid Biomolecule Quantification at Femtomolar Levels”. In: *ACS Sensors* 2.2 (2017), pp. 274–281.
- [349] M. Yamamoto et al. “Molecular floating-gate single-electron transistor”. In: *Scientific Reports* 7.1 (2017), p. 1589.
- [350] J. R. Prance et al. “Electronic refrigeration of a two-dimensional electron gas”. In: *Physical Review Letters* 102.14 (2009), pp. 1–4.
- [351] S. Kafanov et al. “Single-Electronic Radio-Frequency Refrigerator”. In: *Physical Review Letters* 103.12 (2009), pp. 1–4.

- 
- [352] F. Sols. “Aspects of quantum cooling in electron and atom systems”. In: *Physica E* 42.3 (2010), pp. 466–471.
- [353] S. Gasparinetti et al. “Probing the local temperature of a two-dimensional electron gas microdomain with a quantum dot: Measurement of electron-phonon interaction”. In: *Physical Review B* 83.20 (2011), pp. 2–5.
- [354] P. Bhadrachalam et al. “Energy-filtered cold electron transport at room temperature.” In: *Nature communications* 5.Sep (2014), p. 4745.
- [355] A. V. Feshchenko, J. V. Koski, and J. P. Pekola. “Experimental realization of a Coulomb blockade refrigerator”. In: *Physical Review B* 90.20 (2014), pp. 1–5.
- [356] V. N. Ermakov et al. “Thermoelectricity in tunneling nanostructures”. In: *Physical Review B* 92.15 (2015), pp. 1–7.
- [357] Y. Zhang, G. Lin, and J. Chen. “Three-terminal quantum-dot refrigerators”. In: *Physical Review E* 91.5 (2015), pp. 1–9.
- [358] G. Rosselló, R. López, and R. Sánchez. “Dynamical Coulomb blockade of thermal transport”. In: *Physical Review B* 95 (2017), pp. 1–9.
- [359] M. F. Gonzalez-Zalba et al. “Probing the limits of gate-based charge sensing”. In: *Nature communications* 6.Jan (2015), p. 6084.