University of Glasgow

A III-V channel field effect transistor for non-classical CMOS: Process optimisation for improved gate stack function

Olesya Ignatova

Project Supervisors

Dr S. Thoms

Prof I. Thayne

Dr D. Macintyre

A thesis submitted for the degree of
Doctor of Philosophy (Ph.D.)

to the
College of Science and Engineering
School of Engineering

December 2015

Abstract

This thesis describes a collection of studies into the electrical response of a III-V MOS stack comprising metal/GaGdO/GaAs layers as a function of fabrication process variables and the findings of those studies. As a result of this work, areas of improvement in the gate process module of a III-V heterostructure MOSFET were identified.

Compared to traditional bulk silicon MOSFET design, one featuring a III-V channel heterostructure with a high-dielectric-constant oxide as the gate insulator provides numerous benefits, for example: the insulator can be made thicker for the same capacitance, the operating voltage can be made lower for the same current output, and improved output characteristics can be achieved without reducing the channel length further. It is known that transistors composed of III-V materials are most susceptible to damage induced by radiation and plasma processing. These devices utilise sub-10 nm gate dielectric films, which are prone to contamination, degradation and damage. Therefore, throughout the course of this work, process damage and contamination issues, as well as various techniques to mitigate or prevent those have been investigated through comparative studies of III-V MOS capacitors and transistors comprising various forms of metal gates, various thicknesses of GaGdO dielectric, and a number of GaAs-based semiconductor layer structures.

Transistors which were fabricated before this work commenced, showed problems with threshold voltage control. Specifically, MOSFETs designed for normally-off ($V_{TH} > 0$) operation exhibited below-zero threshold voltages. With the results obtained during this work, it was possible to gain an understanding of why the transistor threshold voltage shifts as the gate length decreases and of what pulls the threshold voltage downwards preventing normally-off device operation.

Two main culprits for the negative V_{TH} shift were found. The first was radiation damage induced by the gate metal deposition process, which can be prevented by slowing down the deposition rate. The second was the layer of gold

added on top of platinum in the gate metal stack which reduces the effective work function of the whole gate due to its electronegativity properties. Since the device was designed for a platinum-only gate, this could explain the below zero V_{TH}. This could be prevented either by using a platinum-only gate, or by matching the layer structure design and the actual gate metal used for the future devices. Post-metallisation thermal anneal was shown to mitigate both these effects. However, if post-metallisation annealing is used, care should be taken to ensure it is performed before the ohmic contacts are formed as the thermal treatment was shown to degrade the source/drain contacts.

In addition, the programme of studies this thesis describes, also found that if the gate contact is deposited before the source/drain contacts, it causes a shift in threshold voltage towards negative values as the gate length decreases, because the ohmic contact anneal process affects the properties of the underlying material differently depending on whether it is covered with the gate metal or not. In terms of surface contamination; this work found that it causes device-to-device parameter variation, and a plasma clean is therefore essential.

This work also demonstrated that the parasitic capacitances in the system, namely the contact periphery dependent gate-ohmic capacitance, plays a significant role in the total gate capacitance. This is true to such an extent that reducing the distance between the gate and the source/drain ohmic contacts in the device would help with shifting the threshold voltages closely towards the designed values.

The findings made available by the collection of experiments performed for this work have two major applications. Firstly, these findings provide useful data in the study of the possible phenomena taking place inside the metal/GaGdO/GaAs layers and interfaces as the result of chemical processes applied to it. In addition, these findings allow recommendations as to how to best approach fabrication of devices utilising these layers.

Acknowledgements

I wish to acknowledge the following friends and colleagues whose contributions, whether scientific, technical, or advisory, have been of significant importance to the completion of this work:

Dr Stephen Thoms, Prof Iain Thayne, Dr Douglas Macintyre,
Dr Richard Hill, Dr David Moran, Dr Xu Li, Dr Steven Bentley
Dr Gary Paterson, Dr Matthias Passlack,
Dr Vincenzo Pusino, Dr Allan Clark,
Dr Mohsin Haji, Dr Carla Camasta, Dr Wout Jansen,
and technical staff of James Watt Nanofabrication Centre.

I am also grateful to Dr Harold Chong, Prof Scott Roy and Dr Anthony Kelly for useful discussions and such a positive feedback during my viva.

Declaration and Copyright

Unless otherwise acknowledged, the content of this thesis is original and sole work of the author. No portion of this has been submitted by the author in support of any application for any degree or qualification, at this or any other university or institute of learning. The views expressed in this thesis are those of the author, and not necessarily those of the University of Glasgow.

Sections from this work have been published and/or publicly presented in the following:

- D. Macintyre, O. Ignatova, S. Thoms, and I. Thayne, "Resist residues and transistor gate fabrication", Journal of Vacuum Science & Technology B (2009) vol. 27 pp. 2597-2601; also presented at EIPBN, US, 2009.

- O. Ignatova, S. Thoms, W. Jansen, D. Macintyre, R. Hill, and I. Thayne, "Lithography scaling issues associated with III-V MOSFETs", Microelectronic Engineering (2010) vol. 87 pp. 1049-1051; also presented at MNE, Belgium, 2009.

- O. Ignatova, D. Macintyre, S. Thoms, and I. Thayne, "Gate stack processing effects on III-V NMOSFET performance", presented at HETEC, Greece, 2010.

Olesya Ignatova
December 2015

Contents

List of Figures

List of Tables

Abbreviations

2DEG	**T**wo-**D**imensional **E**lectron **G**as
AC	**A**lternating **C**urrent
AFM	**A**tomic **F**orce **M**icroscopy
ALD	**A**tomic **L**ayer **D**eposition
CMOS	**C**omplementary **M**etal **O**xide **S**emiconductor
C-V	**C**apacitance - **V**oltage
CVD	**C**hemical **V**apour **D**eposition
DC	**D**irect **C**urrent
EBL	**E**lectron **B**eam **L**ithography
ESR	**E**lectron **S**pin **R**esonance
EOT	Equivalent **O**xide **T**hickness
FET	**F**ield **E**ffect **T**ransistor
GGO	**G**allium **G**adolinium **O**xide (GaGdO)
HF	**H**igh **F**requency
IETS	**I**nelastic **E**lectron **T**unnelling **S**pectroscopy
IPA	**I**so**p**ropyl **A**lcohol
IR	**I**nfra-**R**ed
I-V	**C**urrent - **V**oltage
LEED	**L**ow **E**nergy **E**lectron **D**iffraction
LF	**L**ow **F**requency
MBE	**M**olecular **B**eam **E**pitaxy
MEIS	**M**edium **E**nergy **I**on **S**cattering **S**pectroscopy
MIBK	**M**ethyl **I**so**b**utyl **K**etone
MOS	**M**etal **O**xide **S**emiconductor
MOSCAP	**M**etal **O**xide **S**emiconductor **Cap**acitor

MOSFET	**M**etal **O**xide **S**emiconductor **F**ield **E**ffect **T**ransistor
NEXAFS	**N**ear **E**dge **X**-ray **A**bsorption **F**ine **S**tructure
PMMA	**P**oly(**m**ethyl**m**eth**a**crylate)
PL	**P**hoto**l**uminescence
REELS	**R**eflection **E**lectron **E**nergy **L**oss **S**pectroscopy
RHEED	**R**eflection **H**igh **E**nergy **E**lectron **D**iffraction
RF	**R**adio **F**requency
RIE	**R**eactive **I**on **E**tching
S/D	**S**ource/**D**rain
SEM	**S**canning **E**lectron **M**icroscopy
UV	**U**ltra-**V**iolet
XPS	**X**-ray **P**hotoelectron **S**pectroscopy

1 Introduction

This thesis presents an investigation into a III-V heterostructure MOSFET (metal-oxide-semiconductor field-effect transistor) with an InGaAs channel and a Ga_2O_3/GaGdO gate dielectric, designed as an n-channel candidate for high-speed low-voltage CMOS (complementary metal-oxide-semiconductor technology) digital logic circuits. The aim of this work was to improve the gate field control by: a) optimisation of fabrication processes associated with the gate stack of the device, and b) investigation into parasitic elements associated with its gate and source/drain contacts layout. The former was achieved through observation of changes in capacitance-voltage measurements of MOS capacitors (MOSCAPs) as the gate processing parameters were varied, with some of the processes tested on devices. To achieve the latter aim, a specialist MOS capacitor structure was designed, modifying the standard MOSCAP towards increased periphery and much smaller dimensions of its contacts and spacing between them.

The investigated transistor design was first reported in literature in 2005, following the development of a low-defect GaAs/Ga_2O_3 interface, implantation-free source/drain regions, and numerous previous reports of improved electron mobility in doped barrier / undoped channel heterojunctions. The popularity of heterojunction transistors for digital logic applications increased with the need for replacement of silicon-channel (Si) SiO_2-dielectric CMOS technology that was used to construct large-scale digital logic integrated circuits populated by fast-switching low-voltage transistors until recently.

The success of the Si MOSFET stems from its scalability, i.e. the proportional reduction of all its lateral and vertical dimensions, which achieves faster, denser and therefore cheaper to manufacture circuits. However, at the turn of the century it

became apparent that if the aggressive dimensional scaling is to be continued, the gate insulator - the thinnest layer of the MOSFET - would have to be 4-5 atoms thick by the end of the next decade. At these thicknesses, the tunnelling of electrons across the insulator are predicted to reach significant levels in terms of gate leakage current, affecting reproducibility and reliability of devices, therefore an alternative to SiO_2 gate dielectric was to be sought.

In parallel to investigating alternative dielectrics, functional scaling methods were explored, aimed at delivering improved system functionalities, such as higher output current and more efficient switching, at fixed physical dimensions. This spurred research into materials with a higher carrier mobility and structures with lower current leakage, naturally leading to III-V heterostructures with undoped quantum well channels and eventually resulting in the development of the device investigated in this work. The InGaAs-channel GaGdO-dielectric heterostructure MOSFET produced a high output current and good subthreshold characteristics at 1 μm gate length and 100 μm gate width, but when scaled tenfold to 100 nm gate at the same width, a shift of threshold voltage towards smaller values and degradation of transconductance occurred. Moreover, the threshold voltage values measured on a number of wafers were generally at least 0.3 V lower than the theoretical value. This suggested the issues were related to the processes involved in the fabrication of the device and the design of the device introducing parasitic elements into the MOS system. It was thus clear that the fabrication process required optimisation, and the effects of the parasitics an investigation. The details of the work carried out are provided below.

In this thesis, the experimental results are presented in three chapters in the following order:

First, large-area MOS capacitor measurements are presented and discussed, leading to identification of the most promising processes. A section presents a method developed to determine the thickness of Electron Beam Lithography (EBL) post-development resist residual layer.

The second chapter presents the results of electrical characterisation of MOSFETs fabricated using processes based on MOS capacitor measurement outcomes.

Finally, a specialised MOSCAP layout will be presented, combining large area capacitance gate stacks with scaled test structures mimicking the range of dimensions used for the actual devices. A section is dedicated to a description of challenges that were overcome in order to fabricate the structures.

Supporting theory sections are contained within the experimental chapters focusing on non-idealities of the MOS stack, as well as MOS transistor characteristics and key parameters. These are explained via carrier dynamics and equivalent circuit diagrams. The background section expands on the introduction providing more detail about the evolution of the heterostructure MOSFET device. It also complements the theory sections summarising the most common methods to measure defects in the MOS structure, followed by a review of the effects of common fabrication processes on the amount of defects for various semiconductor-oxide structures. For this, semiconductor surface passivation methods are presented, as well as techniques to minimise charge trapping at the oxide-semiconductor interface and/or unwanted charge in the oxide.

The key findings can be summarised as follows:

1) Gate region processing and thermal, mechanical, and radiation-induced damage

- E-beam exposure during gate lithography introduces interface states
- E-beam evaporation of metals introduces interface states
- Oxygen plasma barrel ashing is non-damaging
- Polymer contamination at the gate-oxide interface decreases gate capacitance and degrades device output characteristics
- Post-metallisation anneal at 430°C causes damage at the interface, but removes fixed charge in the dielectric
- Ultrasonic agitation during substrate cleaning introduces interface states

2) Gate region process damage prevention / mitigation

- Slowing down the deposition rate of metals
- Using a single-layer gate metal
- Post-metallisation anneal in N_2 or $O_2 \leq 350°C$, for gate-first fabrication method only

3) Parasitic components associated with the gate region, and other device fabrication elements that could affect scaling

- Periphery capacitance element of the total gate capacitance is dependent on the separation between the gate and the source contact
- Uneven lateral distribution of density of oxygen vacancies in GaGdO
- Lateral diffusion of Ni/Ge/Au metal during 430°C ohmic contact anneal

Expanding on the key findings, this work showed that the effect of radiation damage during fabrication is dependent on which fabrication process is the source of damage. Electron-beam exposure was found to induce a change in the amount of charge traps present at the interface, while oxygen ashing did not cause observable changes in the gate oxide, proving instead to be a necessary step to avoid the detrimental effect of PMMA resist residues. These beneficial effects first observed from analysis of bulk-doped surface-inversion MOSCAPs were confirmed for buried channel quantum-well heterostructure MOSFETs. With regard to metallisation techniques, it was found that electron beam metal evaporation induces damage to GaGdO and InGaAs layers, with the amount of damage directly proportional to the number of metal layers composing the stack. Depositing metals at a slower rate resulted in damage mitigation. It was also found that the thickness and the type of the top layers affect the electrical response of the entire stack, most likely due to the differences in electronegativity of the metals chosen.

From the experiments on thermal treatment of the material stack, a trade-off between the benefits of a thermal step and thermally induced damage was observed, with the best results obtained when the temperature did not go beyond 350°C. Regarding the atmosphere in which the thermal step was carried out, GGO (short for GaGdO) and InGaAs were annealed in nitrogen and oxygen at 350°C and no difference was observed. This is most likely due to the gas molecules which were not provided enough thermal energy to decompose into their elemental components and drive the annealing of defects. The inclusion of a post-gate metallisation thermal step in oxygen atmosphere at 350°C on buried channel quantum-well heterostructure MOSFETs, which were also subject to an annealing step at 430°C in nitrogen atmosphere, was beneficial to device performance as long as the devices were "gate-first", i.e. the nitrogen anneal took place after the oxygen one.

Finally, the results from the specially developed scaled gate capacitor structures compared periphery and area effects on the total capacitance, and provided useful insight into parasitics in a metal gate GGO dielectric MOS structure. First, the experiments showed a varying spatial density of oxygen vacancies in the GGO which would affect performance uniformity of scaled devices. Moreover, scaling of the capacitors was severely affected by lateral diffusion of the ohmic metal after the annealing step, which contributes as another limiting factor to device scaling. The scaled gate capacitors also showed an increase in periphery effect when reducing the spacing between the gate and the source contact.

The results presented in this thesis will be of use to the MOS research community to offer a better understanding of the way fabrication processes influence the charge behaviour in metal-gate / high-κ dielectric / III-V semiconductor structures, and as a tool towards a successful manufacturing of high switching-speed, low-power transistors made of these materials in the future. The optimisation of the gate function for scaled MOSFETs on III-V substrates, alongside development of low-resistance source/drain ohmic contacts, could pave the way for the transfer of the technology on a silicon platform, that could provide improved functionality to a CMOS transistor without affecting the manufacturing costs.

2 Background

2.1 Non-classical MOSFET

As stated in the introduction, the CMOS transistor technology based on the Si/SiO_2 interface, which has seen unprecedented progress since the beginning of its use [1], requires an alternative. The improvements in speed, power and density of CMOS circuits enabled by dimensional [2] and functional scaling now require a complete transformation of the MOSFET in terms of gate dielectric material, channel material and device structure. The new dielectric material needs to be physically thicker whilst having the capacitance of SiO_2 [3], the carriers in the channel need to have a higher than Si effective mobility, and most importantly, the new semiconductor-oxide interface would need to match the quality of Si/SiO_2. The first few sections of this chapter show how this can be achieved, starting from a comparison of the conventional and the improved device design.

2.1.1 Device design

In cross-section, the planar MOSFET has all three terminals on the top surface of the semiconductor with the gate metal separated by a layer of dielectric, as illustrated in figure 2.1. Under bias, the two components of the total electric field in the device are the vertical gate field that concentrates charge in the channel and the horizontal gate-source field that moves that charge through the channel. All this can be simply expressed through the movement of the semiconductor Fermi level in the channel region, as shown in figure 2.1. If the Fermi level is viewed as an indicator of the conductivity of the material, then when it is above the conduction band edge, that region of the semiconductor is populated with charge. For specifications of design

and operation of the device used in this work, refer to the theory section of Chapter 4.

Traditionally, the classical design for the normally-off n-channel CMOS FET was a p-doped bulk silicon wafer having source/drain regions implanted with a high concentration of electrons. To turn on the device, a positive gate voltage is applied which repels positive charge away from the surface forming a depletion region in the semiconductor. When the gate voltage is high enough, the gate field pulls free electrons from the n^+ implanted source and drain regions until all of the surface is populated with electrons and a channel between the source and the drain is formed. At this point, the band-bending at the surface is so significant that the Fermi level overlaps the conduction band. This device is dubbed *inversion-channel* MOSFET since an n-type channel is formed in a p-type material.

As opposed to the bulk Si device, in the non-classical III-V solution, the semiconductor part is a heterostructure, composed of layers of varying bandgap width in such a way that the channel is formed in a quantum well. The channel region is initially undoped, but populated with carriers by diffusion of electrons from the doped layers surrounding it. Here only one type of charge is involved in the formation of the channel. The movement of charge is reversed compared to the classical device with the channel region initially depleted when the device is off. With positive gate bias applied, the gate field pulls the charge towards the surface, until the maximum concentration is reached. Since no carriers are taken from source and drain regions for channel formation, there is no need for source/drain implantation, and the only requirement for the source and drain regions is to have a low-resistance ohmic contact to the channel, which can be achieved by thermally-induced metal diffusion into the top semiconductor layers. Figure 2.1 lists the key differences in the design of the two types of devices and shows their cross-section in ON and OFF states with corresponding conduction band and Fermi energy levels.

2.1.2 III-V heterojunction channel

As briefly mentioned in Introduction, one method to functionally scale CMOS was to replace the Si channel with a material in which the mobility of carriers is higher. This would enable device operation at lower voltages whilst delivering a higher output current omitting the decrease in channel dimensions. This is expressed in equation 2.1 that shows the output drain current at saturation. It is a known fact that group III-V compound arsenides such as GaAs, InGaAs or InAs have considerably higher electron mobilities, especially when undoped [4]. One way to populate an undoped channel region with electrons was by transfer of electrons from a doped semiconductor layer with a higher bandgap to an adjacent undoped layer with a lower bandgap by diffusion. Using this technique, extra high electron mobility was first demonstrated in an GaAs/n-AlGaAs heterojunction metal-semiconductor FET in 1980s [5, 6]. A two-dimensional layer of electrons (known as two-dimensional electron gas *2DEG*) that formed in GaAs at its junction with the AlGaAs had Hall mobility three times higher compared to a bulk-doped GaAs device [7].

$$I_{D,SAT} = \frac{W}{L} \mu C_G \frac{(V_G - V_{TH})^2}{2}, \quad for\ V_G > V_{TH} \tag{2.1}$$

- $I_{D,SAT}$ is the output current density when the device in saturation
- W and L are the dimensions of the channel
- μ is the carrier mobility of the channel
- C_G is the capacitance density associated with the gate dielectric
- V_G is the voltage across the gate
- V_{TH} is the on/off threshold, the gate voltage at which the channel current is 1 μA/mm

(a)

Classical design	Non-classical design
• bulk doped • surface inversion • carriers everywhere • doped semiconductor S/D regions • low-κ dielectric • channel formed on the surface of bulk *doped* material • channel carrier concentration is defined by S/D doping	• selectively doped • no inversion • carriers confined in the channel • metal S/D regions • high-κ dielectric • channel formed in a buried *undoped* quantum well • channel carrier concentration is defined by heterojunction doping

(b)

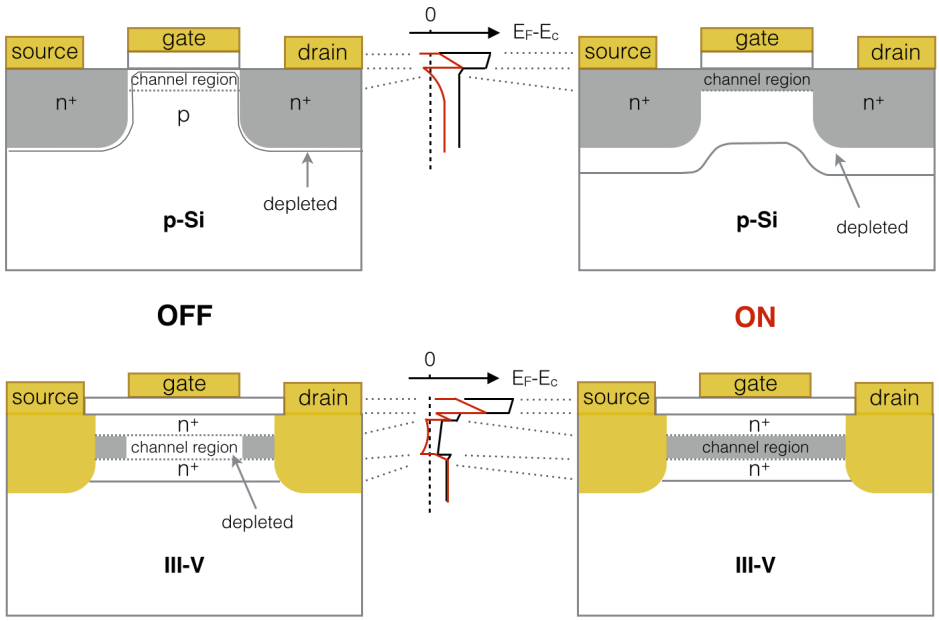


Figure 2.1 The key features (a) and simplified physical structure (b) of two most contrasting types of a normally-off n-channel MOSFET for low-voltage CMOS logic applications. Device cross-sections in on and off states with corresponding conduction-band and Fermi-energy levels are shown for the conventional design with a channel formed in bulk Si (top) and the design studied in this work with a quantum-well InGaAs channel (bottom).

The single heterojunction 2DEG channel evolved into a quantum well channel when the two heterojunction layers were mirrored with respect to the smaller band-gap layer, creating a three-layer structure. With this configuration, better carrier confinement in the channel region was achieved, and with the confinement, the subthreshold current and output conductance were reduced. The material stack was optimised to have a larger conduction band discontinuity to decrease gate leakage current. This resulted in an n-AlGaAs/InGaAs/n-GaAs device with a typical InGaAs channel layer thickness of 10 nm [8].

After this three-layer structure was embedded into a MOSFET by adding a gate dielectric layer between the semiconductor and the gate, the heterostructure underwent further optimisations. One of the most successful designs had AlGaAs/GaAs/InGaAs/GaAs/AlGaAs channel confinement layers [9] and a GaGdO-based dielectric, achieving a high output current and a low subthreshold leakage [10, 11]. It is this device that this work is based on, and the structure details can be found in chapter 4.

2.1.3 High-κ oxide / III-V substrate interface region

As expressed in equation 2.2, the simple solution for the alternative dielectric to have a larger physical thickness whilst maintaining SiO_2 capacitance for the same gate geometry, is a larger relative dielectric constant κ. The alternative high-κ candidates are all metal oxides, falling in the following categories: a) rare earth oxides, such as yttrium oxide Y_2O_3, lanthanum oxide La_2O_3, and gadolinium oxide Gd_2O_3, b) transition metal oxides, such as HfO_2, ZrO_2, and TiO_2, c) post-transition metal oxides, such as Ga_2O_3 and aluminium oxide Al_2O_3, or d) mixed layer oxides like gallium $Ga_2O_3(Gd_2O_3)$. The choice is defined by which layer can provide the best chemical stability at its interface with the III-V semiconductor surface, the best barrier to thermal diffusion of III-V atoms into the dielectric layer, and a wide bandgap for low gate current - all whilst exceeding $\kappa = 3.9$ of SiO_2 by at least a factor of two [12]. The physical properties of the high-κ oxides are listed in table 2.1.

$$C_{ox} = \frac{\kappa \varepsilon_0 A}{t_{ox}} \tag{2.2}$$

- A is the gate area
- κ is the relative dielectric constant of the insulator
- ε_0 is the permittivity of free space
- t_{ox} is the gate insulator thickness

The formation of high-κ dielectric layers on III-V layers can be either implemented *in-situ* or *ex-situ*. In-situ deposition of the dielectric layer takes place in the same deposition system as the III-V substrate, where vacuum is maintained at all times. Ex-situ deposition, on the contrary, is when the semiconductor and the dielectric layers are formed by different methods, and the semiconductor surface is exposed to air while the sample is transferred between the chambers. Albeit surface oxidation problems, the most commonly used methods are ex-situ, as more flexibility is allowed in terms of materials, requiring less complex equipment. In-situ techniques, however, have the advantage of no contamination-related or oxidation-related defects in the semiconductor-dielectric interface.

	SiO_2	Al_2O_3	HfO_2	ZrO_2	La_2O_3
Dielectric constant	3.9	8-9	18-25	18-30	20-36
Bandgap (eV)	9	6.2-8.8	5.6-6	5.8	4.3
Data source	Ref [13]	Ref [13]	Ref [13]	Ref [13]	Ref [13]

	Y_2O_3	Ga_2O_3	Gd_2O_3	$Ga_2O_3(Gd_2O_3)$	$(Gd_xGa_{1-x})_2O_3$
Dielectric constant	12-20	10-11.4	11.4	14.2	20
Bandgap (eV)	5.5	3.2-4.8	5.3-5.9	4.7	5.8
Data source	Ref [14],[15]	Ref [16],[17]	Ref [18]	Ref [19]	Ref [17]

Table 2.1 High-κ oxides and their dielectric properties.

Various methods are used to deposit III-V epitaxial layers and high-κ oxides, all involving a chemical reaction between atomic elements deposited on a given

surface. The main three types are chemical vapour deposition CVD, molecular beam epitaxy MBE, and atomic layer deposition ALD. However, deposition using e-beam evaporation has also been reported. The most widely studied III-V/high-κ interfaces and the corresponding deposition methods are listed in table 2.2.

The in-situ GaAs/GGO structure listed last in table 2.1 and table 2.2 should be described in more detail, being the one used in this work. It was discovered that when Ga_2O_3 dielectric was deposited on an MBE-grown GaAs surface in-situ, a low-defect Ga_2O-GaAs interface formed in the beginning of the Ga_2O_3 deposition process. After a monolayer of Ga_2O chemisorbed on the GaAs surface, growth proceeded via formation of an amorphous Ga_2O_3 film. This dielectric was demonstrated in a MOSFET with 9 nm Ga_2O_3 [25]. However, with Ga_2O_3 only, gate current leakage activation energies were found to be too low due to a negative conduction band offset between Ga_2O_3 and GaAs [17]. This was unsuitable for low power device applications, requiring an additional large bandgap oxide layer [26]. Amorphous GaGdO was a good candidate for this, as it could be deposited in-situ in a dual-chamber MBE system. This led to the bulk Ga_2O_3 layer replacement with a multi-layer stack comprising 1 nm interfacial Ga_2O_3 layer and a bulk $(Gd_xGa_{1-x})_2O_3$ layer [17]. The details of the development process of the high-κ $Gd_xGa_{0.4-x}O_{0.6}$/ Ga_2O_3 dielectric stack on GaAs can be found in [27] and [24]. It was found that for a better oxide-GaAs interface and reduced leakage current density, the concentration of Gd must be below 20% as less Gd moves towards the oxide interface at lower Gd percentage, with the best results achieved with an amorphous $Gd_{0.15}Ga_{0.25}O_{0.6}$ film.

III-V surface	(In)GaAs (MBE)	GaAs (MBE)	InGaAs (MBE)	GaAs (MBE)	GaAs (MBE)	GaAs (MBE)
Dielectric layer	Al_2O_3 (ALD)	HfO_2 (ALD)	ZrO_2 (ALD)	Ga_2O_3 (MBE)	$Ga_2O_3(Gd_2O_3)$ (e-beam evap)	Ga_2O_3/GaGdO (MBE)
In-situ / Ex-situ	Ex-situ	Ex-situ	Ex-situ	In-situ	In-situ	In-situ
Data source	Ref [20],[21]	Ref [22]	Ref [23]	Ref [17]	Ref [19]	Ref [17],[24]

Table 2.2 A summary of the most common III-V arsenide / high-κ oxide interfaces and their deposition methods. Dielectric thicknesses vary between 1 and 10 nm.

2.2 MOS non-idealities

In a perfect MOS transistor, the free charges in the semiconductor channel immediately and fully respond to the field created across the dielectric when bias is applied to the gate. Such a device switches on and off with minimum delay and outputs the maximum available current at saturation. The oxide layer in the ideal device is electrically and chemically balanced, i.e. it does not have any unsaturated bonds or displaced atoms, and therefore it does not have any charge that could respond to the gate bias or interfere with the carriers in the channel. In reality, both the surface and the dielectric layer are non-ideal, simply because they are finite layers terminated abruptly and created by imperfect processes, all resulting in creation of various types of defects.

In general, replacing a Si/SiO_2 semiconductor-oxide interface with a III-V/high-κ interface is very challenging. The Si/SiO_2 interface is formed between a stoichiometric semiconductor surface and its native oxide with a low concentrations of defects, making it easy to modulate the semiconductor surface potential over an appropriate range and move the charge carriers near to the interface with mobilities unaffected. The only factor that can have an effect is the surface orientation. The III-V compound semiconductor surfaces are susceptible to non-stoichiometry and the lattice defects associated with it, and arising from it, the possibility of the native oxide to be a mixed oxide of uncertain composition [28], giving rise to conductive paths and leakage of currents through them. In addition to the non-stoichiometry, there is also surface orientation and surface termination factor.

2.2.1 Classification of defects

From a chemical perspective, every defect, be it a vacancy, an interstitial atom, a dislocation, or line and plane defects in the crystal structure, can be seen as something that induces a change in the material electron balance; in undoped material, a defect would cause ionisation of atoms, deficit or excessive of electrons. Therefore, defects are classified according to the behaviour of charges surrounding

them under the influence of a field. The two main types of defects are either the ones that are rechargeable and temporarily capture electrons, known as *trap states* or *charge traps*, or the ones that are non-rechargeable, the charging state of which is fixed at either positive or negative, known as *fixed charge*. Whether the defect is a fixed anomalous charge or a trap, its incidence is independent of the location in the oxide, and either type can occur at any point across the MOS stack. For example, in some materials, in the area between 0 and 3 nm from the oxide-semiconductor interface into the oxide, there are interface traps, anomalous fixed positive charge and rechargeable electronically-active defects [29].

However, whether the defect is of fixed or trapping nature, it can be further differentiated according to their physical proximity to the semiconductor surface. The main types of MOS defects are schematically shown in figure 2.2. At the few atomic layers of the oxide-semiconductor interface, there are *interface defects*, caused by dangling bonds and lattice imperfections. Of interface defects, there are ones that trap semiconductor charge, known as *interface traps* and *interface states*, and there are fixed charge defects, known as *interface fixed charges*. From the semiconductor surface upwards, there are *border traps* 0 to 3 nm into the oxide and *bulk oxide traps* further deep in the oxide. These defects are broken bonds in the oxide with maximum density near the interface. In the bulk of the oxide there are also *bulk fixed charges*, which are electrons trapped at defect sites.

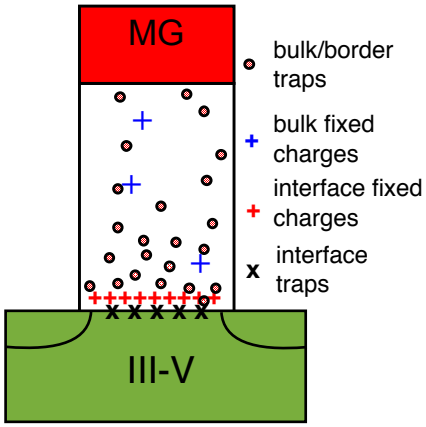


Figure 2.2 Types of defects in the dielectric region of a MOS structure that electrically respond to bias.

Semiconductor bandgap states arising from the interfacial defects act as charge traps and have the most adverse effect on the channel conduction in terms of maximum current and switching action as they effectively 'pin' the semiconductor Fermi level in a fixed position. Fixed charges at the interface and in the bulk of the oxide reduce the field strength and cause reliability problems. Each stage in the fabrication of the MOS structure will either alter the existing defects or create new ones, and the effect of a process on the concentration of defects can be assessed both chemically and electrically.

2.2.2 Measuring defects

This section contains an overview of methods that have been reported in literature to provide electrical and chemical characterisation of semiconductor and dielectric layers before and after a fabrication process. Individually or combined, they provide a measure of process-induced surface and sub-surface disorder and its associated electrically active defect levels in the bandgap. Aside from device fabrication process development, these methods are useful in development of material growth processes for identification of the least detrimental process to the quality of the MOS stack. The techniques can be divided into two major groups: *electrical* for trap state density assessment and *chemical* for structural defect assessment. Electrical characterisation is performed by physically probing the device under test, which requires metal contacts to the surface. It allows measurement of electrically-active defects in response to electrical stimulus. Chemical methods such as spectroscopy and microscopy, on the contrary, benefit from contactless probing, and are capable of providing spatial information about the distribution of structural defects in the material. Combining the two types of techniques allow to determine the percentage of defects participating in carrier removal.

2.2.2.1 Spectroscopy

Spectroscopy is based on illuminating the sample surface with electromagnetic radiation at a range of frequencies and observing the interaction with the

underlying matter. The measured signal is the transmittance or reflectance spectrum of absorbed energy. Observing changes in the height and position of peaks in the absorbed energy spectrum can provide useful information about: a) presence of contaminants on the semiconductor surface, b) presence of ionised atoms in the dielectric due to broken bonds, c) location of defects in the dielectric film, and d) qualitative information about the concentration of defects.

In the order of decreasing frequency (increasing wavelength), radiation sources are X-ray, ultra-violet (UV), visible, infra-red (IR), and microwaves. Lower energy radiation (microwave, long-IR) can only go as far as cause molecular vibrations in the material. These vibrations are picked up as peaks in the measured spectrum, each corresponding to a particular frequency component, which in turn corresponds to a particular chemical species. This makes low energy spectroscopy useful in detecting surface contamination or presence of native oxides on surfaces. High energy radiation (near-IR, UV, X-ray) is used where assessment of bonding between elements is required, both surface and sub-surface. It causes excitation of electrons in inner shells of atoms and is used to analyse semiconductor-oxide interfaces.

Spectroscopy methods can be classified as: high or low energy, optical or electron-beam. For low-energy techniques, electrons induce stronger molecular vibrations than photons, making electron-beam based spectroscopy more sensitive than optical spectroscopy. For assessment of defects in III-V / high-κ stacks, low-energy optical spectroscopy methods are Photoluminescence (PL), Raman, and Infrared (IR) spectroscopy; low-energy electron-beam (20-200 eV) spectroscopy methods are reflection electron energy loss (REELS), low energy electron diffraction (LEED), inelastic electron tunnelling spectroscopy (IETS), electron spin resonance (ESR), medium energy ion spectroscopy (MEIS); high-energy optical methods are X-ray photoelectron spectroscopy (XPS), near-edge X-ray absorption fine structure (NEXAFS); and high-energy e-beam methods are reflection high energy electron diffraction (RHEED).

Dielectric film

Fabrication processes can have a detrimental effect on the properties of a dielectric layer in terms of reducing its bandgap, its crystallinity and concentration of unpaired electrons as a function of depth. The spectroscopic characterisation techniques that have been used to measure these are REELS, XPS, NEXAFS, IETS, and ESR.

Using ESR, Revesz *et al.* assessed the dielectric layer for presence for defects in the form of unstable chemical species. Peaks in the spectrum indicated presence of molecules or atoms with unpaired electrons. No ESR signal (no absorption peaks) was observed if the system was stable and only contains paired electrons [30]. Concentration and energy levels of electrically-active defects in the oxide as a function of depth were measured using ESR by Nishi *et al.* [31] and IETS by He *et al.* [32]. Using REELS, it was possible to determine the effective bandgap energy of a dielectric post-process. A process causing a reduction in the bandgap is undesirable as it would cause an increase in gate leakage current. Kim *et al.* combined XPS with NEXAFS to assess crystallisation of dielectric layers [22].

III-V surface

III-V surface properties such as termination, contamination, and formation of native oxides, are essential to its electrical behaviour and its sensitivity to processing. Using XPS, Hackley *et al.* detected As-O bonds on GaAs surface (associated with interface states), and then treated GaAs surface to HF (hydrofluoric acid) and NH_4OH (ammonia) solution to de-oxidise and passivate (reduce reactivity) the surface [33]. Structural properties of GaAs surfaces like termination atoms and lattice disorder, as well as surface contamination, have been measured by Laukkanen *et al.*, who observed LEED diffraction patterns, capable of detecting as little as a small fraction of a monolayer of contamination [34]. Rahbi *et al.* characterised the effects of hydrogen passivation on GaAs layers using IR spectroscopy [35].

Interface

Dalapati *et al.* made an assessment of the interface quality by measuring bonded semiconductor and oxygen molecules at the interface using XPS [36], whilst Nishi *et al.* characterised defects according to their types by observing peaks during ESR spectroscopy [31]. Defect passivation at interfaces has been validated with optical measurements like IR by Sweeney *et al.* [37] and also with PL spectroscopy. In GGO dielectric stack development, PL spectroscopy showed that the phenomenon responsible for such an effective GaAs surface defects passivation with Ga_2O_3 interfacial layer was formation of Ga_2O between the (001)-GaAs and the interfacial layer [24].

Sub-surface

For sub-surface analysis, IR spectroscopy has been used for a number of applications. Songprakob *et al.* used IR to study carbon as an alternative p-dopant for GaAs [38], Hellman *et al.* used it to optimise MBE growth temperature of GaAs layers by monitoring bandgap shifts [39], and defects in MBE-grown GaAs lattices were investigated through this technique as well by Eickhoff *et al.* [40] and Gledhill *et al.* [41]. Depth profiling of multi-layer III-V stacks was the most effective with higher energy methods such as XPS and NEXAFS. However, the lower-energy ESR technique was used to determine how far surface and interfacial defects extend into the top semiconductor layers. Kang *et al.* concluded that for Si/high-κ stack, it was within one nanometer. Finally, sub-surface characterisation was also achieved by MEIS: by using this technique, Kim *et al.* [22] determined relative concentration of an element as a function of distance from the surface for film elemental composition data.

2.2.2.2 Electrical Characterisation

As opposed to chemical methods that can tell specific information about defects in a given material system, electrical methods determine concentration of electrically active defects, non-specific to the nature of the defects or the type of atoms. Electrical characterisation enables assessment of the effect of a process on the system through: a) densities of *interface-trapped charge*, *border-trapped charge*, and *oxide-trapped charge, b)* densities of *fixed interface charge* and *fixed bulk oxide charge*.

The most important traps for electrical characterisation are those whose energy levels happen to fall within the energy bandgap, hindering switching operation. Charge trapping in these interface states causes diminished surface potential control, removes carriers from the channel and reduces carrier mobility. The energy levels of interface traps are distributed across the semiconductor bandgap and the amount of charge that will get trapped will depend on the semiconductor Fermi level position at the oxide-semiconductor interface. Depending on the type of defect from which the trap originates, it will either capture an electron (an acceptor-like state) or release an electron / capture a hole (a donor-like state). If mostly acceptor-like states are present in the bandgap, it will result in a net negative charge. In the same way, a net positive charge at the interface will mean that most bandgap states are donor-like. The presence of both net negative and positive charge can be detected by electrical methods.

These measure the distribution of surface electric charge by surface impedance measurements that require probing a MOS capacitor to measure density of trapped charges in the oxide; contrary to contactless physical-chemical methods, they cannot be used to characterise semiconductor surface on his own. Electrical and physical-chemical methods are both necessary for a complete assessment of a semiconductor device material system: while the latter methods help identify presence of unwanted atoms or other defect types (e.g. a broken bond, a vacant or an interstitial atom) and assess if a particular defect in the lattice will act as a trapping centre, electrical methods will show the net contribution to charge carrier trapping of all

the defects in the structure. The use of both helps with understanding how and why a particular fabrication process would affect channel conduction.

The main electrical methods used for characterisation of electronic devices material systems can be divided into categories depending on which quantities are monitored: the most common methods are based on measurements of changes in capacitance as a function of voltage applied. Capacitance is the ability of a body to store charge, and its measurement in a MOS system can be related to the presence and density of charge trapping centres. Changes in capacitance as a function of frequency can be used to determine the values of density of interface states D_{it} for energies within the band gap. Conductance methods work in a similar way to capacitance measurements but the measured quantity changes. Finally, current vs. voltage measurements are normally used for leakage current and dielectric breakdown voltage evaluation in MOS systems to study the quality of gate dielectrics. Current-voltage measurements are performed either in MOS capacitor systems or on MOS transistors. The following will report on MOS capacitor and MOS transistor I-V methods, C-V methods and conductance methods. MOS capacitor measurements are the ones that have been used most extensively in this work therefore particular attention will be given to C-V measurements. Here only a general overview of different electrical methods is provided, and a more detailed analysis is presented in the theory section of the experiments chapters.

MOS capacitor I-V

Current-voltage methods are based simply on measuring the current flowing through the MOS stack while sweeping an applied DC voltage from negative to positive values. The measured device can either be a MOS capacitor or a complete MOS transistor, but the focus here will be on MOS capacitors. The main quantity of interest for I-V measurements of MOS capacitors is the leakage current: a large value can be attributed to structural defects or trap centres in the high-κ dielectric. The slope of the I-V curve at low voltages can be used to extract the specific resistivity of the oxide film; typically values greater than 10^{14} Ω.cm are required for acceptable gate leakage currents: ~10^{-3} A/cm^2 for low-power transistor applications

[42, 12]. The I-V characteristic of a MOS capacitor provides information also on the insulating properties of the dielectric film. These can be deducted from the gate leakage current, due to charges moving by tunnelling, and from the breakdown voltage, for which the system switches from an insulating to a conducting state. The breakdown voltage in a MOS stack is generally defined as the value of gate bias for which the current flowing through 1x1 cm^2 area of dielectric exceeds a set value [36].

MOS transistor I-V

I-V characterisation of a complete MOSFET allows for the extraction of more information with respect to a MOS capacitor structure. Full MOSFET I-V characterisation normally yields two sets of curves, measuring the drain source current I_{ds} either as a function of drain-source voltage V_{ds} for a range of gate biases, or as a function of gate-source voltage V_{gs} for operation in saturation and in the linear region (high and low V_{ds}, respectively). The threshold voltage extracted from the I_{ds}-V_{gs} curve can be used as an indicator of the effectiveness of the gate stack in controlling the flow of carriers in the channel, thus assessing MOS stack function. If appropriate processing and fabrication methods have been used, the value of threshold voltage should not deviate from its theoretical value. If the device is designed for enhancement mode operation, i.e. theoretical positive threshold voltage, but the measured value is lower or negative, it can be concluded that process-induced defects in the MOS stack have cause Fermi-level pinning and caused degradation in the gate control function. Other quantities associated with good oxide-semiconductor interface are high saturation current and high transconductance.

Frequency-dependent C-V

The reason for performing C-V sweeps at different frequencies is to detect the entirety of traps present in the capacitors. Traps can be classified by their trapping response time: those that respond to an applied electric field only up to an

oscillating frequency of 1 kHz are called slow traps, while those that are able to trap charges even for higher frequencies are known as fast traps. For high frequency (e.g. 100 kHz), a typical C-V curve for n-doped semiconductor is shown in figure 2.3(a), while figure 2.3(b) displays the traps energies within bandgap covered by a typical multi-frequency sweep. It can be seen that the curve tends to saturate at the minimum and maximum values for high negative DC voltages or high positive DC voltages applied, respectively. The minimum capacitance value is called depletion capacitance (C_{min} or C_d), and it is related to the maximum extension of the depletion region in the capacitor. On the other hand, the maximum capacitance value is known as accumulation capacitance (C_{acc}) and, with a sufficiently large DC bias, it is the oxide capacitance. A difference in accumulation capacitance C_{acc} is observed for different frequencies, commonly called *frequency dispersion*, if there are trap states whose energy levels fall in the semiconductor bandgap. An increased accumulation capacitance value may suggest a presence of a transitional dielectric layer with low resistivity adjacent to the interface [36, 43].

With regard to the transition region, the change in capacitance with applied voltage, i.e. the slope of curve, is related to the presence of defects at the oxide-semiconductor interface. This is explained in more detail at the start of chapter 3. The slope of the curve, often called *stretch-out*, can therefore be used as an indicator of how effectively the gate stack can control the channel charge, and has been used extensively in this work for characterisation of different gate stacks. It can be seen from figure 2.3 that for low frequency C-V measurements (< 1KHz) an n-doped semiconductor will produce a curve that saturates at the oxide capacitance for large DC voltages, while displaying a relative minimum near zero bias. The measured capacitance does saturate to the minimum theoretical value for large applied negative voltages only when measuring at high frequency, since then only the majority carriers response can be detected. Lack of dispersion in minimum capacitance for frequencies > 1 kHz is typical of a structure where interface traps do not respond to high frequency signal. At low AC frequency instead the signal variation is slow enough for minority carriers at the semiconductor surface to follow the variation in voltage, and their contribution can be observed in the C-V curve.

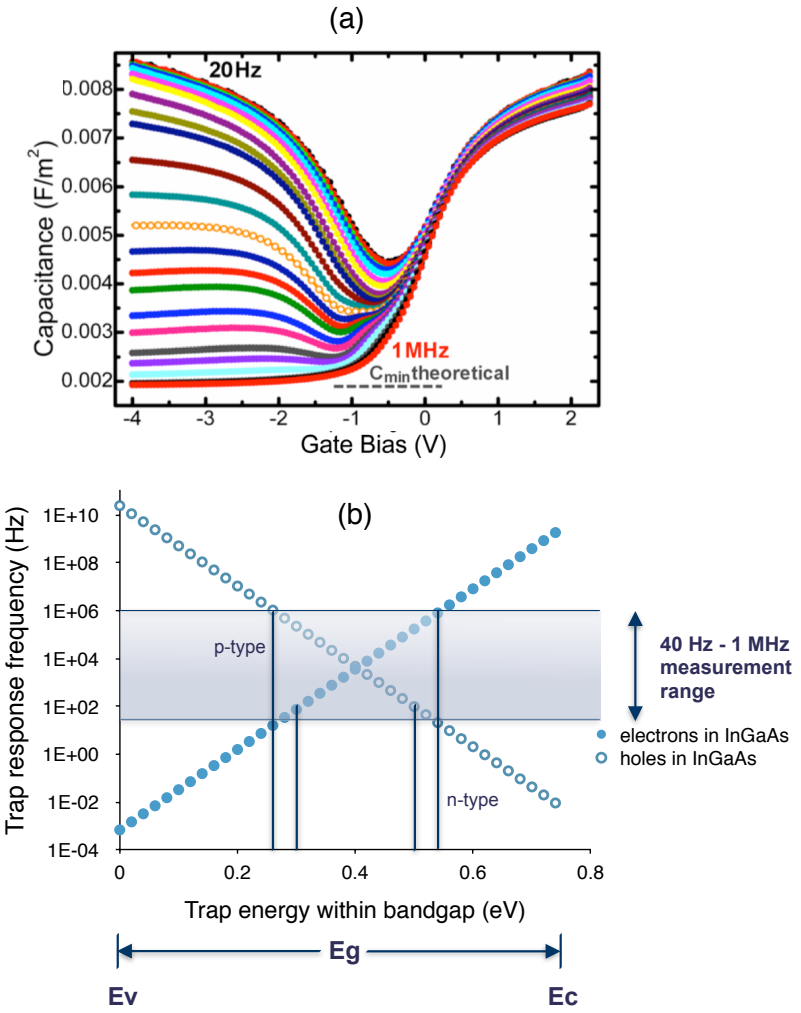


Figure 2.3 (a) Typical multi-frequency C-V response of an n-MOS capacitor [44]. (b) Coverage of bandgap energy levels probed with a multi-frequency C-V measurement.

Multi-frequency C-V sweeps can also be used to assess the presence and characteristics of trapping centres in the oxide, and measuring the trapping time of interface defects, which can range from ns to several seconds, can indicate the location of defects in the oxide [43]. Multi-frequency C-V sweeps were also used to assess the position of interface states within the bandgap, since traps response time is related to their energy level. A comparison of capacitance values in depletion between quasi-static (described shortly) and high frequency for n-MOS and p-MOS stacks yielded the density of interface traps near the conduction (n-MOS) or valence (p-MOS) band, from which the mid-gap value of interface traps density could be extracted by interpolation [43].

C-V measurements have reportedly been used to detect trapped charges densities between 10^{10} and 10^{13} cm^{-2}. The acceptable number of electrically active defects for effective operation of a III-V high-κ MOS device is approximately 10^{11} traps per cm^2. In order to detect this concentration of traps, large-area probing is necessary to have sufficient sensitivity and average out local variations, thus a 250 μm diameter top metal contact is commonly used.

Combination of high and low frequency C-V method

This is a particular case of multi-frequency C-V. All techniques reviewed so far take place at room temperature. A combination of high and low frequency C-V curves recorded for various temperatures was reported to attain an accurate estimate of the density of interface traps. This method is based on measuring multi-frequency C-V curves for substrate temperatures ranging between room temperature and -50°C. This method is based on the fact that accessing different energy levels within the bandgap could either be achieved by sweeping the frequencies (because of the different response time of the traps which depends on their energy level) or by changing the temperature. The density of interface states can then be obtained from an analytical formula. This technique is commonly named the "hi-low" C-V method, and one of its limitations has been shown to be a relevant underestimation of D_{it} for samples with large time constant dispersion [45].

Fixed frequency C-V

Quasi-static C-V

Quasi-static C-V measurements lack the AC component of the applied voltage. The measured current from which the capacitance is calculated is the induced displacement current that derives from applying a linear voltage ramp across the capacitor. Typical sweep rates are in the range from 5 to 500 mV per second, depending on oxide capacitance and minority carrier lifetime. This corresponds to an equivalent frequency approximately 40 times smaller than the lowest AC

frequency attainable with standard laboratory equipment (1 Hz). This method is therefore sensitive enough to assess whether accumulation or inversion are achieved, which was not possible with standard low frequency C-V approach, since as previously stated minority carrier response can only be detected for very slow voltage variations. The quasi-static technique can also produce an estimate of surface potential and surface state density over a large part of the energy gap, while also testing the presence of nonuniformities in MOS structures. If the C-V curves obtained with the quasi-static method display a difference between accumulation and inversion capacitances, it is an indication of a low concentration of interface states near the conduction or valence band edge. If formation of an electron inversion layer is observed in p-type samples, the density of interface states near the conduction band edge is low. Quasi-static C-V analysis has been used by Passlack *et al.* in combination with frequency sweep C-V methods to measure the interface state density as a function of the bandgap energy, but this only worked for structures where interface states do not respond to frequencies > 1 kHz, i.e. when there was no dispersion in the minimum capacitance [43]. The main downside of the quasi-static C-V technique is that it requires measurement of very low currents (picoamps to femtoamps), thus it is sensitive to noise, which can make it challenging to obtain smooth C-V curves.

High frequency C-V

This is one of the most common C-V characterisation techniques, used for quick assessment of the MOS stack quality. Previous research has shown that the stretch-out and hysteresis of high-frequency C-V curves can be used to qualitatively assess the presence of electrically active defects in the system, with more stretch-out indicating more interface states, as shown by Dalapati *et al.* [36], for example. Horizontal shifting of the curve can be an indicator of an increase of the charge in the dielectric.

In this work, the horizontal shift of C-V curves is measured by extracting the voltage at which the slope of the curve is maximum: V gate-maximum-slope, or V_{GMS}. The gate voltage at maximum slope is directly related to the flat-band voltage,

i.e. to the surface potential of the semiconductor-oxide interface. Therefore differences in threshold voltages of a MOS stack correspond to differences in flat-band voltages, and this can be used as a comparative tool to assess the quality of two gate stacks. For an n-type semiconductor, a positive shift of flat-band voltages has been widely used as an indicator of increased amount of fixed charge in the oxide. For example, O'Sullivan *et al.* observed flat-band shifts as a result of thermal processing on Si/SiO_2 systems [46].

As previously stated, the value of the accumulation capacitance for large enough DC biases corresponds to the dielectric capacitance. Starting from this value and using the equivalent oxide thickness (EOT) and physical thickness measurements of the dielectric by transmission electron microscopy, Dalapati *et al.* extracted the effective dielectric constant of high-κ dielectrics on p-doped and n-doped GaAs [36].

A sweep of the DC voltage component in two directions causes the C-V curve to display hysteresis, which can be used to obtain a number of useful pieces of information. The presence of hysteresis is related to the slow discharge time of the charge trapping states. Since the backward sweep is performed immediately after the forward sweep, the states that trapped charges have not yet released it and only the free charge will be detected. Therefore the amount of hysteresis is directly related to the density of active charge traps. Passlack *et al.* showed that for a Ga_2O_3/GaAs MBE-grown stack, 350 mV of hysteresis corresponded to sheet concentration of traps equal to $6x10^{11}$ per square centimetre [43]. A large hysteresis in high-frequency C-V curves in combination with other characterisation methods can confirm the presence of a transitional oxide layer close to the semiconductor surface.

Chapter 3, in which MOS capacitor measurements are presented, contains an expansion of all above information on high-frequency C-V.

Conductance

Standard C-V measurement are carried out by sweeping the DC voltage component of the applied voltage while keeping the frequency constant for each curve. The conductance (G) method instead is based on capacitance measurements carried out whilst keeping the gate at a fixed voltage and sweeping the frequency, and acquiring the curves for several gate biases. Despite the name, conductance method is still based on capacitance measurements, but an equivalent circuit approach is used to derive formulas allowing the conductance of interface states to be calculated and to lead to the mid-gap value of D_{it} [47, 48]. This technique was also used by Passlack *et al.* to detect presence of a transitional dielectric layer at its interface with the semiconductor. Dalapati *et al.* combined the conductance method with standard C-V measurements to calculate a mid-bandgap value of D_{it} by using the maximum capacitance and conductance values from the conductance method and C-V curves at 100 Hz respectively.

2.3 Causes of defects

Certain fabrication processes involve the use of photons and/or charged particles with high incident energies. Upon substrate penetration, these energetic species give rise to defects by causing ionisation of atoms. Such processes include all types of lithography (e-beam, X-ray, photo), metallisation techniques (e-beam and thermal evaporation, sputtering), and plasma-based processes (RIE, etc.). The material damage threshold will be dependent on the typical energies of the process, the type of substrate (e. g. III-Vs are more susceptible to process damage than silicon), and also the doping of the uppermost layer. When this high-energy radiation is inflicted upon the oxide/semiconductor system, the number of growth-related defects, such as net positive charge in the oxide, charge-neutral traps in the oxide and charge traps at the interface, increases. Generally, a process is considered low-damage if incident ion energies are a few eV or less, and damage should be expected from processes using a few hundreds of eV [50].

2.3.1 Metallisation-induced damage

Compared to plasma or lithography processes where incident energies are between a few hundreds to hundreds of thousands of eV, typical incident atomic energies of metal deposition processes are only a few or a few tens of eV. For this reason, when comparing damage caused by metallisation parameters, a sensitive technique is required. Chen *et al.* [50] assessed metallisation damage of III-V substrates by observing degradation in photoluminescence response to argon laser excitation of strained (more susceptible to damage) GaAs/InGaAs quantum wells buried at varying distances from the substrate surface. All three main metallisation techniques used in electronic devices fabrication - e-beam metal evaporation, thermal evaporation and deposition by sputtering - were shown to cause damage to the underlying substrate. Evidence of damage was found after both argon sputter deposition and e-beam evaporation of 10 nm Ti, and after thermal evaporation and e-beam evaporation of 10 nm Au, especially significant at the depth of 30 nm below the surface. Although normally considered a non-damaging metallisation method, thermal evaporation was shown to induce trap levels in Schottky diodes when highly energetic vaporised metal atoms exhibiting a distribution of kinetic energies transfer that energy to the semiconductor surface upon landing on it, resulting in some disordering of the surface. In addition to that, e-beam evaporation process generates x-rays due to its source of radiation, which further degrades the substrate. In Chen's work, the e-beam evaporation method was found the most damaging. However, the type of metal played a role in the distribution of damage, with Ti-deposition induced damage displaying a significant decrease between 30 and 60 nm from the surface, with no change observed in Au and Al. This was explained by the fact that for the same deposition rate the electron beam current is varied for different metals. The effect of current variation was also revealed when Ti was deposited by e-beam evaporation at different rates, where higher deposition rates are achieved through higher electron beam currents. Reducing the deposition rate from 0.5 nm/s to 0.08 nm/s showed a 50% reduction in damage caused to the substrate, even considering the six-fold increase in deposition time required. Chen's findings were confirmed by Burek *et al.* [51], who evaluated metallisation damage

on Al_2O_3/InGaAs and HfO_2/InGaAs structures and found significant damage from thermally-evaporated Ni and e-beam evaporated Pt in both stacks, with e-beam evaporation causing significantly more damage. In Burek's work the damage was evaluated by measuring MOSCAP capacitance and conductance and analysing C-V(f) and G(f,V) plots. With regard to the electrical properties of the metal-semiconductor interface, Auret *et al.* [52] concluded that these are determined by the first few monolayers of metal deposition.

2.3.2 Thermal damage

Thermal process damage to a semiconductor surface and its dependence on the gaseous atmosphere in which the process is carried out was first reported by Revesz *et al.* [30]. Defects were introduced on the silicon surface when the material was subject to thermal treatment at 1150°C. With regard to III-V/high-κ stacks, the main issue with thermal processing is atomic inter-diffusion between layers. Cabrera *et al.* [53] investigated indium diffusion by examining the top surface layer of an HfO_2/$In_{0.53}Ga_{0.47}As$ sample before and after receiving a post deposition anneal in forming gas (a mixture of N_2 and H_2). Both desorption and diffusion of arsenic and indium through the HfO_2 interfacial region to its surface were observed. Samples annealed in N_2 instead of forming gas yielded similar results, and XPS measurements confirmed their findings. Despite the results from [53] suggesting that the hydrogen in the forming gas promotes arsenic desorption, the role of hydrogen is still not well understood. Similar indium diffusion was also observed for other systems, such as Al_2O_3/InGaAs gate stacks, by Weiland *et al.* [54]. For those systems also out-diffusion of Ga and As through an ALD Al_2O_3 layer was observed, after annealing in a N_2 environment with temperatures ranging from 400°C to 700°C. This suggests that elemental out-diffusion is more problematic for ALD Al_2O_3 layers on InGaAs than for HfO_2 on InGaAs systems, where only out-diffusion of indium was observed. Indium diffusion occurrence is common to other III-V/high-κ stacks: Dong *et al.* observed indium out-diffusion from an InP substrate through HfO_2 and Al_2O_3 after post deposition annealing [55].

2.4 Reducing MOS defects by process optimisation

2.4.1 Wet surface clean

Surfaces are inherently reactive, due to unpaired valence electrons in surface atoms. Surface and interfacial defects arise from the presence of native oxides, surface contamination, dimers and dangling bonds at the semiconductor surface, as well as the presence of Coulomb scattering and surface phonon effects. This can result in Fermi-level pinning and poor electrical performance. Normally, the ionised atoms on the surface react with oxygen in the air forming a layer of native oxide, and thus electrically passivating the surface. This layer is commonly removed by wet or dry chemical treatment and the surface is passivated in a more controlled way. Comparative studies of the surface before and after surface clean are a standard way to obtain information about the effectiveness of the clean, and the number of defects on the surface.

In order to achieve an acceptable oxide-semiconductor interface quality for dielectrics deposited ex-situ on III-V surfaces, it is required to use a dielectric layer with good passivating properties to passivate the dangling bonds. Sometimes a separate 'interfacial' layer is deposited prior to the bulk dielectric layer deposition. If that is not possible, the surface can be passivated with a chemical treatment prior to loading the sample in the dielectric deposition chamber. In any case, the first step before surface defects passivation is removal of native oxides, commonly using hydrofluoric acid (HF), hydrochloric acid (HCl), piranha type etches and standard organic solvent cleans.

Passivants suitable for III-V surfaces will depend on the origin of the defect, which determine the local charge distribution. In Si, for example, the largely isolated dangling bond is the dominant type of surface defect, which can be easily passivated with atomic hydrogen. In III-V arsenides, due to variability in surface termination atoms, many types of defects and subsequently many types of charge

distributions could be present. Each type of defect will be effectively passivated by a different substance, with the most common ones being hydrogen, halogens, sulphur, and selenium. Amongst the most effective are sulphur-based solutions, such as S_2Cl_2 and H_2S, with the best results reported for ammonium sulphide ($(NH_4)_2S$) [56]. Aqueous ammonium sulphide has been shown to both fully remove the native oxide and chemically and electrically passivate surfaces of InP, GaAs and InGaAs.

Fermi level pinning occurs due to unstable oxygen coverage of the semiconductor surface. Sulphur is less chemically reactive and has the same electron number in its outer shell as oxygen. A few monolayers of sulphur on GaAs surface can prevent GaAs from oxidation and passivate it in a more stable way.

Optimisation of surface cleaning and surface passivation processes is always required for III-V materials, because a particular treatment can have different effect depending on arrangement of atoms in the surface layers. For example etching processes can produce differential etching rates and surface roughening, and for thermally-assisted treatments like forming gas annealing care has to be taken not to decompose III-V substrates.

XPS studies are often part of the passivation process optimisation: combined with surface roughness AFM scans, they were reportedly used to find the optimal processing parameters for sulphide treatment of InGaAs surface [56]. Furthermore, they were also used to study the reactivity of GaAs surfaces after treatment with hydrofluoric acid (HF) and ammonia solution (NH_4OH), which simultaneously removed the native oxides and passivated the surface against the formation of defect-inducing As-O species [33]. Kim *et al.* [22] successfully treated a GaAs surface with buffered HF as well, prior to deposition of 3.5 nm of HfO_2 deposited by ALD.

2.4.2 Surface plasma treatment

Room temperature plasma treatments can be categorised according to their purpose: cleaning the surfaces off organic contamination or passivation of surface defects. For the former, oxygen plasma (O_2) ashing in either a barrel asher or an etcher is commonly used, whereas for passivation, hydrogen (H_2), nitrogen (N_2) and oxygen-based (O_2, N_2O) atmosphere allows termination of dangling bonds or oxidation of the surface, respectively [57]. Oxygen plasma ashing has been shown to effectively remove post resist development residue, however the oxygen plasma RF power must be carefully tuned in order not to cause damage to the material. There is not much reported evidence towards the use of oxygen plasmas for fabrication damage mitigation on arsenides. Nitrous oxide (N_2O) plasma treatment on a GaN surface was reported to reduce D_{it} by an order of magnitude [58], but the mechanism is not well understood, and it is unclear if the same treatment would be beneficial on arsenides as well.

2.4.3 Annealing for damage mitigation

It was previously mentioned in the thermal damage section that annealing at temperatures above 1000°C could induce damage to the substrate. However, if temperature are kept below 500°C thermal annealing can be used to remove ionising radiation damage, with different temperatures applying to passivation of different types of defects. Oxide charge and interface traps can be removed at 450°C or lower, while neutrally-charged oxide traps need much higher temperatures. Thus an annealing step at an insufficiently high temperature will result in ionising damage not being completely removed.

During fabrication of the gate stack, it is common to use annealing to reduce defects in the bulk and at the interfaces of the dielectric and semiconductor materials. Annealing is generally used twice: after the deposition of the oxide (post-oxidation anneal) and after the deposition of the metal (post-metallisation anneal). Post-oxidation anneal reduces the oxide charge density and passivates the interfacial

states that III-V/high-κ interfaces are prone to because of the tendency of III-V surfaces to non-stoichiometry, as mentioned earlier in the text. Post-metallisation anneal mitigates the damage caused by metal deposition processes that cause ionisation and introduce defects. A typical post-oxidation anneal is carried out at 500°C in nitrogen atmosphere, whereas for post-metallisation anneal the temperature is generally lower (400°C) and the thermal step is carried out in forming gas ambient. The dependence of the anneal outcome on the annealing ambient is something that was already reported in [30], where a post-oxidation anneal at 1150°C in helium resulted in no measurable defects, whereas annealing in hydrogen at the same temperature created more defects. Despite what is observed for silicon by [30], incorporation of hydrogen makes thermal annealing more effective for some material systems, as it will be discussed shortly. When sufficient active hydrogen species are introduced into the oxide, the temperature required to anneal out the radiation-induced charges can be significantly reduced. In the following the characteristics of thermal annealing for the most commonly used gases (nitrogen, hydrogen and oxygen) will be discussed.

2.4.3.1 High-temperature nitrogen anneal

Annealing in nitrogen atmosphere can be used to drive crystallisation of dielectric layers. Kim *et al.* [22] showed that for a deposition temperature of 290°C the as-deposited HfO_2 layer was only partially crystallised and had a 1 nm interfacial layer. However, adding post-deposition annealing step in nitrogen atmosphere at sufficiently high temperature supplied enough kinetic energy to achieve a well-ordered and fully crystallised film. The best results were obtained by heating the film to 600°C. Above this optimum temperature GaAs and HfO_2 reacted and produced an interfacial Ga_2O_3 layer which diffused to the film surface, increasing the film thickness, and reducing the bandgap energy of the dielectric. Annealing in nitrogen atmosphere was also used by Chen *et al.* [50] to mitigate substrate damage from e-beam deposition of Ti contacts. With a temperature of 450°C only partial recovery (~30%) of luminescence of a quantum well buried 30 nm below the surface was achieved.

2.4.3.2 Hydrogen- and oxygen-assisted

Adding a small part of hydrogen to nitrogen gas (what has already been mentioned as forming gas) during the anneal has been shown to have a defect-reducing effect. The most common composition of forming gas is 5% H_2 and 95% N_2. Kim *et al.* [22] reported that 400°C post-metallisation forming-gas annealing of Pt/Al_2O_3/InGaAs capacitors passivates border traps in the oxide, which was supported by the improvement in the multi-frequency C-V characteristics. Burek *et al.* [51] clarified that improvement in C-V is due to reduction of metallisation-induced trapping defects in the stack, which is restored by the annealing. Trinh *et al.* [59] showed that a wet surface treatment followed by a pure hydrogen annealing at 500°C was effective in reducing InGaAs native oxide and subsequently minimising D_{it}. Beneficial effects from forming gas annealing of III-V surfaces were also reported at 440°C [60] and 450°C [61]. Post-metallisation forming gas annealing at even lower temperature (400°C) on MOS capacitors comprising n-GaAs/Ge/HfO_2/TaN gate was shown to reduce midgap D_{it}, as the reduced frequency dispersion in multi-frequency C-V demonstrated [62].

The use of oxygen and oxygen-argon mixture during the metal annealing step of n-GaAs/GGO MOS capacitors were studied by Passlack's group. Pure oxygen was the most effective at producing positive flat-band voltage shifts, related to an increased effective work function. The change in the effective work function was attributed to the presence of induced electrical dipole layer at the metal/GGO interface. The rise of the effective work function after the anneal suggested a reduction in the dipole layer. Since oxygen annealing caused the most significant change, it was postulated that it is related to the role of oxygen at the metal/GGO interface or in the bulk of the dielectric in such a way that annealing in oxygen-rich atmosphere could be driving oxygen atoms to fill oxygen vacancies in the material. Because of this positive effect, a significant part of this work focused on oxygen annealing studies.

3 Effect of process variations on electrical response of MOS capacitors

3.1 Overview

Here the changes in high-frequency capacitance-voltage characteristics of MOS capacitors in response to variables in the MOS transistor gate region fabrication module are presented and discussed. The MOS stack comprised a metal gate electrode, a Ga_2O_3/GdGaO dielectric stack and a GaAs surface. The analysis of the high-frequency C-V curves was of comparative nature, with the aim to identify the most promising processes for application to the actual device. When comparing C-V plots, the focus was on observing its five main characteristics:

- C_{acc} - accumulation capacitance per unit area (F/cm^2)
- C_{min} - depletion capacitance per unit area (F/cm^2)
- Stretch-out - peak value of dC/dV, the slope of the curve (F/V/cm^2)
- V_{GMS} - gate voltage value at maximum slope (V), and
- Hysteresis (V) - difference in the forward and reverse sweep gate voltage taken at midpoint of the C-V curve, as indicated in figure 3.1.

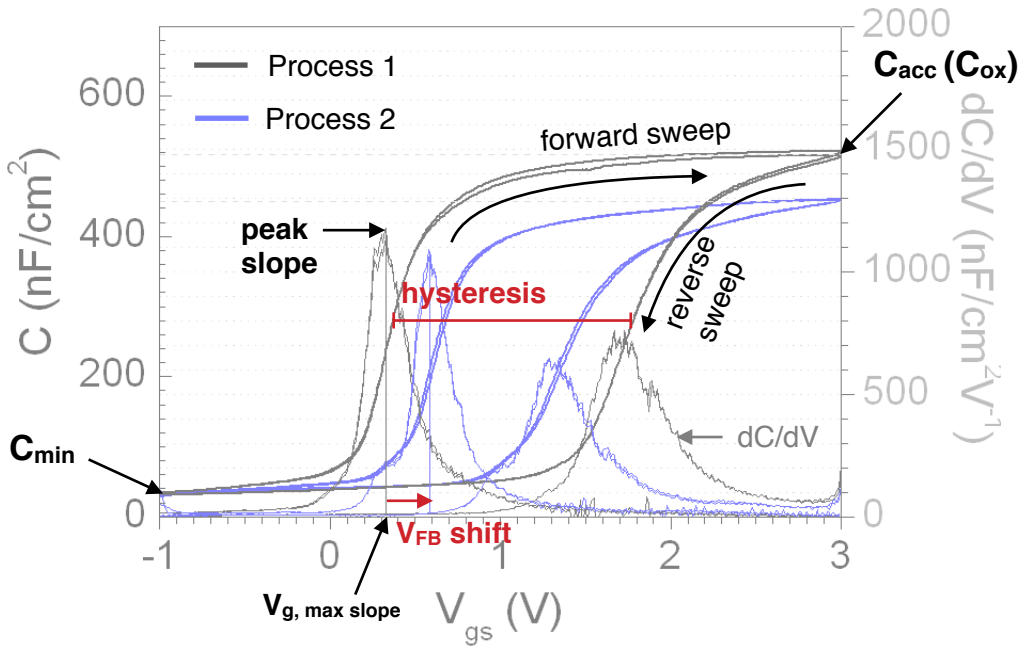


Figure 3.1 Typical high-frequency (1 MHz) C-V curves of MOS capacitors fabricated using two different processes on the same material, showing the main C/V curve and its slope dC/dV. The gate is swept from inversion to accumulation (negative to positive voltages) and back.

3.2 Theory

An overview of different methods for electrical characterisation of defects in MOS stacks can be found in section *Measuring defects* of the background chapter. Here, only the theory relevant to high-frequency C-V measurements of MOS capacitors is presented.

3.2.1 C-V response

The function of the MOS gate in a MOS transistor is placement of charge in the channel via a bias swing on the gate. The equivalent circuit for the C-V measurement is shown in figure 3.2. The applied voltage has a DC component and a small signal component oscillating at the measurement frequency. The concentration of charge underneath the gate will react to the change in gate bias in order to maintain the overall charge neutrality in the system. Figure 3.3(a) illustrates a typical Q-V response.

In an n-doped MOS structure at negative gate voltages electrons are repelled from the semiconductor surface leaving fixed positively ionised atoms, and at positive gate voltages, the electrons from the bulk migrate and accumulate on the surface. Due to this movement of charge in the semiconductor, the behaviour of a voltage-dependent capacitor is observed, consisting of two capacitances in series: the fixed dielectric capacitance C_{ox} and the variable capacitance element, defined by the width of the charge depletion layer, C_{dep} as shown in figure 3.2. In terms of charge, the MOS capacitance is expressed as C(V) = dQ/dV, i.e the MOS C-V curve is obtained by differentiating the Q-V curve, as illustrated in figure 3.3(b).

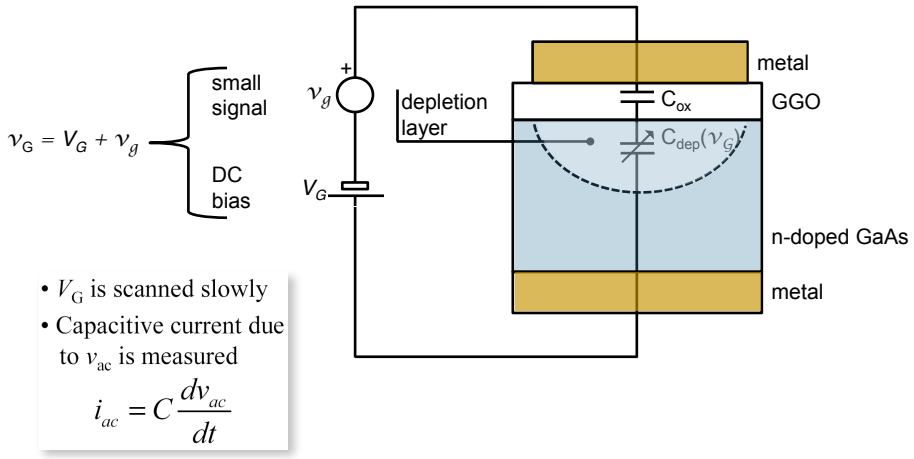


Figure 3.2 The equivalent circuit diagram of a bulk-doped vertical MOS capacitor set up for C_G-V_G measurement.

The change in capacitance results from the modulation of the width of the surface depletion layer by the gate field. If the frequency of the small AC signal applied is lower than 1 kHz, in inversion it will be the minority carriers in the depletion region near the oxide-semiconductor interface mainly contributing to the measured capacitance, yielding the curve labelled *LF* in figure 3.3. At frequencies above 1 kHz (i.e. in *HF* mode) the response of minority carriers is too slow to oscillate with the applied field, therefore only the majority carrier response will be captured in the measured C-V curve, as shown in figure 3.3 by the *HF* curves. As shown in the figures, three different operation regimes can be identified in the measured C-V curves: accumulation when the maximum capacitance is reached, depletion in the

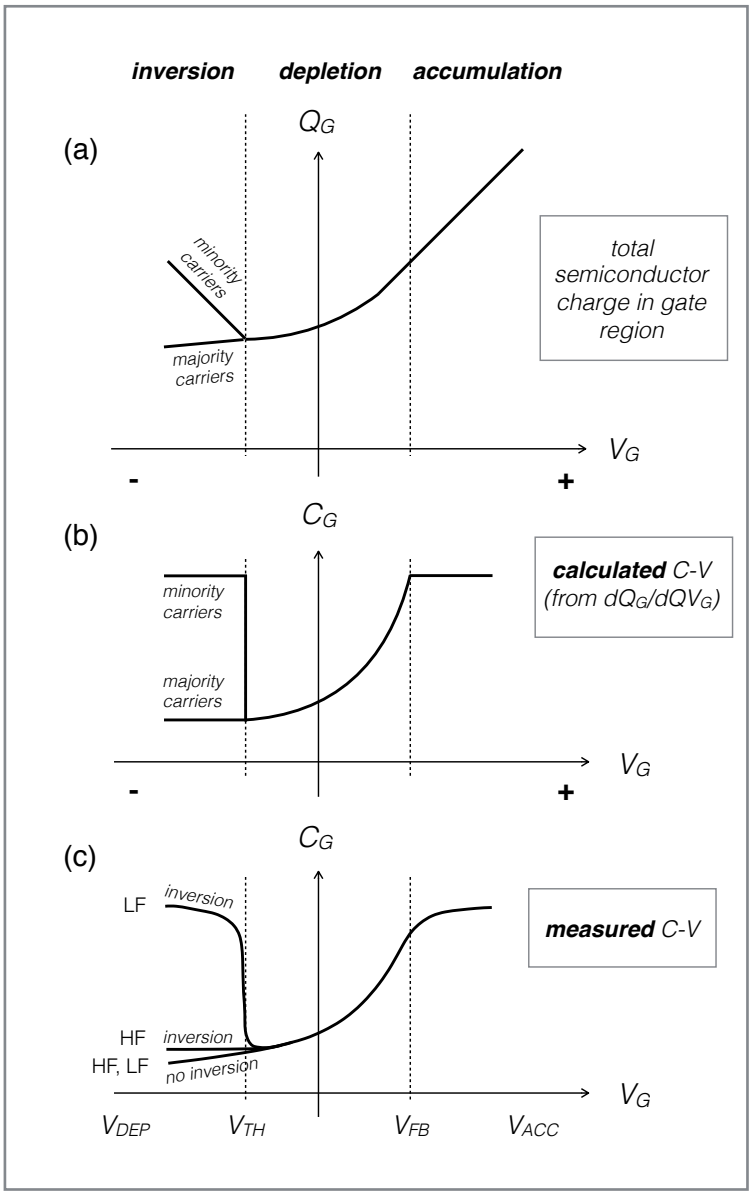


Figure 3.3 The three regions of a voltage-dependent MOS capacitor: accumulation, depletion, inversion. (a) Total gate charge density Q_G as a function of gate voltage V_G. (b) Gate capacitance C_G as a function of gate voltage V_G calculated from dQ_G/dV_G of graph (a), highlighting minority carriers and majority carriers in inversion region. (c) A typical measured C-V, showing the majority (measured with a high frequency HF small signal bias) and minority (measured with a low frequency LF small signal bias) carrier response when the surface is inverted, and carrier response (HF, LF) when inversion of the surface does not take place (*deep depletion*) due to overly fast DC bias voltage giving insufficient time for inversion charge generation.

transition region and inversion when the minimum capacitance is measured. These regimes will be explained in more detail in the next sections. In this work, most C-V curves were measured with a small signal component at 1 MHz frequency. This is because, due to the design of the final device not being based on an inversion channel operation, only the response of the majority carriers (i.e. electrons) is of interest. Therefore, here the C-V is used as a probe for the electrons trapped by mid-bandgap defect-induced states, either in the oxide or at the oxide-semiconductor interface.

3.2.2 C_{acc} and C_{min}

When a high positive voltage (2-3 volts for these devices) is applied across the device, free electrons have accumulated at the surface of the semiconductor, and the electron concentration has reached maximum. This corresponds to a capacitor with zero depletion region, meaning that the total measured capacitance equals the oxide capacitance, as figure 3.4(a) illustrates. If the electric field across the device is reduced, not all free carriers accumulate at the semiconductor surface, and the total capacitance will result of the series between the oxide capacitance and the depletion capacitance, rapidly decreasing while voltage is reduced. If the applied voltage is reduced to high negative values, the extension of the depletion region will reach its maximum, thus yielding a constant measured capacitance at a C_{min} value which will not be affected by further increases in the strength of the electric field.

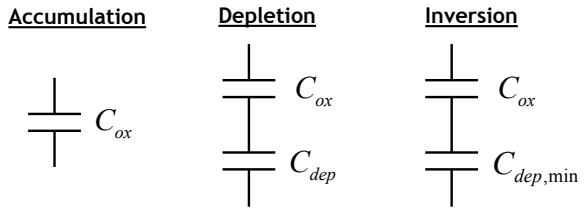


Figure 3.4 High-frequency C-V: In accumulation, the incremental charge is located at the semiconductor surface, in depletion, the charge is located at a distance from the semiconductor surface, and in inversion, the minority incremental charge is located at the semiconductor surface and the majority incremental charge is located at the bottom edge of the depletion region.

The value of C_{min} is given by the series of the oxide capacitance with the minimum depletion capacitance, and can be calculated using the expression in (3.1).

$$C_{\min} = \frac{C_{ox} C_{dep,\min}}{C_{ox} + C_{dep,\min}} \tag{3.1}$$

Since the capacitance indicates the amount of charge stored at the plates with respect to voltage drop between the plates (figure 3.3), the physical origin of the accumulation capacitance difference between two samples is either more charge is accumulated at the surface at the same gate voltage, or there is a difference in the physical separation between the capacitor plates, or in other words, the equivalent dielectric thickness has changed.

3.2.3 Stretch-out

It has been said how in the depletion region the total measured capacitance is given from the series of the oxide capacitance C_{ox} and a variable capacitance related to the extension of the depletion region C_{dep}. The variable component itself arises from the parallel of two components: one is the capacitance related to the free semiconductor carriers, and the other is a parasitic resistance R_{it} and parasitic capacitance C_{it} in series, as illustrated in figure 3.5. The parasitic component is linked to the density of interface traps at the surface of the semiconductor, and it is fixed throughout the depletion. A higher density of interface traps will increase the value of the variable capacitance related to depletion, meaning the curve will reach accumulation for higher voltages and the stretch-out (or slope) of the curve will be lower. On the contrary, if less charge is trapped, the transition from depletion to accumulation will happen over a smaller range of voltage. It thus becomes obvious how the influence of a fabrication process variable on the stretch-out of a curve can be used to assess if the process introduces additional interface states. The slope of the C-V curve is related to the majority carrier concentration in the bulk of the semiconductor (see equation (1) in [63]) for non-heterostructure bulk materials, or in the 2D electron gas layer of the quantum well junction for multi-layer structures.

3.2.4 V_{TH} and V_{FB}

The threshold voltage V_{TH} is the gate voltage at which the depletion charge has reached its maximum and the surface inverts accumulating minority carriers at the surface. The flat-band voltage V_{FB} is the gate voltage required to compensate for the band-bending arising from the built-in potential due to workfunction differences between the metal and the semiconductor, and to flatten the bands. Extracting both of these from the $dQ_G(dV_G)$ shown in figure 3.3(b) is only possible for low frequency C-V due to a sharp transition of capacitance trace at these points. In real high-frequency C-V it is more problematic, because of curved corners and sometimes large stretch-out, and because at high frequencies it is not obvious at which point the inversion takes place. For these reasons, an arbitrary value of V_G in the region between the V_{FB} and V_{TH} was used to enable a quick assessment of the flat-band / threshold voltage shift as a result of processing (as indicated in figure 3.1), namely, V_{GMS} - the gate voltage at which the slope of the C-V curve is maximum. The point of the maximum slope marks the gate voltage at which the rate of charge accumulation is the highest. If at the same gate voltage, the V_{GMS} is shifted, a change in the electric field in the oxide has taken place. An observed change in V_{GMS}, i.e. a horizontal shift of the C-V curve as shown in figure 3.6, could be linked to the following: a) immobile charges in the dielectric due to structural defects, and b) the value of the effective workfunction of the gate metal. This way, a positive/negative shift would correspond to an increase/decrease of process damage induced fixed positive charge, or decrease/increase of the effective workfunction of the gate due to changes in material properties of the gate metal.

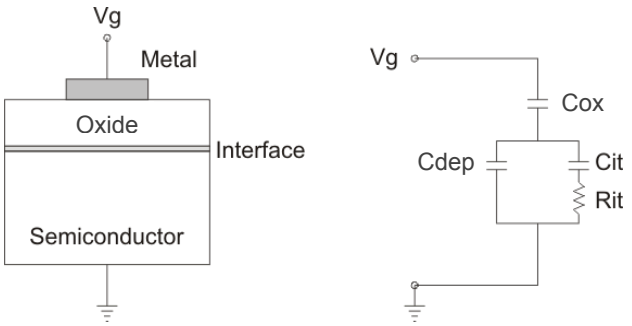


Figure 3.5 The equivalent circuit diagram of a bulk-doped vertical MOS capacitor including parasitic elements due to interface defects.

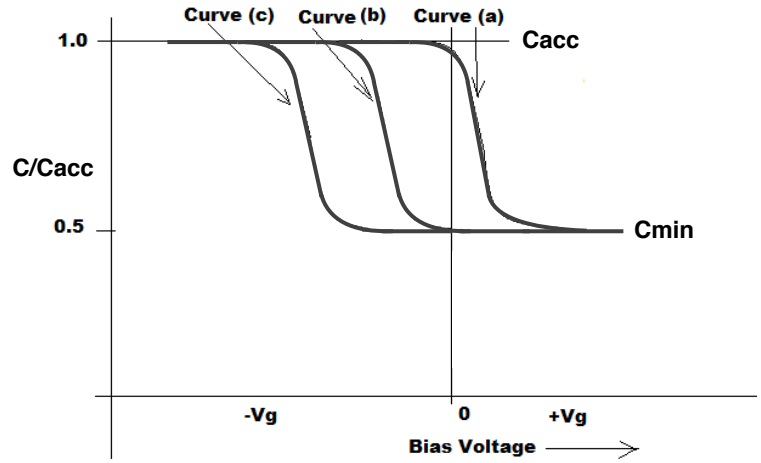


Figure 3.6 Flat-band voltage shift: (a) ideal C-V curve; (b) parallel negative shift due to a combined effect of band bending (i.e. effective workfunction of the gate) and fixed / mobile / trapped charges in the oxide; (c) effect of trapped interface charge.

3.2.5 Hysteresis

The hysteresis phenomenon is observed when a double sweep of the gate voltage is performed, where a reverse sweep immediately follows forward sweep over the same voltage range. Hysteresis is calculated as the difference in values of the gate voltage at mid-capacitance $C_{mid} = (C_{acc} - C_{min})/2$ extracted from the forward and reverse sweep. Hysteresis occurs due to trapping and de-trapping of electrons at mid-bandgap energies that are not emptied fast enough. For a high-frequency AC signal measurement where f = 1 MHz, most of the traps states (fast and slow) are filled during the forward sweep. When ~1 min later (for ~5V/min sweep), during the reverse sweep, the C_{mid} gate voltage value is higher, some charge is still trapped. If, as a result of processing, more charge-trapping defects have been induced, it will be reflected in a larger hysteresis, suggesting that more charge was trapped during the first sweep. Thus, the hysteresis is a qualitative measure of interface state density, and an increase in the hysteresis would indicate an increase in the number of energy states in the forbidden gap of the dielectric and degradation of the interface quality.

3.3 Material

In the following three sections the behaviour of mobile charge in the semiconductor under the influence of the bias applied between the gate and the semiconductor separated by a dielectric layer will be analysed. Three types of material design and capacitor contact design were used: one with the active region formed in a thick bulk-doped n-GaAs layer with charge accumulating at the oxide-semiconductor interface and two with a quantum well active region formed in a thin InGaAs layer away from the oxide-semiconductor interface where charge is supplied by a delta-doping layer and confined in the channel by AlGaAs barriers (see figure 3.7). The first two structures are built on a conducting substrate and comprise doped layers for vertical fields and gate electrode and source/drain ohmic contacts formed on the top and the bottom surfaces of the wafer, i.e. *vertical* structures, whilst the third structure has a semi-insulating substrate and non-conducting bottom layers with both contacts formed on the top surface of the sample for lateral field direction across the capacitor, i.e. *lateral* structure. All structures had a 500 μm thick GaAs substrate (doped or undoped), a few micron thick GaAs-based III-V epitaxial layer structure and a thin high-κ dielectric layer. The key feature of the material is that all semiconductor layers and the dielectric layer were grown by molecular beam epitaxy (MBE) *in-situ*, i.e. without exposing the sample to air at any point of the process. This ensured that the interface between the dielectric and the semiconductor was formed in a controlled manner and its quality was maintained for all wafer growth runs.

From the substrate upwards, the structure referred to as *bulk vertical structure* (figure 3.7(a)) consisted of a 200 nm heavily-doped GaAs buffer layer (Si, $5\text{x}10^{24}$ m^{-3}), a 100 nm undoped $Al_{0.45}Ga_{0.55}As$ layer, a 200 nm heavily-doped GaAs layer (Si, $5\text{x}10^{24}$ m^{-3}) and a 1500 nm n-doped GaAs layer (Si, $2\text{x}10^{22}$ m^{-3}). The function of the buffer layer is to reduce the density of dislocations due to mismatch between crystalline lattices of the substrate and the layers above. The heavier doping in the substrate and around the AlGaAs ensured a reduced series resistance through the conducting

path extending from the back surface to the oxide-semiconductor interface. The lower doping concentration GaAs immediately beneath the dielectric ensured a great extension of the depletion region to make the material more responsive to the voltage sweep during C-V measurements. The gate dielectric stack comprised a 1 nm Ga_2O_3 template layer and a GdGaO layer ranging between 20 and 30 nm. The thickness of Ga_2O_3 is 1 nm for all wafers grown and the thickness of the GaGdO layer varied between 6 and 30 nm depending on the type of structure.

In the structure referred to as *vertical heterostructure* (figure 3.7(b)) the layers were identical to the bulk structure, with additional epitaxy between the 1500 nm n-GaAs layer and Ga_2O_3: a 10 nm undoped $In_{0.3}Ga_{0.7}As$ layer, followed by a few nanometers of $Al_{0.45}Ga_{0.55}As$ encompassing a monolayer of $1x10^{12}$ m^{-2} silicon to provide delta-doping to the InGaAs layer, topped with an ultra-thin 2ML undoped GaAs interfacial layer to allow the dielectric to be grown with an unpinned interface and good morphology. The AlGaAs layer acts as a barrier, reducing the leakage of electrons provided by the Si delta-doping to the gate metal. The thickness of the GdGaO/GaO dielectric stack for these structures was 10 nm. The reason for using the heterostructure is that it is more affected by Fermi level pinning than the bulk structure due to the nature of charge modulation mechanism in quantum wells, as explained in Chapter II. Since the pinning of the Fermi level is mainly caused by interface states, any changes in the C-V hysteresis would be a good indicator of improvement/degradation of the interface quality. It was for this reason that the double-sweep C-V was only performed when measuring heterostructures. Therefore, in the section containing C-V measurement data, all bulk structure measurements were single gate voltage sweep measurements, with no hysteresis taken.

Finally, the structure referred to as *lateral heterostructure* (figure 3.7(c)) is similar to the vertical one, but grown on a non-conducting substrate. This structure was used for MOS transistor fabrication and measurement, described in detail in the next chapter. The top layers down to the quantum well layer are identical to the vertical heterostructure.

The main difference between the two heterostructures is that in the lateral one, in addition to the doped AlGaAs layer above the active InGaAs layer for free carrier provision, there is also one below the active layer, also encompassed in an AlGaAs layer. Between it and the substrate is an AlGaAs superlattice. A superlattice is a series of layers of alternating composition and is included to reduce the number of impurities that can migrate from the substrate to the active regions of the wafer during the wafer growth. The lateral heterostructure had the thinnest dielectric thickness of the all three structures - 6 nm. In the quantum well structure the active depletion region is away from the oxide-semiconductor interface and the dielectric is thinned, therefore, measuring this structure enables separating the effects of processing on the integrity of the dielectric film from effects of interface damage.

The two vertical structures were mainly involved in the experimental work due to the simple processing required to deposit the capacitor contacts. Circular Pt/Au gate contacts were directly deposited onto the top surface of the sample through a shadow mask. Then the sample was flipped onto the other side and Ni/Ge/Au ohmic contact metal was deposited on the back surface of the sample as a blanket layer. The resulting n-type contact was low resistance without the need for annealing. Such rapid feedback process was the main advantage of the vertical structures. For lateral MOSCAPs fabrication both contacts were deposited by e-beam lithography patterning and metal lift-off on the top surface of the sample with the ohmic contact surrounding the gate contact. An electron-beam metal evaporator was used for all metallisation, and the gate contacts were always 250 μm diameter circles.

During the course of the project, many wafers were grown of each type of structure with a certain degree of variability between them, e.g. random dopant concentration variation and nonuniformities. It was ensured that all comparison studies were performed on samples from the same wafer, thus data plotted on the same graph comes from the same growth. Due to the variability in the wafer growth, no direct comparison between all data from the same layer structure design was possible.

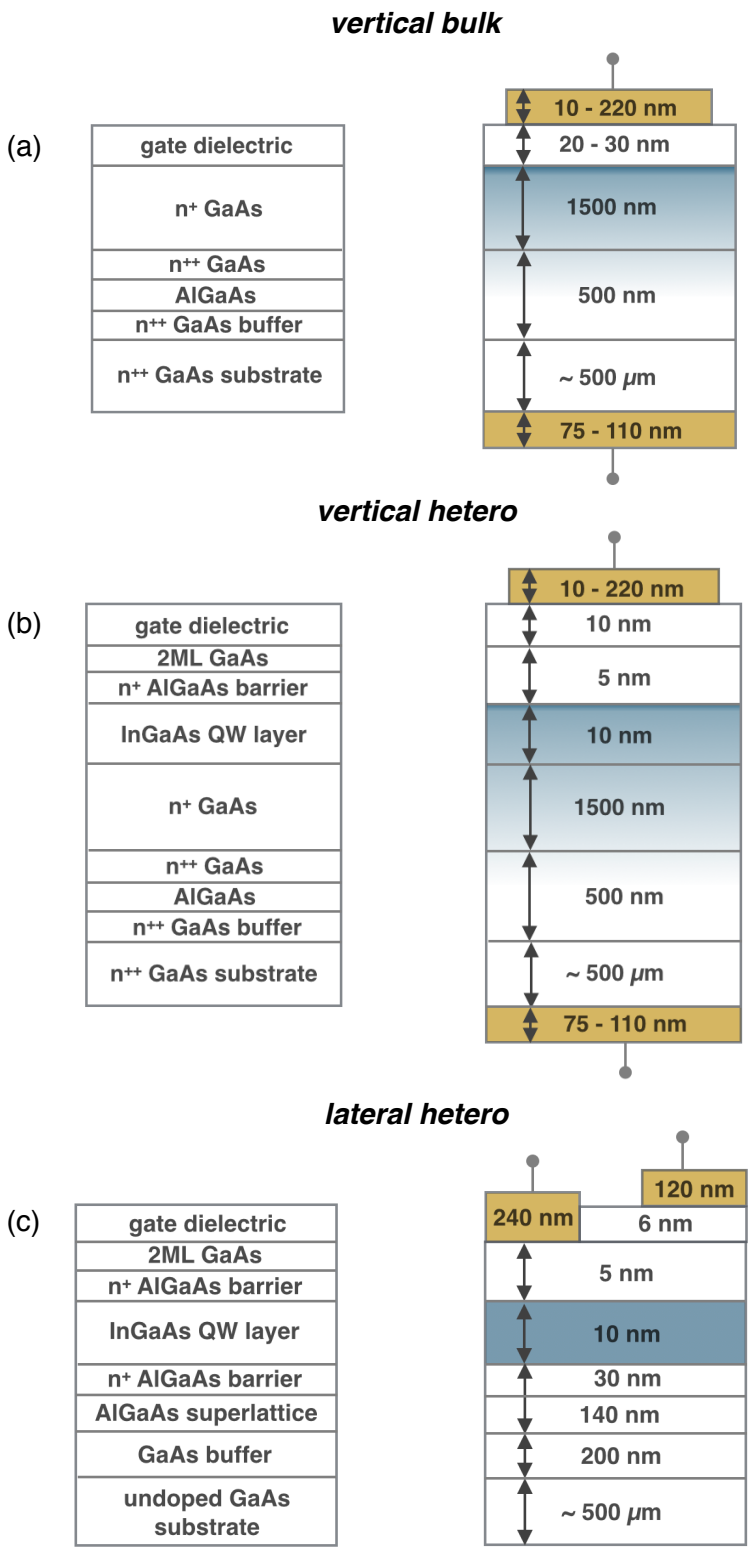


Figure 3.7 Summary of the epitaxial metal-oxide-semiconductor capacitor structures used in this chapter post-MBE growth (left) and with gate and semiconductor contacts deposited. Schematic diagrams on the right are coloured in such a way as to roughly indicate the concentration of electrons when a positive gate bias is applied.

3.4 Experimental results

The experiments carried out all assessed the dependence of charge behaviour on: 1) e-beam lithography patterning, 2) gate metal deposition, 3) thermal treatment of complete MOS stack. The following lists detail how the charge behaviour has been studied in each of the three cases.

1) Dependence of charge behaviour on variables of EBL processing:

- Electron irradiation, assessed through removal of EBL steps and replacement with shadow masks.
- Analysed effect of post-development ash step and exposure dose on post-development resist residue.
- Difference between post-development resist residue in exposed areas (gate) and residue post removal of resist in unexposed areas (source and drain).
- Absence of pre-metallisation surface treatment, direct post-growth measurements.

2) Dependence on various metallisation parameters:

- Varying the composition of gate metal layers
- Using different types of gate metal altogether
- Effect of various metal deposition rates

3) Dependence on thermal treatment:

- Influence of duration and temperature of a rapid thermal annealing (RTA) step in oxygen atmosphere carried out after the gate contact deposition.
- Effect of carrying out the two RTA steps required for the full device fabrication in various gaseous atmospheres, with two cases examined: nitrogen (inert) RTA first, then oxygen (reactive), and vice versa.

- Impact of varying metal thickness on gate contacts which were thermally annealed in oxygen atmosphere.

Below, the experimental activities are explained in further detail, highlighting the main conclusions from each section.

3.4.1 Charge response to varying EBL process parameters

3.4.1.1 Electron beam irradiation

The effect of electron irradiation during e-beam exposure of the gate pattern on charge trapping in the oxide was assessed through removal of EBL steps and replacement with direct deposition of gate metal through a cobalt shadow mask. Both samples were subjected to a standard acetone/IPA clean, and then one sample had gate contact deposited directly on the dielectric surface through a shadow mask and another went through the full lift-off process starting with e-beam exposure of PMMA, MIBK:IPA development and an oxygen plasma ash. The parameters (dose, O_2 ash power and time) of the e-beam lithography process were optimised so that the amount of PMMA residues on the surface was minimised and so that it is reasonable to assume the differences between the C-V curves only due to the effect of electron beam irradiation.

From the C-V results in figure 3.8 it can be seen that the sample which went through the entire lithographic process shows similar values of capacitance density at accumulation C_{ox} to the sample where the gate metal was deposited directly on the surface, suggesting there were no substantial changes in the dielectric properties of the oxide stack after exposure to the electron beam. The measured C_{acc} value is 290 nF/cm², lower than its theoretical value of 348 nF/cm² by ~17% . On the other hand, there is an obvious change in slope which is higher for the electron irradiated sample, and a ~125 mV shift of the curve towards more negative voltages. These results both indicate a reduction in the number of charge traps and a decrease in the positive charge. This can be explained in the following way: the secondary electrons

generated by electron-electron scattering provide a mechanism for charge transfer. These slow moving (~10-100 eV) electrons may be influenced by trapped charge. They then act to neutralise any existing fixed charge, since the material is rendered, temporarily, conducting. It is unlikely for this process to happen also in the bulk of the oxide since, as previously stated, similar values of C_{ox} are observed. One could argue that the differences in C-V could be due to oxygen ash removal of resist residues, which the non-e-beam irradiated sample was not subject to. However, as discussed later, the effect of an O_2 ash is negligible compared to that of electron beam irradiation.

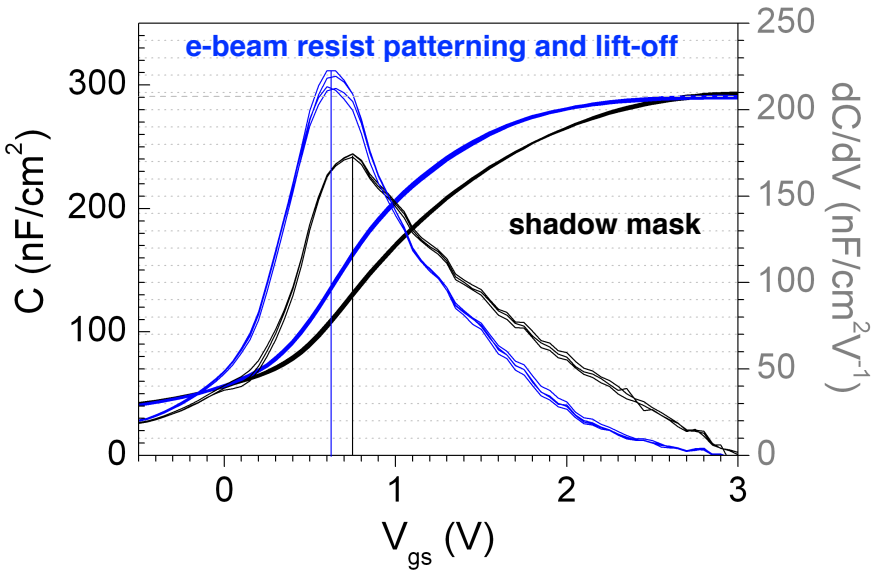


Figure 3.8 The effect of using e-beam lithography process to pattern the gate of a vertical bulk n-GaAs MOS capacitor with 25 nm GGO and Pt/Au gates, shown by comparing high-frequency (1 MHz) single-sweep C-V response of capacitors with gates fabricated by e-beam patterning and metal lift-off (blue trace) and gates fabricated by metal deposited through a shadow mask (black trace).

3.4.1.2 PMMA resist residue in the gate region

It is well known that PMMA is not fully removed in the development process forming a residue on the substrate surface. Using the method shown in the next chapter, it was found that the amount of PMMA present on GaGdO surface after development depends on the exposure dose. Figure 3.9 shows how the differences

in the polymer layer between the oxide and the gate metal are reflected in behaviour of semiconductor charges at the interface between the oxide and the semiconductor. As described in the next chapter, PMMA residues result in complete detachment of the deposited gate metal from the dielectric surface underneath, and it remains so at all exposure doses. However, the PMMA residues comprise of two elements: a 1-3 nm thick layer and granular blobs of PMMA on top of the thin layer. A dependency of the amount of granules on the exposure dose was found, and the differences in C_{acc} shown in the C-V plots in figure 3.9 where the accumulation capacitance is the sum of oxide capacitance and residue capacitance in series, confirming that at higher doses the average thickness of the PMMA residue decreases. The fact that no other change but the C_{acc} has been observed in the C-V characteristics suggests that the only parameter affected with the change in exposure dose is the average thickness of the residual PMMA layer.

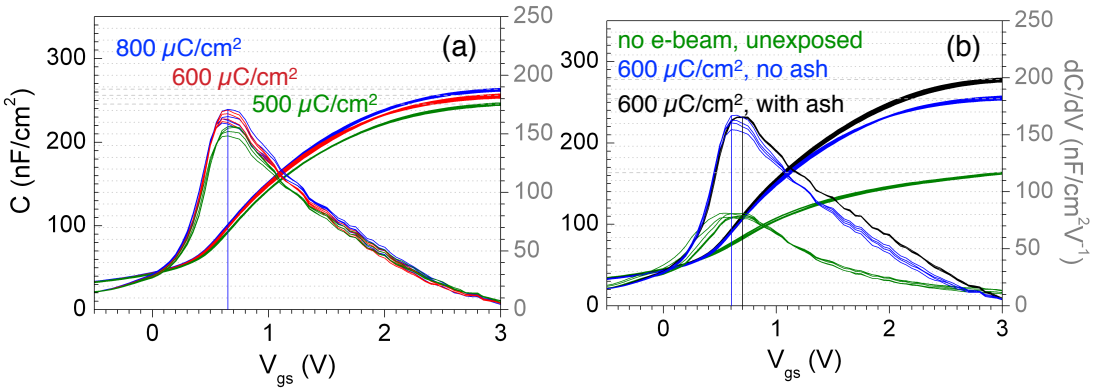


Figure 3.9 The effect of PMMA residues in the gate region of vertical bulk n-GaAs MOS capacitors with 25 nm GGO and Pt/Au gates when using e-beam lithography process to pattern the gate, shown by comparing high-frequency (1 MHz) single-sweep C-V response. (a) Varying the e-beam exposure dose. (b) Comparing the residues in the gate region (exposed) with the residues outside of the gate region (unexposed) with a control residue-free surface (with ash).

The dependance of such thickness for three different scenarios is depicted in figure 3.9. The three samples compared all underwent electron beam lithography at the same dose of 600 $\mu C/cm^2$, but then only two of them were developed (with

standard IPA:MIBK development), whereas the third one had its resist layer stripped in acetone, and the C-V contact pattern later defined by shadow mask. Of the two developed ones, one of them was subject to oxygen plasma ash, which removed any PMMA residues from the surface. In fact the ashed sample, indicated in the figure by the black line, was the one with the highest C_{acc} value, thus the thinnest total dielectric layer since all of the PMMA residue was removed. The blue line instead illustrates the result from the sample which was not ashed after development, and confirms a thicker residual PMMA layer. Finally, the sample where resist was stripped by acetone without undergoing development or oxygen ashing was the one with the thickest residual layer, as shown by the green line. Based on these results, it can be concluded that the presence of a thin layer of polymer between the gate metal and the gate dielectric does not have any effect on the amount of charge trapping at the interface, being physically too far from the interface to have an electrically observable effect. However, the residual layer did have an effect on MOSFET device characteristics, as it will be shown in the next chapter.

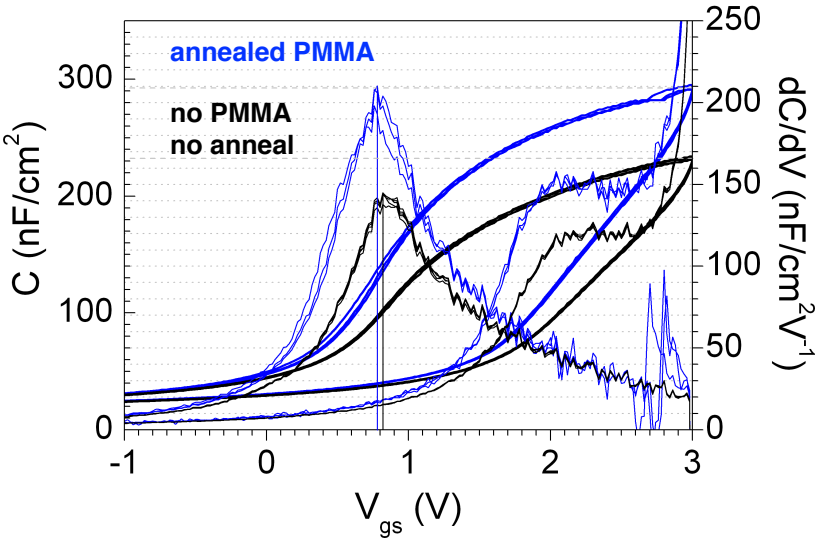


Figure 3.10 High frequency (1 MHz) double-sweep C-V characteristics of bulk vertical n-GaAs structure with 25 nm GGO and Pt/Au gate. Investigating the effect of unexposed resist stripped in acetone and annealed at 430°C during the ohmic contact fabrication. Gate metal was deposited through shadow mask.

The effect of thermal treatment on a surface where resist residues are present has also been assessed, as figure 3.10 shows. The black curve displays the results obtained from shadow mask patterning on a PMMA-free surface, while the blue curve shows the measurement obtained from a sample which was first spun with PMMA, then had the resist layer removed with 1 hour warm acetone soak and finally annealed. This experiment aimed at showing the effects of an annealed layer of resist on the gate-dielectric interface, by mimicking what happens in the gate region when the ashing step is omitted after lift-off of source and drain metal in a gate-last fabrication process. As can be seen from figure 3.10, almost no change in hysteresis or flat-band voltage was measured, indicating no degradation of the gate stack. The unusual fact portrayed by the figure is that the value of C_{acc} in the sample with the resist is much higher that that of the sample without the resist, which suggests that the decrease of C_{acc} normally observed when an additional thickness is present has been compensated by a large increase. This has been interpreted as related to the thermal treatment, with the following explanation: as a result of MBE process there are some lattice imperfections in the amorphous GaGdO layer; thermal treatment gives the atoms in the material system enough energy to redistribute themselves into a more ideal lattice, thus increasing the dielectric constant which is reflected in the higher C_{acc} value. In fact, Dalapati *et al.* [36] observed a densification of high-κ films at post-deposition annealing temperatures below 500°C which improved the dielectric properties.

3.4.1.3 Oxygen plasma for removal of resist residue in the gate region

Oxygen plasma exposure is a technique commonly used to de-scum the surface from organic contaminants. It is most commonly used after development of a resist layer to remove residues from lithographic processes. Several experiments were carried out to explore the effect of oxygen plasma exposure on the GaGdO dielectric layer. The tool used was a barrel asher, and unless otherwise stated the sample surfaces were always ashed for 60 seconds with an RF power of 40 Watts - the minimum power required to achieve stable plasma on the machine.

The black trace in figure 3.11 shows the C-V curve obtained from a sample patterned directly with a shadow mask, whereas the blue trace was obtained for a sample subject to oxygen ash before shadow mask patterning. It is evident from the unaltered hysteresis and the near-coinciding accumulation capacitances and slopes that no increase in surface charge trapping has occurred as a result of the low power oxygen plasma ashing. A small change in the flat-band voltage as indicated by the shift of the curve generally means a reduction of fixed positive charge in the oxide. In a similar way to what happens for electron beam irradiation, the reduction in positive charge can be explained by the oxygen plasma acting a source of free electrons for the oxide surface; those electrons which enter the oxide layer partially neutralise the positively charged ions in the oxide, thus reducing the net fixed positive charge. Contrary to electron beam irradiation, where the high energy electrons reach the interface layer and alter its properties, the oxygen ash does not affect the interface, since the concentration of trapped charge remains unchanged.

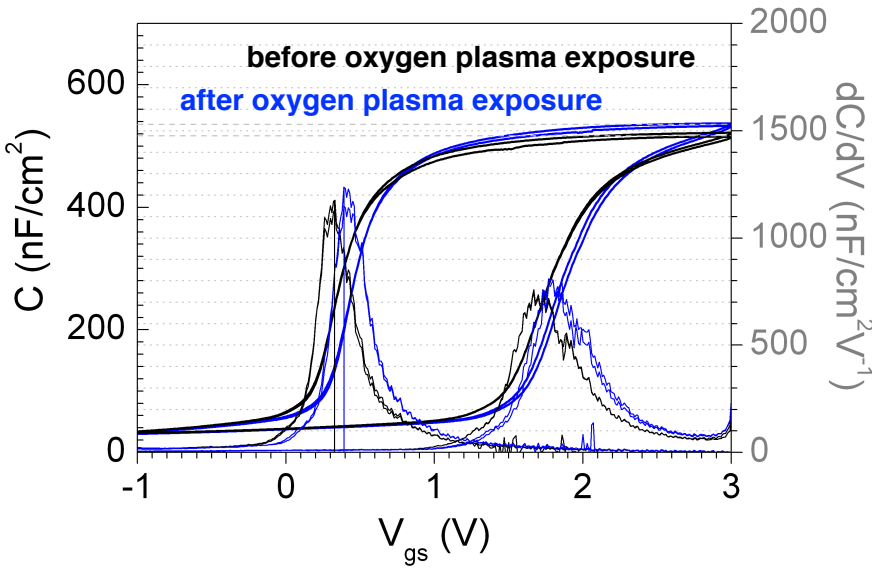


Figure 3.11 The effect of subjecting the GGO surface to oxygen plasma as part of barrel ashing technique, shown by high-frequency (1 MHz) double sweep capacitance-voltage characteristics of vertical heterostructure samples with 25 nm dielectric. The surface was subjected to standard solvent clean prior to oxygen ashing.

The previous experiment analysed the effect of oxygen ash on a resist-free surface. As stated in the previous paragraph, oxygen plasma can be used to remove PMMA residues from the GaGdO surface after resist development. The experiments for the assessment of PMMA residues were carried out on the bulk material structure, where the channel is formed by inversion at the semiconductor surface. These experiments were repeated on a hetero-structure where the channel is located deeper in the semiconductor, separated from the oxide interface by barrier layers. As previously explained in section *Material* of this chapter, this kind of material structure is more sensitive to the effect of the imperfections at the oxide-semiconductor interface because a significant number of trap-inducing defects at the oxide-semiconductor interface can pin the Fermi level below the conduction band of the quantum well layer.

Figure 3.12(a) shows a comparison between a C-V curve obtained by shadow masking a resist-free surface, indicated by the black line, and one obtained from a sample which was coated with PMMA, had the resist removed by acetone strip and was ashed before metal deposition. A great reduction in hysteresis and a positive threshold voltage shift can be observed between the two curves, meaning a reduction of positive fixed charge, consistent with the effect of oxygen ash as interpreted previously from the results on the bulk material. Figure 3.12(b) instead shows a comparison between two samples which were also both coated with PMMA and subsequently had the resist stripped in acetone, but with only one of them undergoing oxygen plasma treatment, represented by the black line. In this case the ash caused a great change in the value of C_{acc}, confirming once again the effective removal of the PMMA residues. Also, a considerable shift of the gate threshold voltage (nearing 1 Volt) is observed, nearly an order of magnitude greater than that observed for the bulk structure subjected to oxygen ash (figure 3.10). If the shift of the C-V curve in the positive direction is attributed to a reduction in the positive net charge in the dielectric stack, then this result would translate into a poorer dielectric quality due to a higher presence of ionised atoms in the heterostructure dielectric film with respect to the bulk material.

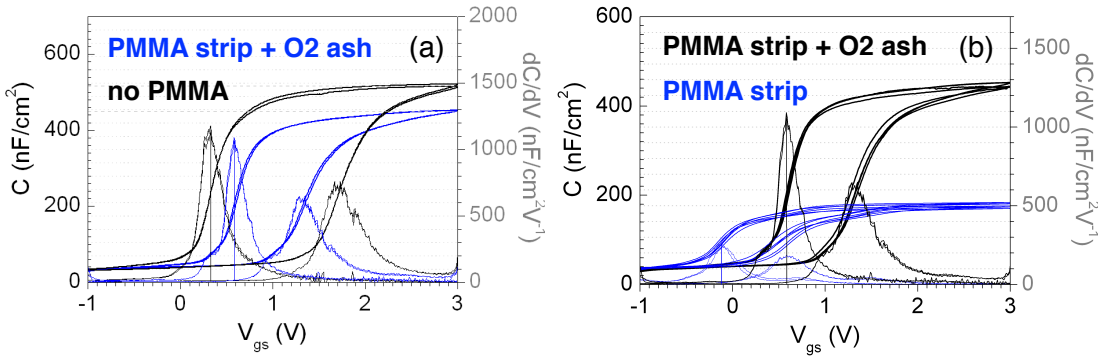


Figure 3.12 High-frequency double-sweep C-V measurements of two vertical heterostructure samples, comparing (a) unexposed PMMA removed with acetone and an O_2 ash against a virgin surface, and (b) unexposed PMMA removed with acetone, with and without an O_2 ash.

3.4.1.4 Wafer cleaning

The baseline MOSCAP fabrication process used ultrasonically-agitated cleaning in standard organic solvents as a default surface preparation procedure. Two samples have been processed with and without organic solvent clean of the dielectric surface, subsequent to which, the gate metal was deposited directly on the substrate using a shadow mask. The electrical characterisation result and the layer structure are shown in figure 3.13.

Two noticeable differences can be observed between the two samples: the upward shift of capacitance values of the curve that received wafer surface clean, and its increased hysteresis. Based on these observations, it can be assumed that the acetone was effective in removing inorganic particles from the dielectric surface and the difference in capacitance values across the gate bias range is due to the difference in the physical separation between the capacitor plates. Since the oxide-semiconductor interface was not exposed during the cleaning procedure, the increase in the hysteresis observed in the cleaned sample could be due to an increase in interface states induced by the structural damage at the interface between GaAs and the oxide stack, caused by high intensity (> 20 kHz) waves that are used in ultrasonic techniques to generate pressure fluctuations in order to create

cavitation bubbles which upon collapsing release enough energy to dislodge and disperse particles. The bubbles in the ultrasonic bath have been previously reported to cause damage to delicate nanoelectronic structures [64]. Ultrasonically-assisted cleaning was the default procedure in the processing of MOSCAPs, and since the detrimental effect was discovered at a later stages of the project, all samples reported here were cleaned in the ultrasonic bath.

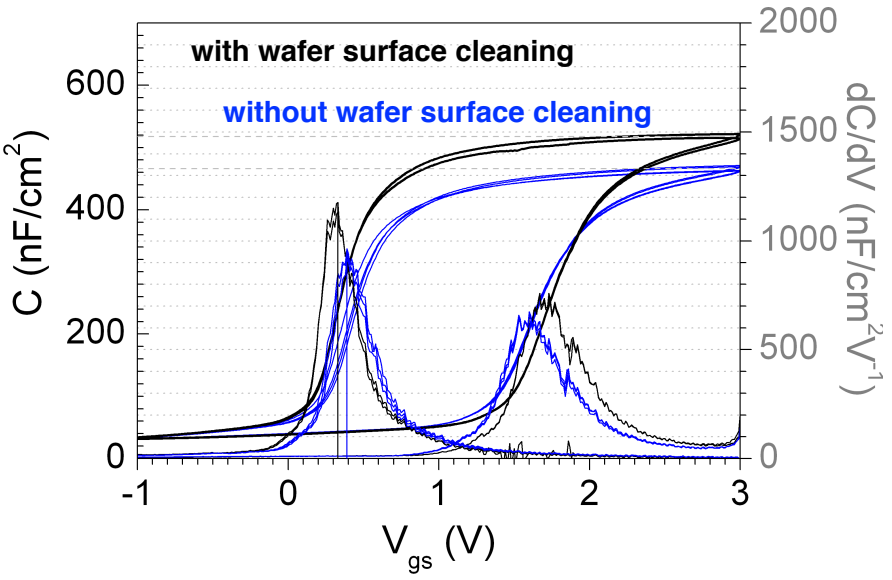


Figure 3.13 The GGO/GaAs sample with and without ultrasonically-agitated clean, shown by high frequency (1 MHz) double-sweep C-V characteristics of vertical heterostructure material with 25 nm GGO. The cleaning procedure was 5 minutes in acetone at room temperature, followed by 5 minutes in IPA.

3.4.1.5 Section summary

In summary, the radiation damage during e-beam exposure was found to induce a change in the amount of charge traps present at the interface. The nearly identical accumulation capacitances obtained from ashed and unprocessed resist-free surfaces indicates that the gate oxide did not react with oxygen during the ashing process. Also the experiments shown prove that removing the barrel oxygen plasma ashing step causes a large drop in capacitance, due to a presence of a residual PMMA layer on the dielectric surface, the thickness and surface roughness of which was reduced with increasing exposure dose.

3.4.2 Charge response to varying metallisation parameters

There is a disagreement in literature about how metallisation process affects charge response. Auret *et al.*, for instance, state that the electrical properties of the metal-semiconductor interface are determined within the first few monolayers of metal deposition [52]. In the following it was assessed if altering several metallisation parameters had observable effects on the measured C-V curves. Theoretically, for a given gate bias a change in the metal stack should not affect the electric field sufficiently to show measurable differences in the amount of charge accumulating on the surface of the dielectric, due to the high conductivity of metals, regardless of which metal is chosen. However, all materials have different electrochemical potential, which means that their workfunctions change when in contact with the dielectric. Thus, because of the Fermi level alignment in a multi-material stack, a change in the workfunction of the gate metal stack corresponds to a change of the surface potential of the bottom layer, which is observed electrically as a shift of the C-V curve. The following experiments were designed to see if a shift in the C-V curve is the only observable difference caused by changes in the metal stack, and to determine if an optimal metal gate stack could be identified.

3.4.2.1 Varying the composition of gate metal layers

Two different approaches were adopted: first, only platinum (Pt) and gold (Au) were used, because of the high workfunction of Pt for which the MOSFET device material layers were designed. On top of Pt, a thicker layer of Au was required to decrease the total resistivity of the contact. Secondly, different metals were studied, with the experiments discussed in the next paragraph.

It was of high interest for this work to try to reduce the thickness of Pt in the Pt-Au stack, the reason being the ease of etching a thinner Pt layer with dry etch techniques in a subtractive type of gate contact formation. The effect of a metal stack with a Pt thickness reduced to 1 nm is shown in figure 3.14. It can be seen than a decrease in Pt thickness from 10 nm to 1 nm shifted the flatband voltage in the

negative direction. The observed effect can be related to the proximity of Au to the dielectric surface. First, the Au layer is only a few atomic layers away from the oxide surface; second, the Pt layer was deposited by standard rate (0.3 nm/s) e-beam evaporation, therefore its thickness might not be uniform in the contact area (the exact value for the non-uniformity of coverage is not known). This could mean that, as far as the dielectric surface is concerned, there is Au partially in direct contact with it. As a result, a change in the gate workfunction takes place, reflected in the flatband voltage shift.

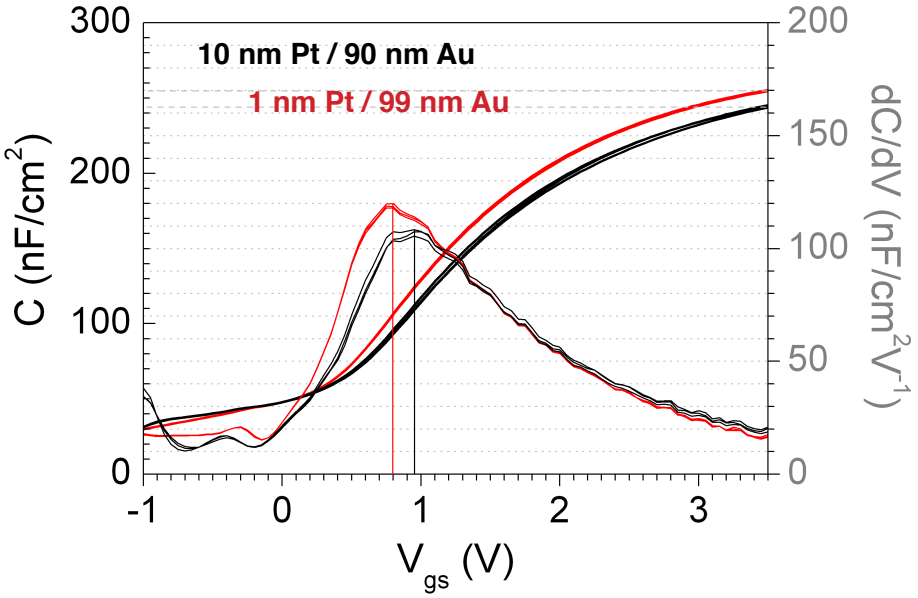


Figure 3.14 High frequency (1 MHz) single-sweep C-V characteristics of bulk n-GaAs with 31 nm GGO where the thickness of Pt in the Pt/Au gate metal stack changed between 10 nm and 1 nm, and the Au thickness was changed accordingly to achieve 100 nm total contact thickness.

After comparing a nearly Au-only stack with a Pt-Au one, the latter was compared with a Pt-only layer, with the results shown in figure 3.15. The most obvious change between the two plots is the value of the accumulation capacitance, which is considerably higher when Pt is the only metal used. Since the resistivity of platinum is approximately 4.4 times that of gold, this is thought to be due to the higher impedance of the 100 nm Pt layer compared with the composite Pt-Au one, where the thickness of Pt is only a tenth of the Pt-only layer. As discussed in the theory

section of this chapter, the accumulation capacitance difference between two samples may be interpreted as more charge accumulating at the surface at the same gate voltage. Thus comparison of gate metal stacks with different impedance will have different associated parasitic capacitances (discussed later in Chapter 5), and the total C_{acc} will consist of a contribution from the oxide and a contribution from the metal stack. The greater value of C_{acc} confirms a higher parasitic capacitance for the Pt-only case, and therefore the need to cap the metal stack with a thick layer of gold.

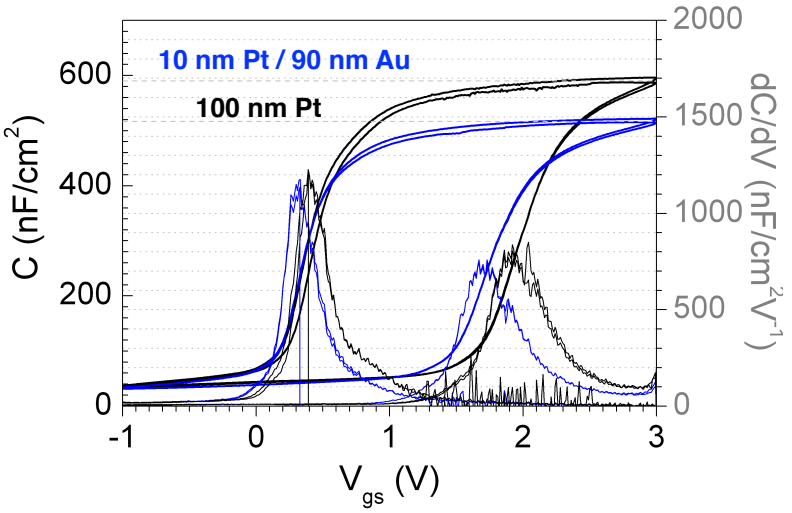


Figure 3.15 High frequency (1 MHz) double-sweep C-V characteristics comparing 100 nm Pt gate (black) and 10 nm Pt / 90 nm Au gate (blue) MOS capacitors fabricated on vertical heterostructure material with 10 nm dielectric.

The results shown in figure 3.15 were measured on the hetero-structure material from figure 3.7(b). A comparison between a Pt-only stack and a Pt-Au one was compared also for the bulk structure, as the plots in figure 3.16 show. In this case, there is no change in the value of C_{acc} observed, and this can be explained with the different dielectric thicknesses between the two epi-structures. The bulk material has a considerably thicker dielectric, at 25 nm, compared with the heterostructure whose GdGaO/GaO layer is 10 nm thick. This will translate into a higher value of oxide capacitance for the latter structure, which will also be less sensitive to parasitic capacitance in the metal stack.

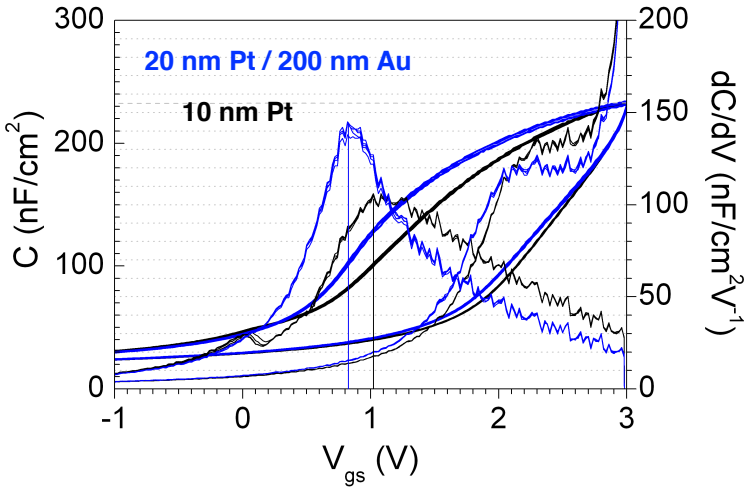


Figure 3.16 High frequency (1 MHz) double-sweep C-V characteristics comparing 10 nm Pt gate (black) and 20 nm Pt / 200 nm Au gate (blue) MOS capacitors fabricated on vertical bulk material with 25 nm dielectric.

3.4.2.2 Using different types of gate metal altogether

After comparing the effect of different thicknesses of Pt and Au in a Pt-Au stack, the possibility of using a metal other than gold as the low resistivity layer to deposit on top of the Pt layer was explored. Given that the final fabricated devices had to be annealed in oxygen atmosphere as a way to raise the workfunction of the gate metal, nickel (Ni) seemed a viable alternative to Au. Platinum and gold are both noble metals, and their effective workfunctions will remain unchanged after a treatment in oxygen atmosphere [65], unlike that of nickel that was reported to result in a higher workfunction facilitated by oxygen diffusion through the metal [66]. Resistivity values for Pt, Ni and Au are provided in table 3.1.

Figure 3.17 shows the different C-V responses measured for Pt (10 nm) / Au (90 nm) and Pt (10 nm) / Ni (90 nm) gates, where the latter showed a positive ΔV_{FB}. As explained in the theory section of this chapter, this could either suggest a reduction of immobile positive charge in the dielectric, a change in the effective workfunction of the gate. Since the wafer and the fabrication process were identical for both samples, the effective workfunction of Pt/Au must be different to that of Pt/Ni. It

seems that top Ni causes an increase in the effective workfunction of the gate, i.e. a decrease in the semiconductor surface band bending. This suggests that, contrary to previous reports stating that the electrical properties of metal-semiconductor system are determined within the first few monolayers of metal [52], it is instead determined by the whole metal stack. The effective workfunction of Pt/Ni is smaller than that of Pt/Au, meaning the subsequent band bending yields a thicker depletion layer for the same applied voltage, thus less carriers modulated during the forward sweep, and causing an increase in the stretch-out of the curve. This effect was observed for all the experiments carried out on *bulk vertical* structures: as summarised by the results in table 3.2, a change in V_{GMS} was always measured. The different values of V_{GMS} differential further confirm that the behaviour observed could only be ascribed to phenomena related to electron-beam metal evaporation process.

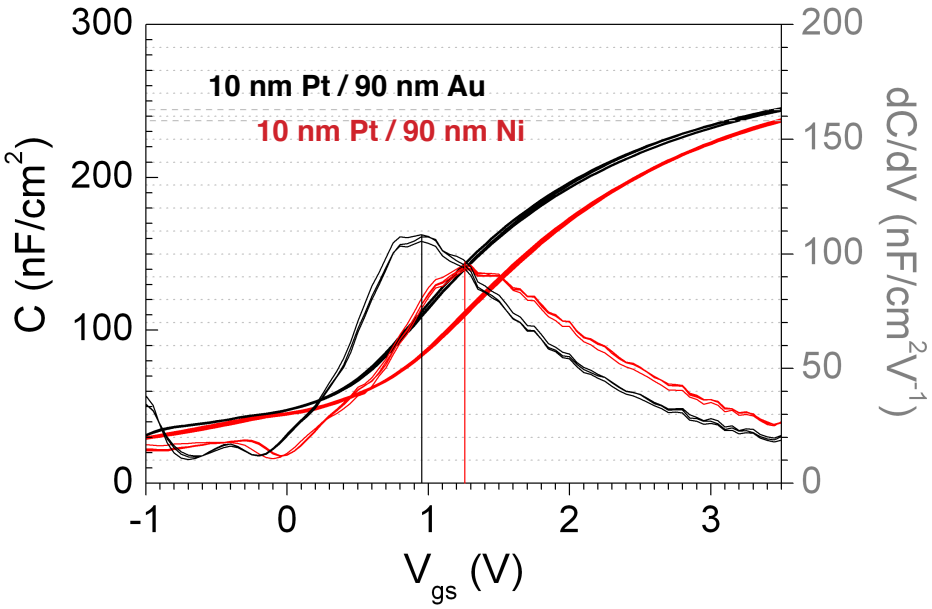


Figure 3.17 High frequency (1 MHz) single-sweep C-V characteristics of bulk n-GaAs with 31 nm GdGaO/GaO dielectric stack where the Au layer on top of Pt in the Pt/Au gate metal stack changed to Ni.

	Resistivity (Ω.m)	Workfunction (eV)
Au	2.24E-08	5.1 - 5.47
Pt	1.1E-07	5.12 - 5.93
Ni	6.85E-08	5.04 - 5.35

Table 3.1 Key parameters of the materials used in the gate metal stacks. Workfunction values are taken from CRC Handbook of Chemistry and Physics 2008.

3.4.2.3 Effect of various metal deposition rates

As previously stated, the deposition of thin layers of Pt is prone to uniformity issues if the deposition rate is too fast. The standard deposition rate used for the samples discussed so far is 0.3 nm/s, therefore requiring only approximately three seconds to deposit a 1 nm film. A decrease of the deposition rate was thus targeted to increase the time required to attain 1 nm to at least twenty seconds at a new rate of 0.05 nm/s, thus attaining a better thickness control and layer uniformity. As the data shown in figure 3.18 show, it is clear that slowing down the deposition rate is greatly beneficial to the resulting structure, as proved by the positive shift in the flatband voltage. In both cases the total thickness of the metal was 100 nm, of which 1 nm of Pt and 99 nm of Au. For the sample deposited with the higher deposition rate, as previously discussed there will be uniformity issues and Au atoms in contact with the dielectric surface. This will not be the case for the sample deposited with slow deposition rate, and the change in the C-V curves mirrors the change in the workfunction. The change in maximum capacitance values should not be taken as a real C_{acc} difference since no full accumulation was achieved at the measurement range for any of the curves. However, a change in accumulation capacitance is possible, given that a slower deposition rate means more exposure to contamination which could yield a lower C_{acc}.

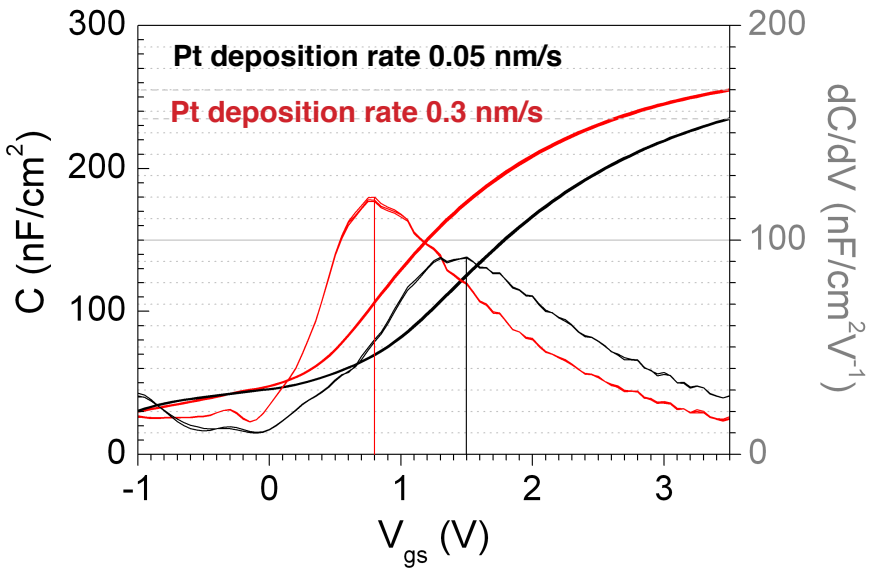


Figure 3.18 High frequency (1 MHz) single-sweep C-V characteristics of bulk n-GaAs with 31 nm GdGaO/GaO dielectric stack and a metal stack consisting of 1 nm Pt and 99 nm Au, deposited at two deposition rates.

3.4.2.4 Section summary

The results from this section are summarised in table 3.2. The common feature of the experiments was a positive shift of V_{fb}, always accompanied by an increase in stretch-out of the curve. A double C-V sweep was performed only in one case, providing a decrease in hysteresis together with the positive V_{FB} shift, and C-V curves overlapping during the reverse sweep. Both a reduced slope and a reduced hysteresis are conflicting results if analysed in terms of trapped charge. However, the reverse sweep overlap indicates that there is no difference in the amount of charge since the same amount of it is trapped. Thus the changes in the forward sweeps could not be ascribed to charge trapping, but only to changes in the effective workfunction, the only other variable in the processes compared.

Gate metal stack	Material	C_{acc}	Slope	ΔV_{FB}	Hysteresis
Pt 1 / Au 99	bulk	1	1	0	N/A
Pt 10 / Au 90	bulk	0.95	0.9	+0.15 V	N/A
Pt 10 / Au 90	hetero (vert.)	0.88	-	0	1.4 V
Pt 100	hetero (vert.)	1	-	+ 0.07 V	1.6 V
Pt 20 / Au 200	bulk	-	1	0	1.3 V
Pt 10	bulk	-	0.7	+ 0.20 V	1.0 V
Pt 10 / Au 90	bulk	1	1	0	N/A
Pt 10 / Ni 90	bulk	0.97	0.87	+ 0.30 V	N/A
Pt 1 / Au 99	bulk	1	1	0	N/A
Pt 1 / Au 99 (slow)	bulk	0.92	0.75	+ 0.70 V	N/A

Table 3.2 A summary of key parameters extracted from C-V plots of the Metallisation section in order of appearance starting from the top of the table. The accumulation capacitance and peak slope values are normalised to their maximum values. The metal layers are in nanometers and are presented in the order of deposition.

It can be also noted from the summary table that a negative flat-band voltage shift is observed every time Au is added to the gate metal, or the thickness of Au in the Pt/ Au stack is increased. A possible explanation to this could be the group electronegativity concept as applied to the effective workfunctions of metal gate electrodes on high-κ gate oxides [67, 68]. Pauling electronegativity values for Pt, Au and Ni are 2.28, 2.54 and 1.91, respectively. An addition of a more electro-positive Au to the Pt layer strengthens the local dipoles generated on the GGO surface, reducing the effective workfunction of the gate and therefore causing a reduction of V_{FB}. The opposite is true for the addition of Ni to Pt layer. In Ref. [68], two multi-component gate electrode materials with 0.6 electronegativity difference exhibit an up to 0.4 eV increase in the effective workfunction.

Based on the surface inversion capacitor results, reducing the deposition rate of the bottom metal in the gate proved to achieve the most significant positive $\Delta V_{FB.}$ In this case nominally identical gate stacks were compared, thus the results are not related

to changes in the effective workfunction. The improvement could be due to an improved uniformity of the deposited layer, as well as the slower metallisation process being less damaging. This will be explored in the next section that looks into post-metallisation annealing.

3.4.3 Charge response after thermal treatment

As mentioned in the background chapter of this thesis, post-oxidation and post-metallisation annealing are commonly used techniques to reduce trapping of charge in MOS stacks. Here, the effects of post-metallisation anneal are explored by subjecting samples to thermal treatment in either inert (N_2) or reactive (O_2) gaseous atmosphere with varying temperatures and durations. The aim was to achieve reduction of growth-related and metallisation-induced defects by means of one anneal only. In all experiments described below, the annealing was performed in an RTA (rapid thermal anneal) tool immediately after the deposition of 250 μm diameter Pt/Au-based contacts through a shadow mask, followed by blanket deposition of Au/Ge/Ni/Au back contacts. This ensured only the gate metal was exposed to the thermal treatment.

3.4.3.1 Annealing in oxygen atmosphere

In the previous section it was shown that the thickness of Pt in 100 nm Pt/Au gate has an effect on the MOS high-frequency C-V response. Here, 100 nm Pt/Au gates with Pt thicknesses varying between 1 nm and 10 nm were annealed for 60 s in O_2 at 430°C. The resulting C-V response is shown in figure 3.19. Despite the differences observed between the as-deposited samples with minimum and maximum Pt thicknesses, after thermal treatment their C-V curves overlap, as can be seen in figure 3.19(c). This effect is confirmed by including 3 nm and 5 nm Pt data and observing the same result (figure 3.19(d)). It can therefore be concluded that the effect of thermal treatment on the gate stack is such, that the ratio of metals

composing the gate stack is an irrelevant factor to the quality of the interface in a process where a post-gate deposition annealing step is employed.

As discussed earlier, when 1 nm Pt was deposited at a slower rate (0.05 nm/s), it produced a very different high-frequency C-V response, the most significant change being the ΔV_{FB} of +0.7 V. Possible thin Pt layer uniformity issues may have been the reason for that, improved by reducing the rate and consequently increasing the time of the deposition. These differences were annihilated by the 430°C O_2 treatment, as shown in figure 3.20, which presents the 1 nm Pt deposited at 0.05 nm/s and 1 nm - 10 nm deposited at 0.3 nm/s after the anneal. However, the post-anneal slow-rate C-V showed an increased maximum slope (figure 3.20 (b)). If considering the anneal as a method to recover the damage caused to the material during the metallisation, as discussed in chapter *Background*, the independence of the fast-rate Pt thickness on the post-anneal results versus an improvement of the slow-rate Pt after the anneal, could suggest that slow deposition is less damaging to the GGO/GaAs stack. This has been previously reported by Chen *et al.* [50], who found GaAs-based heterostructures producing better optical response when the rate of Ti contacts deposited on them was slowed down. From the data in figures 3.19 and 3.20, it can be concluded that 430°C O_2 post-metallisation annealing is a useful step to be included in gate fabrication process, and as long as the total thickness of the contact and the metal deposition rates are kept constant, Pt fraction can be changed from 0.5 to 10% with negligible effect on gate stack performance if the annealing step is used.

The next set of experiments looked at MOS capacitors where the metal stack consisted of Pt only, before and after thermal treatment in oxygen atmosphere. The C-V curves in figure 3.21 display a reduction in stretch-out, an increase in maximum capacitance, and a positive shift of the flatband voltage after annealing, similar to the outcome of the Pt/Au contact annealing (figure 3.19(a)).

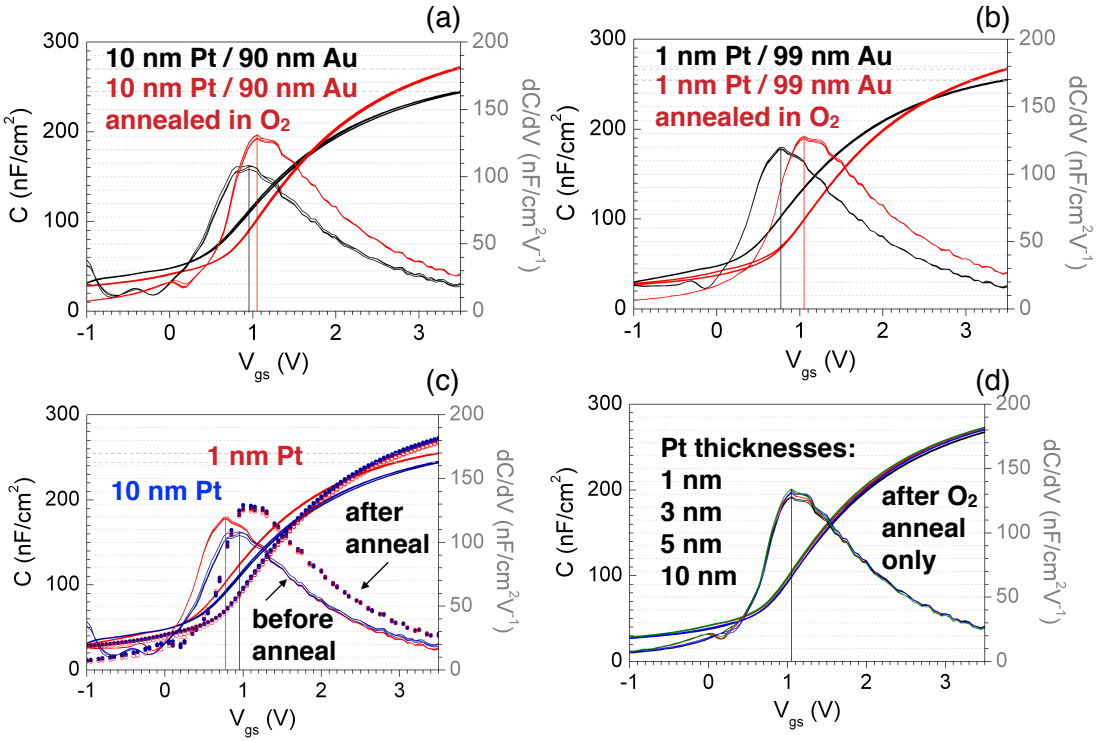


Figure 3.19 High frequency (1 MHz) single-sweep C-V characteristics of bulk n-GaAs structure with 31 nm GGO showing the effect of 60 s in 430°C O_2 RTA on Pt/Au gates with varying Pt thickness. The response of 10 nm Pt (a) and 1 nm Pt (b) before and after annealing are shown separately (a, b) and superimposed (c). Post-anneal responses of all four samples with Pt thickness varying from 1 nm to 10 nm are shown in (d).

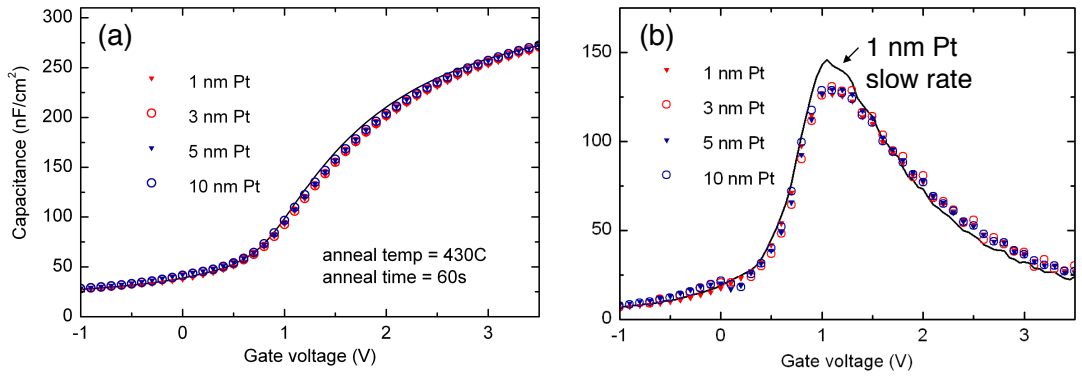


Figure 3.20 High frequency (1 MHz) single-sweep C-V characteristics (a) and dC-dV (b) of bulk n-GaAs with 31 nm GGO showing the effect of annealing on 100 nm Pt/Au gate stacks with different Pt thicknesses showing 1 nm Pt deposited at 0.3 nm/s (scatter) and 0.05 nm/s (solid line) rate.

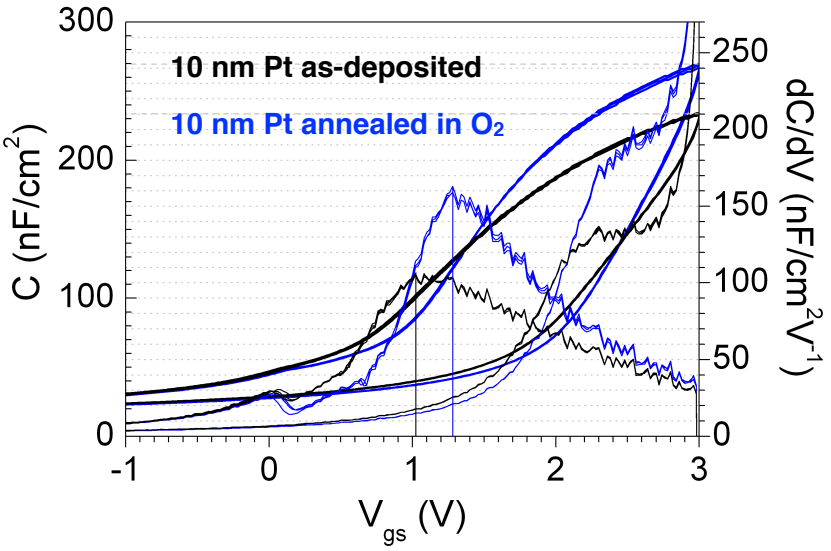


Figure 3.21 High frequency (1 MHz) double-sweep C-V characteristics of bulk n-GaAs structure with 25 nm GGO and 10 nm Pt before and after 60 s in 430°C O_2 RTA.

The following paragraphs will look at the effect on MOS capacitors of RTA steps at different temperatures and/or for shorter times, and also assess if the benefits of RTA are exclusive to a process in oxygen atmosphere or if they can be obtained also with use of inert gases.

3.4.3.2 Effect of anneal time and temperature

All annealing processes in the experiments described so far were conducted at 430°C for 60 seconds. These parameters were chosen for a simple reason: they were identical to the nitrogen anneal that was used in the MOSFET process flow after ohmic contact metal deposition, which could provide a possibility to combine the two. The test described next was done after it was postulated that the 430°C temperature is optimised for a nitrogen process, but could be causing damage in combination with the reactive oxygen. The annealing process optimisation involved reducing the temperature to 300°C and also reducing the initially chosen duration of the 430°C anneal, and measuring the effect of it on Pt (10 nm) / Au (90 nm) gate MOS capacitors. The C-V curves plotted in figure 3.22 indeed confirmed that a lower temperature was better for this process with an increase in the maximum slope observed as the temperature decreased. It can be also seen that cutting the

annealing time in half greatly improved the C-V of the 430°C anneal. From the V_{GMS} values analysis, higher values are achieved for increasing temperatures. In combination with the previous observations, it could be concluded that higher temperatures help with reducing fixed positive charge in the oxide, but are more damaging to the semiconductor. For this reason, the optimised O_2 RTA process was 30 s at 350°C.

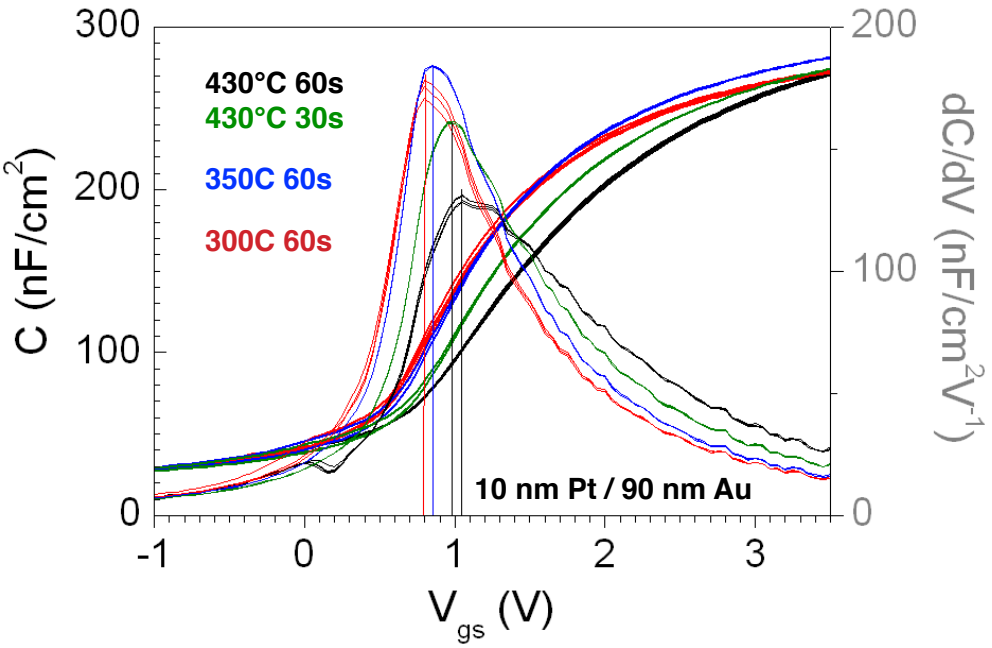


Figure 3.22 High frequency (1 MHz) single-sweep C-V characteristics of bulk n-GaAs with 25 nm GdGaO/GaO dielectric stack where the structure was annealed in O_2 for different times and temperatures after the Pt/Au gate metal deposition.

3.4.3.3 Annealing in nitrogen atmosphere

In order to assess if the beneficial effects of the thermal treatment were dependent on the gas used during the RTA step, samples deposited with a thin layer of Pt were annealed at 350°C for 30 seconds either in nitrogen or oxygen atmosphere. The results, displayed in figure 3.23, show superimposing C-V curves, albeit previously reported differences between 350°C Ar- and O_2-based treatments of 10 nm Pt gates, where O_2 caused 0.5 V V_{FB} shifts. It is possible that not enough energy was supplied

to break molecular O_2 and N_2 into atoms so that a reaction with the defects in the dielectric can occur. In combination with the data shown in figure 3.19, where C-V curves overlap after anneal for different Pt/Au ratios used in the gate metal, it is possible that the changes in the C-V characteristics that are observed after 350-430°C annealing are not caused by changes in the dielectric film, but result from intermixing of gate metals and subsequent equalisation of workfunction value at the metal-oxide interface.

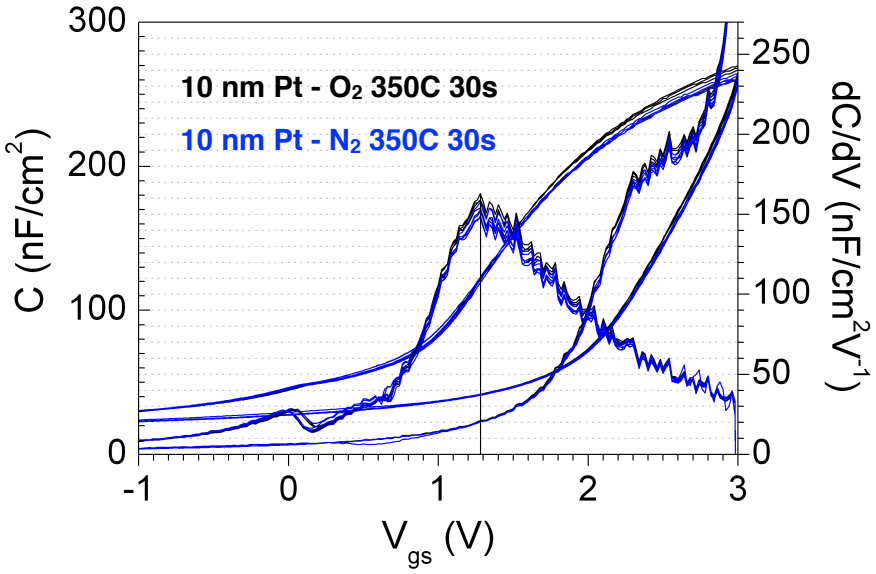


Figure 3.23 High frequency (1 MHz) double-sweep C-V characteristics of bulk n-GaAs structures with 25 nm GGO and 10 nm Pt comparing O_2 and N_2 as annealing atmosphere.

3.4.3.4 Investigation of the effects of the two RTA steps required for the full device fabrication

It is useful to perform MOSCAP measurements on a layer structure specifically designed to de-couple the MOS layers (the gate stack in a MOSFET device) from the rest of the complex epitaxial structure needed for a high-mobility transistor device. This simplified structure allows the use of simple fabrication processes (shadow masks, etc.) in order to understand how the fabrication will affect the control of

charges on the surface of the semiconductor. So far, we have examined the effect of thermal treatment after deposition of the gate metal. However, due to the requirements of process integration the fabrication of the actual device will imply another thermal step to occur in order to form a low resistivity ohmic contact to the source and drain regions. It was therefore of importance to determine the effect of multiple thermal processes on the gate stack. In order to carry out this assessment, the *vertical heterostructure* samples were employed containing the full device layers. Since the substrate was semi-insulating, both contacts of the MOS capacitor were defined on top of the sample, with the top contact being the gate metal stack and the bottom contact consisting of the source and drain ohmic metals, defined on an area where the gate dielectric was etched. All metal contacts were defined by electron-beam lithography and lift-off technique. The fabricated geometry is depicted in the inset of figure 3.24(b). In both cases the gate metal was deposited first and underwent a post-metallisation RTA step at 350°C for 30 seconds in oxygen atmosphere, while the ohmic contacts were annealed in nitrogen atmosphere as standard.

The layer composition of the Ni/Ge/Au contacts are slightly different between 3.24(a) and 3.24(b) with the Ge20/Ni15/Au40 contact having a reduced lateral alloying effect (see section 5.3.2.2). The data in 3.24(a) were obtained during annealing conditions tests for the Ge20/Ni15/Au40 contact and show a particularly interesting result. The 30°C change in annealing temperature produced a drastic effect on the C-V characteristics, mainly expressed in a much greater value of C_{acc} for the higher temperature, with the rest of the C-V curve parameters remaining generally unchanged between the two cases. This result can be explained by the bottom contact of the capacitor having lower resistivity for a higher annealing temperature, with at least 430°C needed for a good ohmic contact. At lower annealing temperature the ohmic contact has a residual parasitic capacitance which affects the measured value of C_{acc}.

Figure 3.24(b) depicts the effects on the C-V curve of an added annealing step in oxygen atmosphere after both the gate and the ohmic metal annealing were carried out. The introduction of an additional annealing step in oxygen gas does not seem

to affect the measured curves, and it can be concluded that the benefits of annealing the device after metallisation do not increase with the number of anneals performed - at least at 350°C it does not. Comparing the two curves in 3.24(a) with the blue curve in 3.24(b), the dominance of annealing temperature over annealing duration is evident.

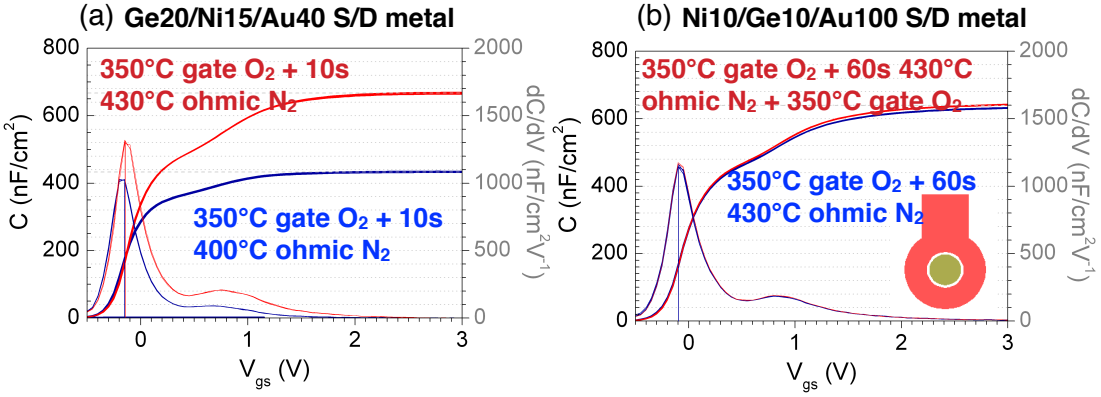


Figure 3.24 High frequency (1 MHz) single-sweep C-V characteristics of GaAs-based heterostructure with 6 nm GGO dielectric comparing the effect of the two RTA steps on gate and ohmic metal required for a complete gate-first MOSFET process with a gate post-metallisation anneal step: (a) Ge20Ni15Au40 annealed at 430°C and 400°C; (b) Ni10Ge10Au100 annealed at 430°C with an added post-ohmic gate anneal.

3.4.3.5 Section Summary

The conclusions drawn from the results of this section are based on a) all anneals at 430°C improving C-V, regardless of the composition of the gate stack, b) O_2 and N_2 anneals at 350°C displaying no difference, albeit previous reports of varying annealing gases, c) similarities between a two-O_2-anneal one-N_2-anneal and a one-O_2-anneal one-N_2-anneal process flows. First of all, it can be concluded that the 350°C temperature level fails to provide enough energy to break up O_2 and N_2 atoms and affect the substrate in any way. When the sample is heated to 430°C the workfunctions of the different metal stacks all collapse at a similar value, thus

yielding a conduction band offset in the semiconductor and C-V results with little variation, yet before the annealing this is not the case. The possibility of Pt and Au metal layers intermixing during the heating is unlikely because platinum is known as a good diffusion barrier. One possible reason for the observed phenomenon could be the difference in the duration of the deposition of each metal layer. The extent of damage caused to the substrate will vary with layer thicknesses as the currents required for the deposition will differ. This will be reflected in the C-V response of as-deposited capacitors. During the heating, the metallisation-induced defects are annealed out, and the C-V curves are merged. In the dielectric layers, the damage recovery could be due to self-rearranging of atoms taking place as a result of thermal energy supplied by the annealing process. However, in the event of metallisation-induced damage being insensitive to small variations in beam currents, the explanation could be that the metal intermixing does occur when the high diffusion coefficient of gold overpowers the high diffusion barrier properties of platinum, especially for thin Pt layers, making the contact a more ideal conductor. In this case, the initial differences in C-V are because of the gate effective workfunction being tuned by the electronegativity properties of the metals, as discussed earlier.

4 Gate fabrication challenges and the effect of resist residues and post-metallisation annealing on device performance

The device used in this work for process optimisation experiments was based on a layer structure developed at Freescale Semiconductor by M. Passlack *et al.* [9]. It was grown by M. Holland in the West of Scotland Science Park and the fabrication process was developed in the University of Glasgow by R. Hill, D. Moran and I. Thayne [10, 11]. Given a brief overview of design and operation of a normally-off buried quantum-well channel n-type heterostructure MOSFET was provided in chapter *Background*, while the explanation of its layer structure in chapter 3, this chapter starts with a description of the device layout and fabrication process.

4.1 Device layout and fabrication process

A MOSFET is a three-terminal device where ohmic contacts are required for the source and drain regions and a rectifying contact is required on the gate region between the source and the drain. In a CMOS chip the ultra-high density of transistors is achieved by stacking metal contact layers upwards, isolated by insulators. The contacts are narrow lines shared between multiple devices. The effect of a process variation on device performance or reliability needs to be tested on a complete device, and this requires probing each device separately. Because of this, a simple planar layout which enables rapid process turn-around and access to individual devices is much more suitable.

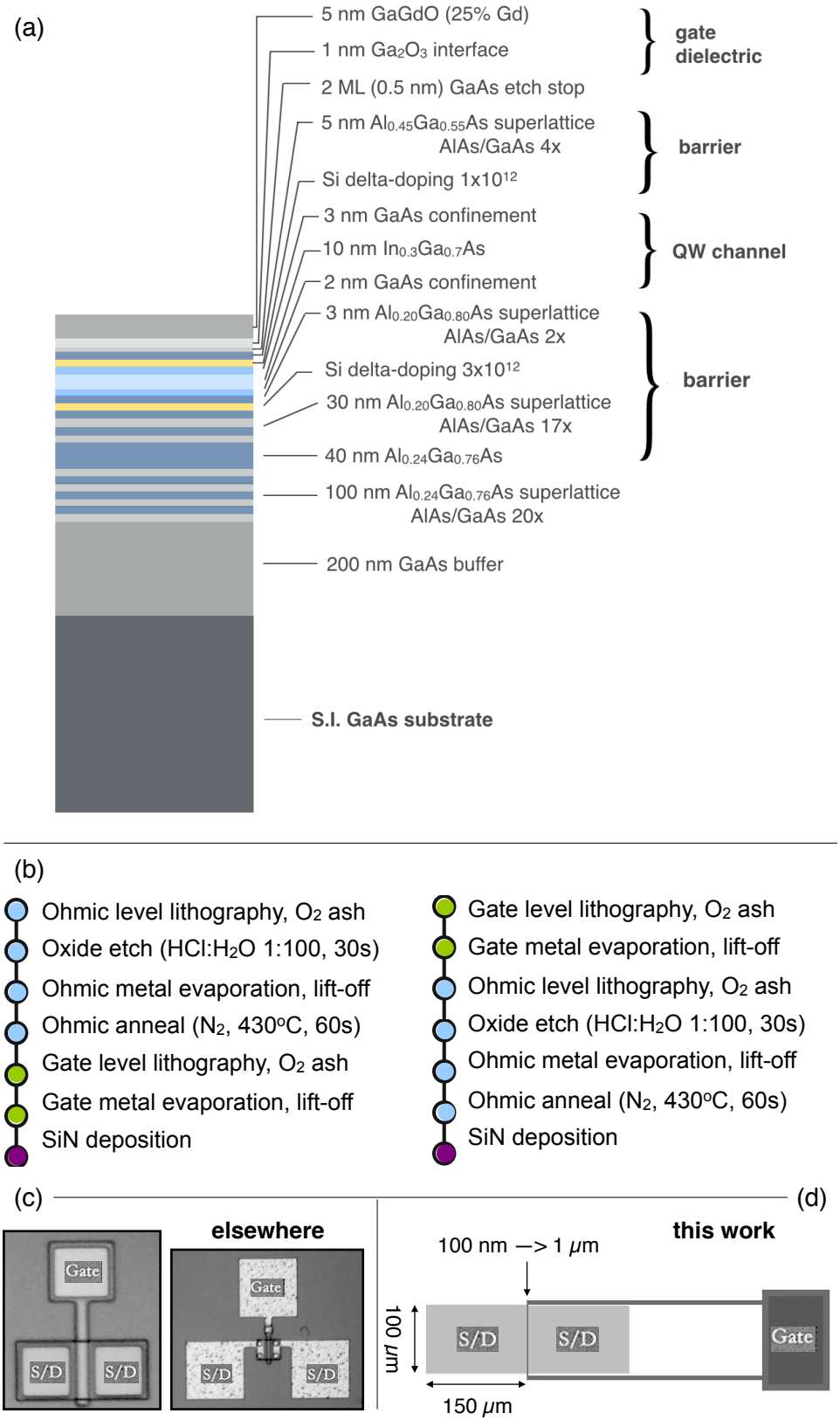


Figure 4.1 (a) III-V device layers grown by molecular-beam epitaxy. (b) Process steps for gate-first and gate-last fabrication flow. (c) Examples of test MOSFET contacts layout with isolation. (d) Design used in this work, where no isolation was required.

For this reason, layout for a test MOSFET normally has probe pads for gate and source/drain contacts on the same plane, and the contacts are normally just continued into a larger area probe pad. Such a device would inevitably use up a larger area, and normally, test patterns for devices consist of arrays of isolated devices, each having their own probe pads for the three terminals. Figure 4.1(c) shows two different versions of test device layout, varying in gate width and active region isolation size, where the active region either extends over the source/drain probing pads or covers the active device area only. The layout used in this work, shown in figure 4.1(d), required no isolation as the gate contact wrapped around the drain, isolating it from the source.

The key dimensions were:

- Source/drain contact width = 150 μm
- Source/drain contact height = 100 μm
- Gate length L_g = 100 nm to 1 μm
- Distance between gate and ohmic contact $L_{g/o}$ = 1 μm

The fabrication process of the rapid feedback device only had two layers of lithography: the gate level and the ohmic level. When the gate contact was patterned first, the process was called *gate-first*, whereas *gate-last* process had source/drain ohmic contacts patterned first. All patterns were written by electron-beam lithography, using metal markers for level-to-level alignment. The metal layers were lifted-off rather than etched, which was better for surface damage. The default gate metal was 20/200 nm Pt/Au, and the metal used for the source/drain ohmic contacts was always 20/20/200 nm Ni/Ge/Au. Since both of these types of metals were equally effective as alignment markers metal, providing enough contrast between the metal and the substrate surfaces, it was possible to combine the markers lithography level with either the gate or the ohmic level, thus keeping the fabrication process to the maximum of two lithography levels at all times. The baseline fabrication flows for gate-first and gate-last devices are shown in figure 4.1(b). The last step is blanket Si_3N_4 deposition necessary to passivate the surface between the gate and the source/drain contacts, i.e. the access region.

4.2 Gate contact patterning issues

All metal contacts in the device shown in figure 4.1(d) were fabricated by lift-off technique. In lift-off, the features are first patterned in a resist (a polymer layer sensitive to radiation), then a sheet of metal is blanket deposited over the resist pattern, followed by removal of the metal outside contacts area by dissolving the resist underneath the metal in those areas in acetone. Electron-beam lithography was used for resist exposure, due to strict alignment requirements regarding placement of the gate pattern in the centre of the gap between the source and drain contacts.

Two issues are immediately apparent with fabrication by e-beam lithography and lift-off: a) surface contamination by polymer associated with resist processing, and b) proximity effect associated with electron-beam exposure that affects dense patterns. These were explored, and the results are presented below.

4.2.1 E-beam resist residue in the gate region

A very common and easily available e-beam resist suitable for lift-off process is polymethyl methacrylate (PMMA). Residues left on semiconductor surfaces after e-beam exposure and development of PMMA became a popular research topic in early 2000s when the cleanliness of the semiconductor surface became important to achieving good semiconductor-dielectric and dielectric-metal interfaces. A few groups studied Si and SiO_2 surfaces after PMMA development in the standard IPA-MIBK solution [69, 70], and found residual resist to be a 0.5 to 1 nm layer covered with granular island structures that were highly dependent on the exposure dose. The granular structures were well visible in an SEM (scanning electron microscope), but the layer underneath had to be measured with photoelectron spectroscopy. Here, a method was developed to measure the residual PMMA layer that employed two electron beam exposures and a subsequent step height measurement to quantify PMMA residues in the gate region, demonstrated on both Si and GaGdO surfaces. As a preliminary step, the clearing exposure dose for PMMA resist was

determined. In order to carry out this step, a common method consists of defining a series of rectangular patterns subject to increasing electron dose. Each rectangle must be spaced at least 200 μm from the next one in order to avoid being additionally exposed with the electrons backscattered during the neighbouring pattern exposure. After resist exposure and development, the clearing dose was determined by means of a profilometer scan of the pattern, as shown in figure 4.2(a). At the clearing dose, surface scans using an atomic-force microscope (AFM) in tapping measurement mode showed the granularity illustrated in figure 4.2(b).

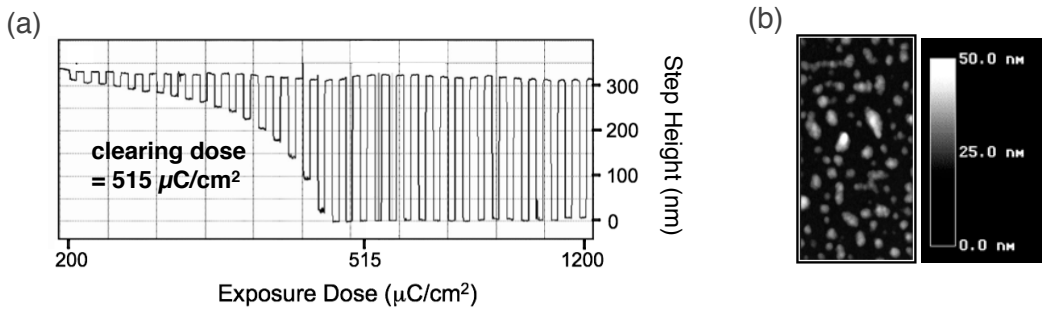


Figure 4.2 Residual PMMA on Si surface at clearing dose. (a) Exposure test for identification of the clearing dose using vertical profile analysis of an array of 100 x 50 μm^2 rectangles exposed with 200 - 1200 μC/cm^2 electron dose at 100 keV in 300 nm 120K PMMA and developed in IPA-MIBK solution, measured with a surface profiler; b) AFM scan of 3 x 1.5 μm^2 area of the surface exposed to clearing dose after development.

The surface granularity dependence on electron beam exposure dose and molecular weight of the resist was assessed, and the resulting AFM surface scans are shown in figures 4.3(a) and 4.3(b), for PMMA with molecular weight 120K and 495K, respectively. The size of granules at clearing dose was found to be larger for the lower molecular weight resist. The granules became more scattered and diminished in size with increase in exposure dose and with increase in molecular weight. The amount of overexposure required to remove the granules from the surface varied with the molecular weight: with 90% increase required to clear 495K PMMA and 53% increase to clear 120K PMMA.

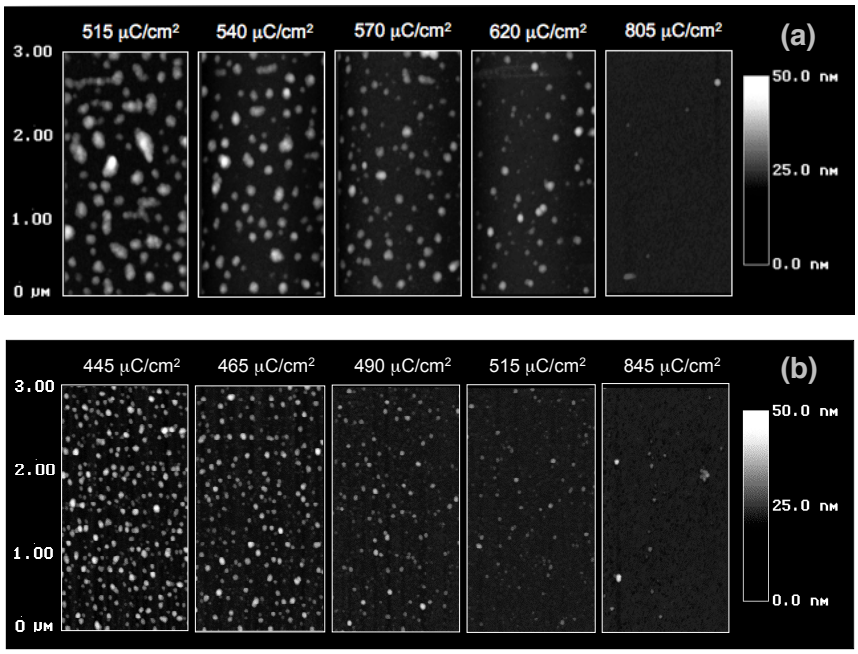


Figure 4.3 Difference in granularity of residual PMMA on Si surface after exposure and development of (a) 120K molecular weight PMMA and (b) 495K molecular weight PMMA. The dose range shown starts from the respective clearing doses.

Albeit the disappearance of the granularity at high electron beam doses, a thin layer of residual PMMA was reported under the granules, which, as previously stated, could only be measured with photoelectron spectroscopy. Here the presence of such layer was assessed with a different method, described in the following. First a mask consisting of PMMA gratings was defined on the sample surface by means of e-beam exposure, resist development and oxygen ashing, this last step to guarantee a residue-free reference surface to act as zero level. Without removal of the developed PMMA, the sample underwent a second electron beam lithography step, patterning both previously unexposed areas and some exposed ones. After the second development step but this time without any oxygen ashing, profile scans were taken across surfaces containing both previously cleaned and residue-free areas and freshly developed ones, with the step measured indicating PMMA residue. The process is schematically illustrated in figure 4.4. The effectiveness of oxygen plasma in providing residue-free surfaces is well established. It is a standard process used to remove resist residues prior to etching or metallisation [71, 72].

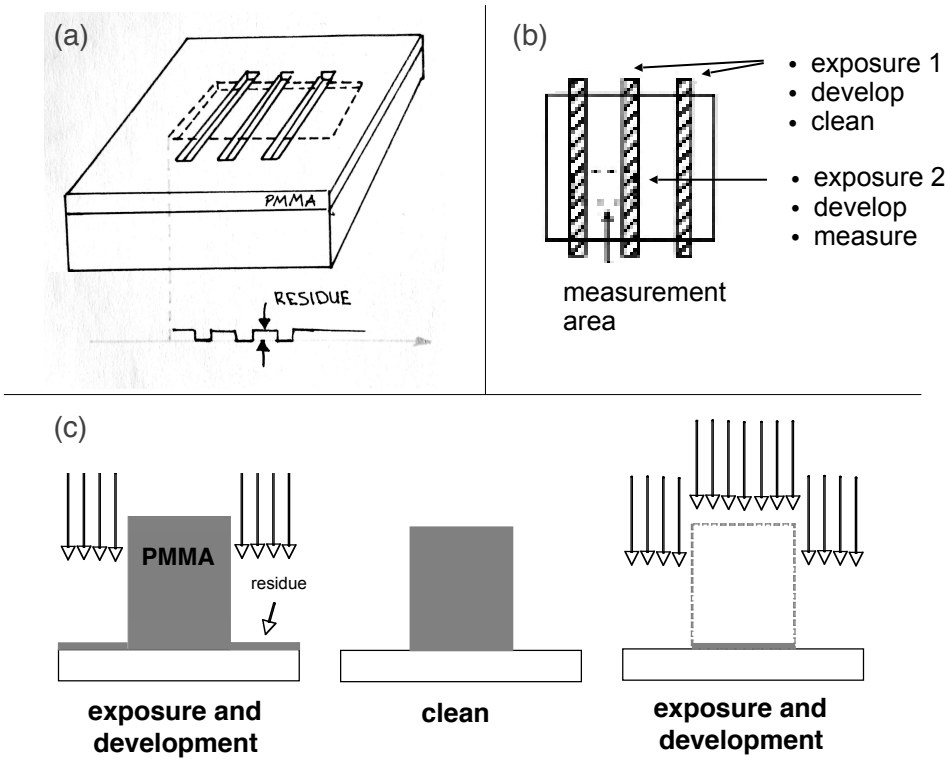


Figure 4.4 Schematic diagrams of the 'two-exposures' technique to assess post-EBL process resist residue on the substrate surface: (a) and (b) show the final structure highlighting the AFM scan area, where vertical lines are written in PMMA first and cleaned with O_2 plasma to achieve a residue-free surface at the bottom of the trenches, followed by a second e-beam exposure and development where a rectangle is written covering the lines without any cleaning step, as shown in (c); the step height difference between the 'residue-free' lines and the space between the lines is the thickness of the residual PMMA film.

The advantage of the two-exposures method is that it provides simultaneous measurement of residual layer thickness and its surface roughness measurement, as figure 4.5 shows. All measurements presented here were based upon patterning a single layer of PMMA on either silicon or GaGdO surface. On silicon, a residual PMMA thin film of approximately 2.5 nm hosting a granular structure of 10-15 nm islands was observed. On GaGdO surface, the film was approximately 1 nm (scans not shown), but the granular structure was observed to be independent of dose.

An important finding regarding surface cleanliness post PMMA exposure was revealed by this technique. Upon inspection of the post-EBL surface in a microscope, the surface may appear clean if a sufficient enough exposure dose was applied for clearing the granules, and give a false impression, as there is always a thin resist layer existing on the surface, and its thickness varies from 1 to 3 nm if the underlying surface is a dielectric or a semiconductor, respectively. A cleaning step, such as a short low power oxygen plasma ash (somewhere between 40W and 80W, depending on the system), should therefore always be included after development to ensure the surface is clean and no reaction with the surface layers will occur during the following processing steps, e.g. in the case of a thermal annealing. However, oxygen ashing enlarges the size of the patterned features, thus its impact can be detrimental for small dense patterns.

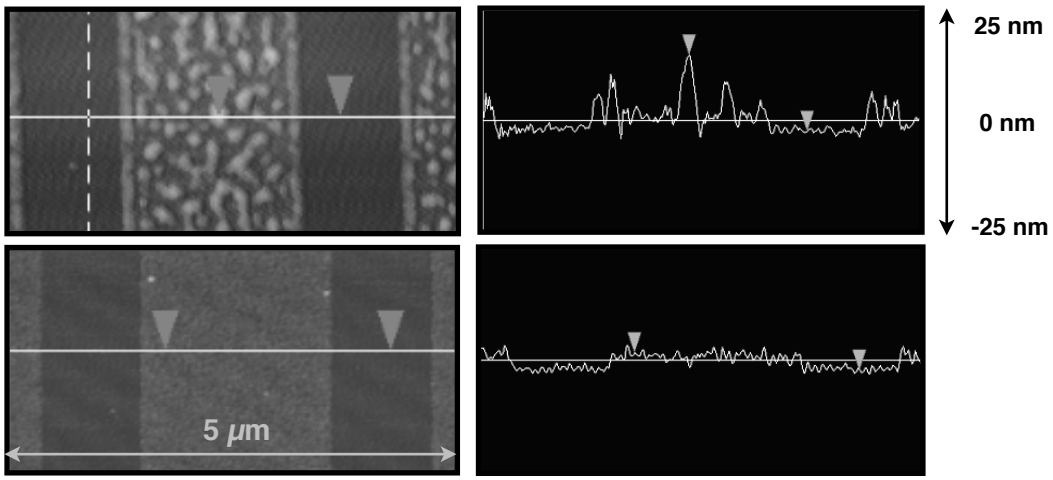


Figure 4.5 Step height measurements of the PMMA residual layer on Si surface: surface scans of measured areas and corresponding height profiles for exposure doses of 350 $\mu C/cm^2$ (underexposed) and 800 $\mu C/cm^2$ (overexposed). The film is 2.5 nm thick and the height of the granules is 10-15 nm.

The usefulness of a technique to measure post-development residual layer might not be obvious, since it can be easily removed by oxygen ashing, a widespread post-gate lithography practice in the gate region when using e-beam resist. However, knowing exactly the thickness of the residual layer is important when there is a concern about oxygen plasma damaging the underlying material, as is often the case with III-V materials, susceptible to plasma damage. Since the nature of the

residue determines the power and the duration of the ash, this technique allows to choose the minimum required settings.

4.2.2 Non-uniformity of gate resist and gate metal linewidth dependance on e-beam exposure parameters

In the baseline gate-last process, the source and drain contacts are defined first, consisting of 240 nm thick Ni/Ge/Au layers, 2-3 μm apart depending on the gate length. As illustrated in figure 4.6(d), this causes a sample non-planarity, which will result in thicker resist within the small gap between the source and drain contacts when the gate resist (320 nm film) is spin-coated on the sample. Resist thickness variation along the source/drain gap was measured by spinning 140 nm PMMA over 150 nm thick source/drain contacts with 300 nm, 540 nm, and 1000 nm gaps. A 500 nm wide PMMA line was written in PMMA across the source and drain gap (shown in figure 4.6(c) as yellow lines), so that resist thicknesses in different points along the length of the gap could be extracted from an AFM surface scan (shown in figure 4.6(c) as yellow boxes). Proximity effect correction was used for all lithography. From figure 4.6(a), a 1:2 ratio was measured between open areas and the smallest source and drain gap, highlighting a possible issue when defining the gate. In order to produce reliable gate contacts the electron exposure dose must ensure that the resist will be cleared after development even where it is at its thickest. The need for a constant linewidth defined in the developed resist independently of the film thickness arises because the dimension of the final metal line is determined by the topology of the resist uppermost layer.

An experiment was designed to determine which electron beam dose could provide a certain gate metal linewidth in two different thicknesses of PMMA, approximately corresponding to the maximum and minimum values extracted from the measurement in figure 4.6(b). The results are shown in figure 4.7. The three families of curves were obtained for three digital linewidths, nominally 10, 20 and 80 nm. It can be seen from the measured linewidth values on the y-axis that the minimum dimension achieved was 30 nm independently of the dose or of the designed size.

From all the curves it can be noticed that the dose providing a certain linewidth is approximately the same for the two thicknesses of PMMA tested, especially at high doses.

S/D gap (nm)	on substrate	1000	540	300
Resist thickness (nm)	140	184	208	282

(a)

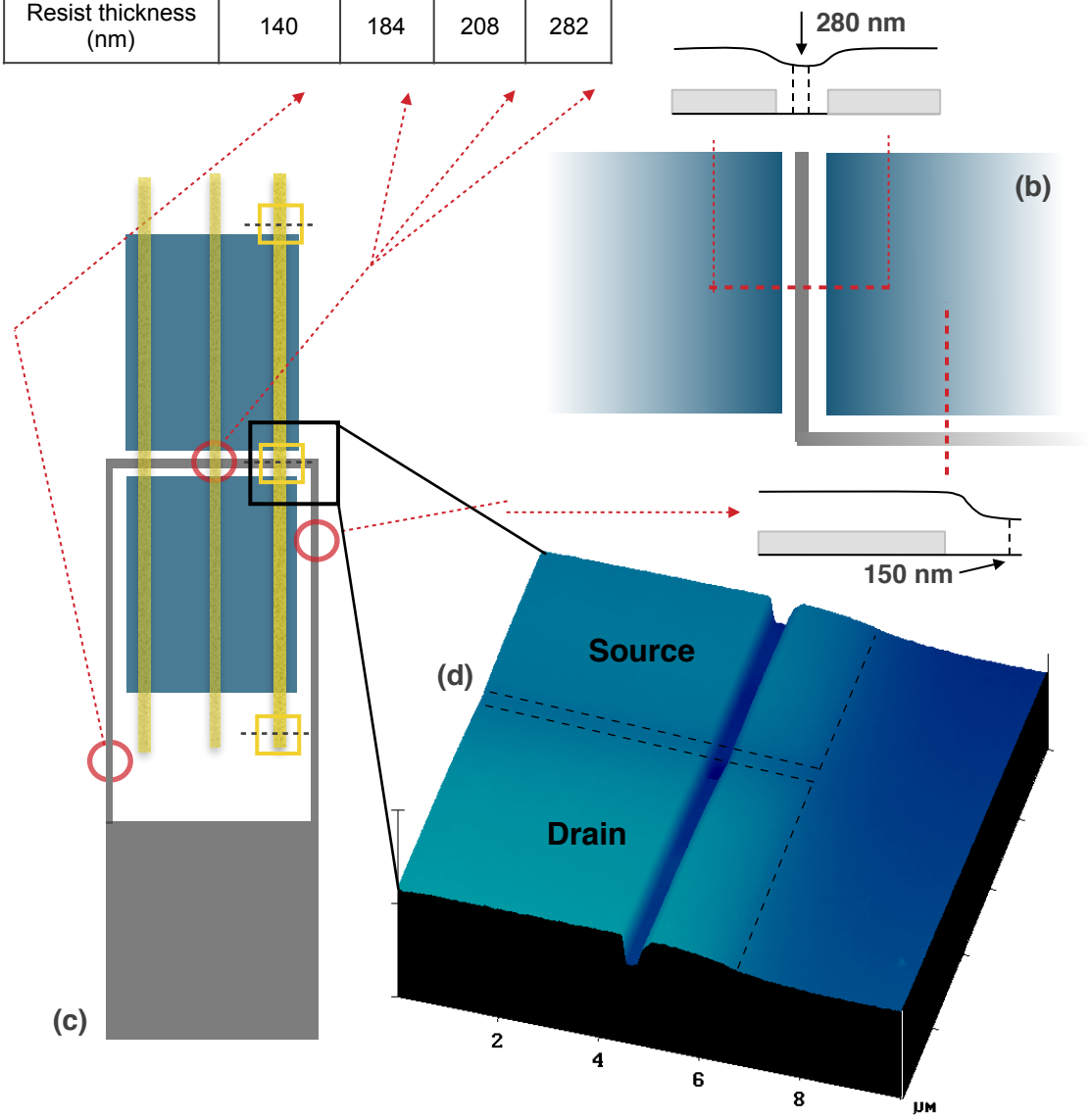


Figure 4.6 Gate resist over metal S/D contacts: (a) table of resist thickness measured by AFM in the gap between the source and drain metal contacts half way through the width of the device as a function of source-drain gap; both the S/D metal thickness and the PMMA thickness were 140 nm; (b) diagram of resist profile for two cross-sections in the S/D region; (c) patterned layout for resist thickness measurements; (d) AFM surface topography showing resist thickness fluctuation at the edge of the two closely-spaced S/D contacts with a trench across the S/D gap to enable resist thickness measurement in the gap.

All data shown in figure 4.7 were obtained with the beam current of 4 nA on a 100 kV column, corresponding to a beam diameter of 9 nm, while the beam step size was 5 nm. The effect of writing the gate metal with different beam currents was also assessed, with figure 4.8 depicting the results for a 20 nm linewidth patterned on a 280 nm PMMA layer. From the various beam diameters compared it can be seen that the smallest ratio between actual and designed linewidth is achieved for the smallest beam at the lowest exposure dose tested, as it could be intuitively expected. If a 64 nA beam (having a diameter of 33 nm) is instead used with a high dose, a linewidth of 60 nm can be achieved by a 10 nm designed value, which can be of use when writing larger gate lengths to reduce the writing time.

From the designed experiments it can be concluded that even with 1:2 resist thicknesses difference between the resist film in open areas and over the small gaps between source and drain, the clearing dose of the thinner PMMA will still be effective in defining the gate line between the ohmic contacts. The change in resist profile over the source and drain metal will not have a detrimental impact on the gate definition. With regard to the actual linewidth defined in the resist in comparison with its nominal value, the smallest which could be achieved was still twice the designed width, but as long as the ratio is known the desired value can be obtained by changes at layout stage.

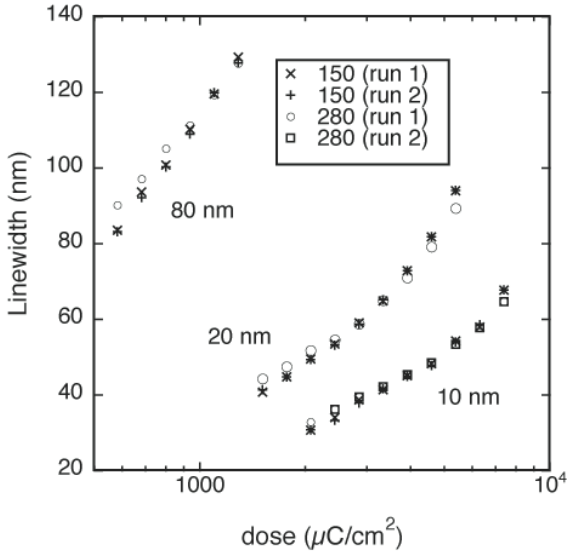


Figure 4.7 Linewidth against exposure dose for PMMA thicknesses 150 and 280 nm written with the beam current of 4 nA, showing results for 10, 20 and 80 nm digital linewidths. Metal - Pt20/Au100.

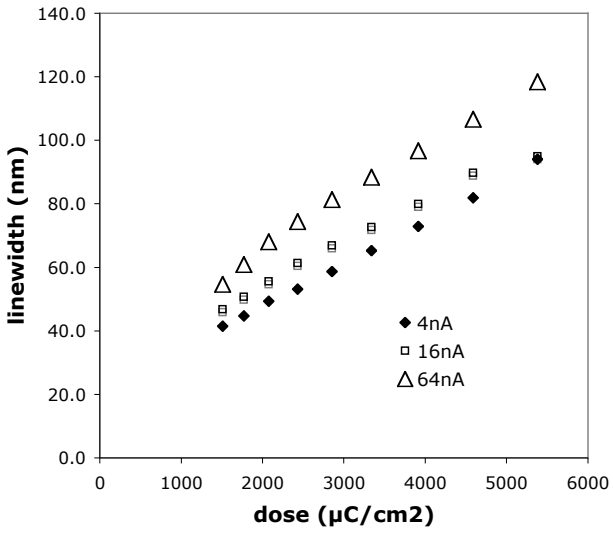


Figure 4.8 Linewidth against exposure dose for PMMA thickness of 280 nm and beam currents 4, 16 and 64 nA, showing results for the 20 nm digital linewidth. Metal - Pt20/Au100.

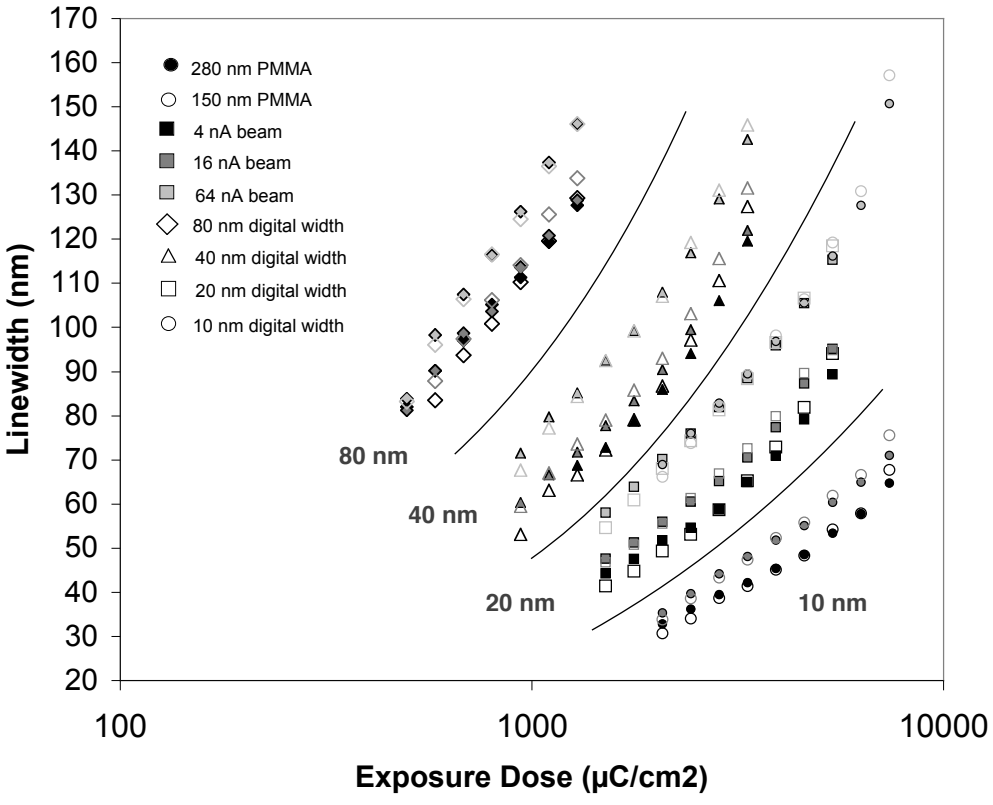


Figure 4.9 Metal linewidth post-lift-off against exposure dose for two PMMA thicknesses exposed at a range of doses and beam currents at 100 kV on a thick layer of n-GaAs. Metal layers are 20 nm Pt and 100 nm Au. The bi-layer PMMA films consisted of a thicker lower molecular weight bottom layer and a thinner higher molecular weight top layer, with concentrations chosen to achieve 150 nm and 280 nm total film thicknesses.

4.3 Effect of gate region contamination and gate metal thermal anneal on transistor characteristics

Of many experimental outcomes and observations deduced in this work so far, two of them were related to negative effects of resist contamination in the gate region underneath the metal and positive effects of post-metallisation annealing. These are explored further in this section by either removing the oxygen ash step before gate metallisation or adding an O_2 anneal step after the gate metallisation for both gate-first and gate-last processed MOSFETs. The step sequence for the gate-first and the gate-last processes is shown in figure 4.10.

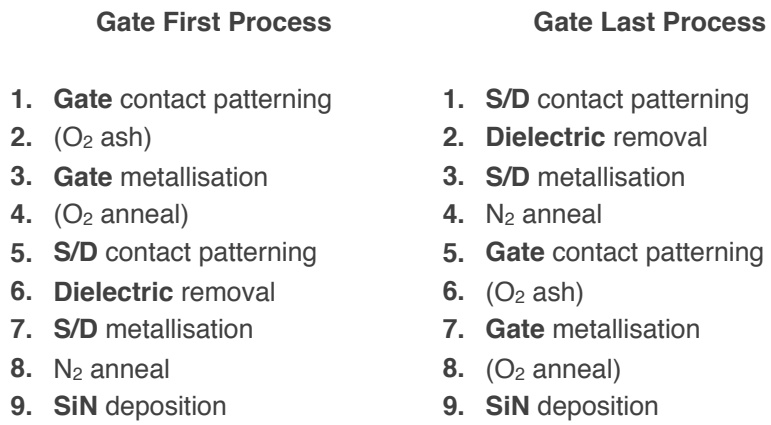


Figure 4.10 Sequence of steps for gate-first and gate-last fabrication process indicating the optional steps in brackets.

4.3.1 Resist residues on dielectric surface in the gate region

In order to assess the effect of resist residues on the gate region, two sets of gate-first devices were fabricated, having identical process flows except for the plasma ashing of the gate region, which was omitted for one of the sets. The ashing step consisted of subjecting the sample to two minutes of oxygen plasma at an RF power of 110 Watts. This power is different from the previously mentioned 40 W as it was

performed on a barrel asher of a different make and model. In both cases, the power setting was chosen to achieve ~5 nm/min resist removal rate. No thermal annealing of the gate contact was carried out for any of the devices that were part of the gate region resist residues study. Figure 4.11 compares the effect of ashing the gate region in terms of drain current I_d and transconductance g_m vs. V_{gs} for two gate lengths of 1 μm and 300 nm and a drain-source voltage V_{ds} of 0.05 V. The spread in measured values between different devices on one sample is negligible for the set of MOSFETs with ashed gate region, whereas it can be clearly seen when oxygen ashing was not applied. The spread in values is due to the random distribution of resist residues on the dielectric surface underlying the gate metal. This is supported also by the fact that MOSFETs with a shorter gate length of 300 nm (figure 4.11(a)) showed less spread in the measured values, because considering a similar residue surface density, smaller areas are more likely to experience similar residues distribution. The same trend was also observed in the 300 nm and 1 μm L_g devices at V_{ds} = 1 V (not shown). With regard to the changes in channel conduction, Figure 4.12 compares the output characteristics of the two sets of devices with or without ashing of the gate region for a gate length of 300 nm (a) and 1 μm (b). Devices where the gate region was not subject to oxygen plasma provided lower off-state and sub-threshold leakage currents, suggesting a more effective gate function. It is likely that the oxygen plasma step caused damage to the thin (~6 nm) gate dielectric, resulting in a poorer gate control of the channel. A trade-off thus emerges between uniformity of MOSFET performance on the wafer requiring resist residues removal and limiting the damage inflicted to the dielectric surface by oxygen plasma.

The maximum output current measured in the saturation region provided different results between ashed and un-ashed gate regions depending on gate length: MOSFETs with 1 μm gate length showed I_d reduction for ashed gate region similar to that observed for off-state and sub-threshold current. The saturation drain current for MOSFETs with 300 nm gate length instead reached approximately 450 μA/μm if the gate region was not ashed, while only 300 μA/μm if the same region was subject to oxygen plasma, a reduction by more than 30%. This suggests some charge trapping for high electric fields applied to the gate electrode, and could be

due to the introduction of fixed charges to the oxide layer during the oxygen plasma step.

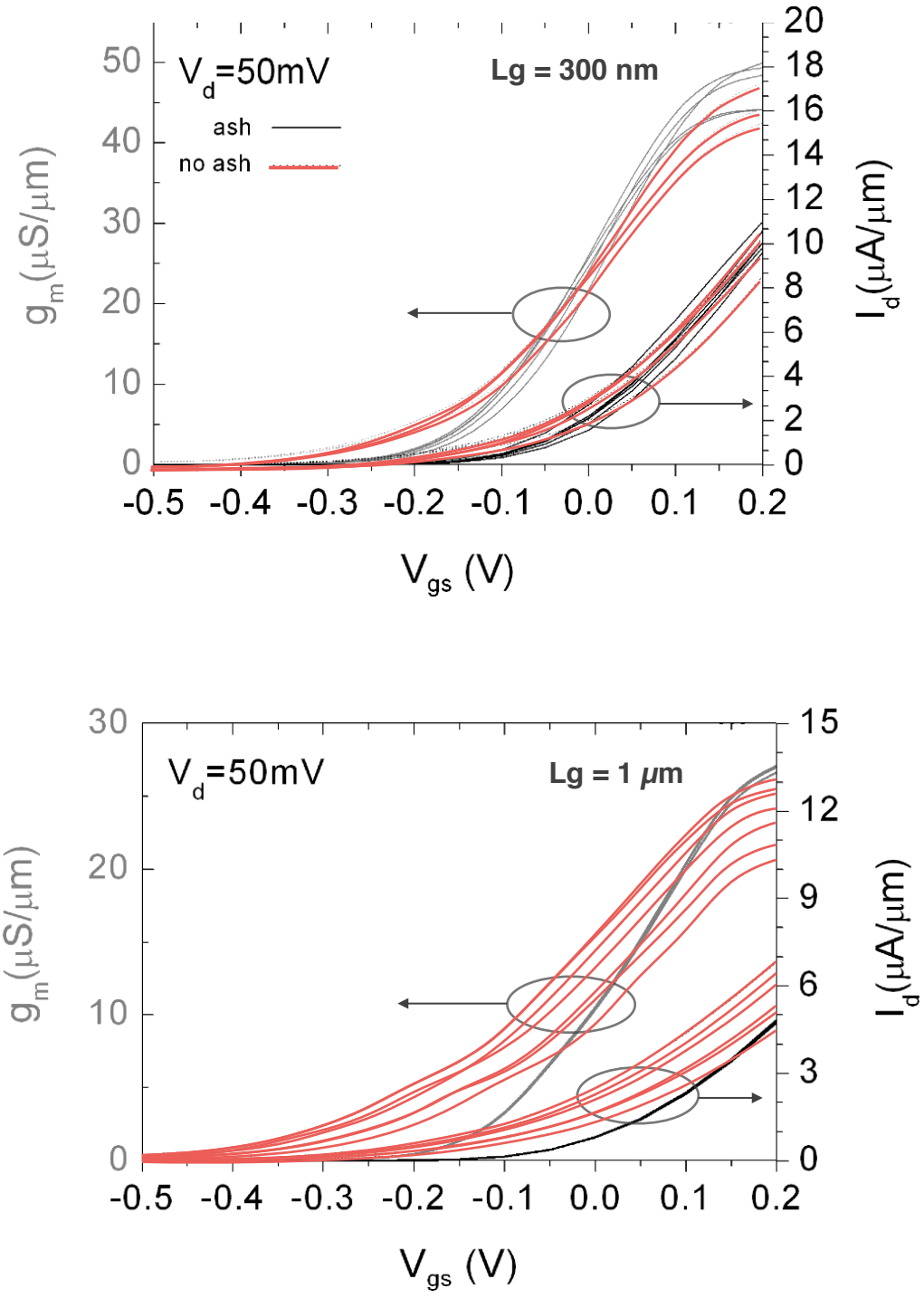


Figure 4.11 I_d and $g_{m,max}$ curves as a function of V_{gs} comparing gate-first MOSFETs with and without gate region O_2 plasma ash for 1 μm gate length and 300 nm gate length devices.

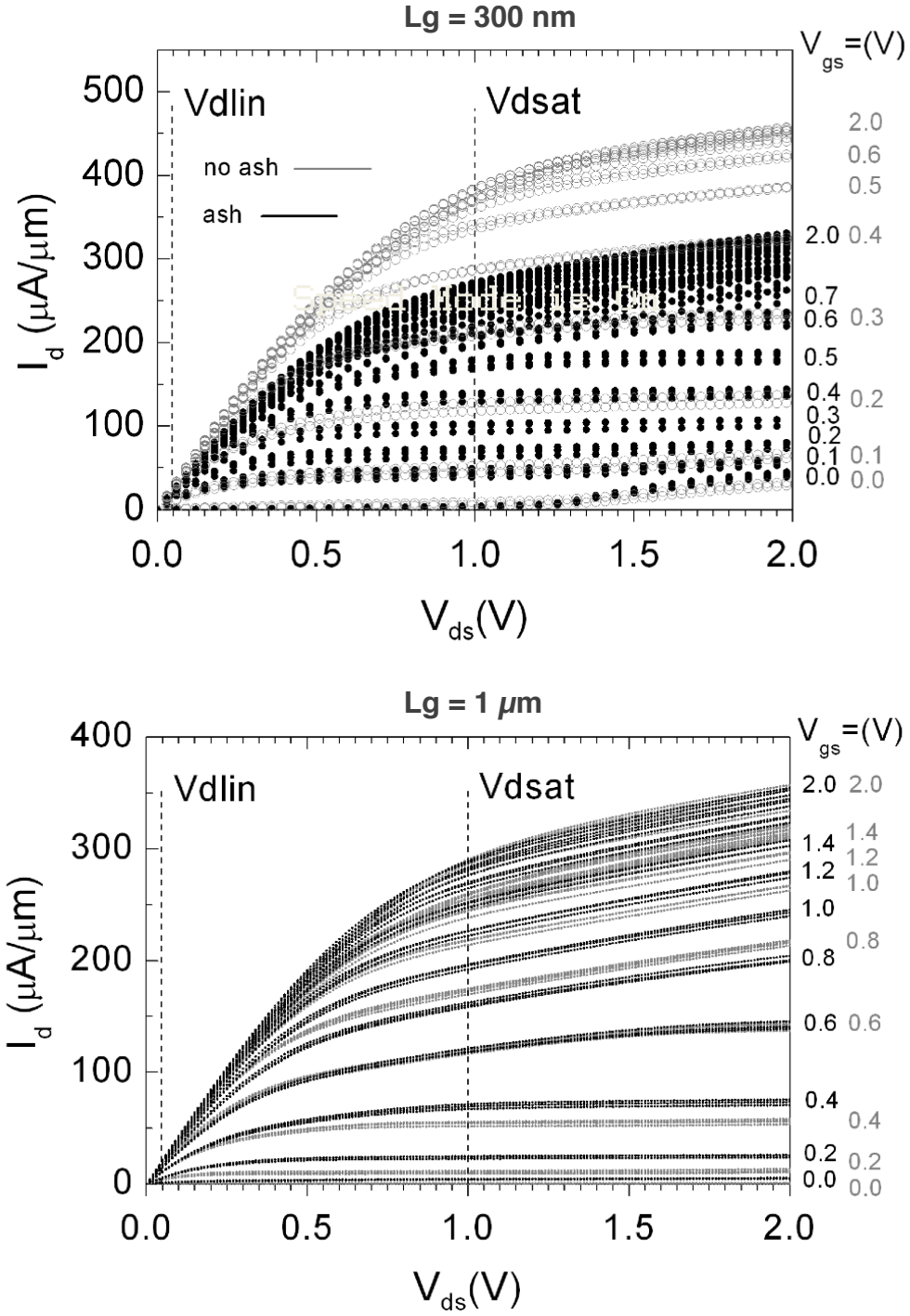


Figure 4.12 I_d vs V_{ds} curves for various V_{gs} comparing gate-first MOSFETs with and without gate region O_2 plasma ash for 1 μm gate length and 300 nm gate length devices.

From the analysed data, it can be concluded that ashing of the gate region provided an improved device uniformity, however at the expenses of a degradation in device performance, more pronounced for sub-micrometer gate lengths. This is in contrast with the findings from similar experiments on MOSCAP material presented in the previous chapter, where an oxygen ashing step with the same parameters was applied without casing any noticeable damage to the dielectric. It must be noted, however, that the dielectric thickness for the MOSCAP material was 20 nm, whereas for the aforementioned MOSFETs the gate dielectric was 6 nm thick, more prone to plasma damage.

For future MOSFETs fabrication, a trade-off between resist residue removal and dielectric damage must be sought, and an assessment of the minimum oxygen plasma RF power still effective for residue removal without damaging the dielectric on the actual device material could be object of future work.

4.3.2 Post gate metallisation thermal anneal

The MOSFETs discussed in the previous paragraph were only subject to a 430°C thermal anneal in nitrogen atmosphere after source/drain contact definition, whereas the gate metal was not annealed. Another set of experiments was carried out to determine the influence of a 350°C gate annealing step in oxygen atmosphere on device performance both for gate-first and gate-last fabrication processes. The effect of thermal anneal of the gate contact on gate-first devices can be seen from the output characteristics of figure 4.13, showing the results measured for two devices with gate length of 1 μm (a) and 300 nm (b). In both cases the introduction of the gate annealing step in oxygen atmosphere slightly degraded device performance, causing a reduction in the output current $I_{d,sat}$, with smaller gate lengths more affected. A 300 nm gate length gate-first MOSFET showed a reduction of approximately 40 μA/μm, whereas a gate-first device with a 1 μm channel had the output current reduced by an average of 25 μA/μm. With regard to the linear increase of the output curve in saturation for high V_{ds}, a decrease was observed in gate-first devices with annealed gate region, meaning MOSFETs less affected by

channel width modulation. Since the slope of the output curve in saturation is entirely process-technology dependent, a more ideal behaviour indicates a better process, suggesting the thermal anneal of the gate contact in a gate-first process is beneficial to device performance despite the slight reduction in output current.

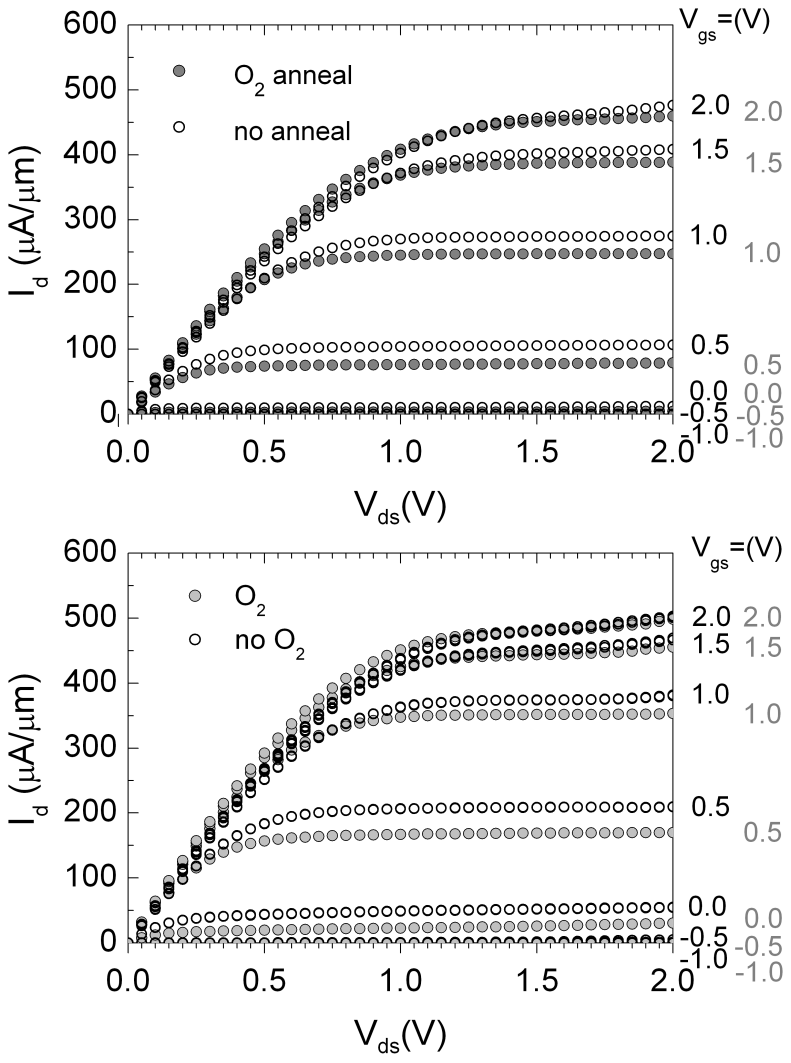


Figure 4.13 I_d vs V_{ds} curves for various V_{gs} for 1 μm (a) and 300 nm (b) gate length MOSFETs showing the effect of thermal anneal of the gate contact on gate-first devices. All traces are average of multiple devices.

This is further confirmed by other parameters extracted from the measurements, shown in table 4.1. Here, the threshold voltage was taken from an $I_d(V_g)$ curve as the value of the gate voltage when the drain current is 1 μA/μm, whereas the subthreshold slope was extracted as the change in the gate voltage per decade of drain current in the linear part of a log $I_d(V_g)$ plot. The values of threshold voltage

were always negative for all fabricated devices, thus indicating possible material growth issues, e.g. variations in delta-doping layer concentration and in layer thickness. The thermal anneal of the gate metal, however, shifted V_{th} towards positive values, both for 1 μm and 300 nm gate length, an improvement for an enhancement-mode transistor. The thermal anneal of the gate provided a beneficial effect also in terms of sub-threshold slope, with smaller values provided by MOSFETs having their gate annealed; this effect was more pronounced for sub-micrometer channel lengths. Summarising, for gate-first devices annealing of the gate metal in oxygen atmosphere generally caused a small decrease in output current. However, it also improved device performance, by causing a positive shift in threshold voltage, a reduction in sub-threshold slope and an output characteristic less affected by channel width modulation. It can thus be concluded that for gate-first devices thermal treatment of the gate metal in oxygen atmosphere is recommended to achieve better MOSFETs.

Gate-first	**L_g = 1 μm**		**L_g = 300 nm**	
Post gate metallisation O_2 anneal	No	Yes	No	Yes
V_{th} [V] (V_g at I_d = 1 μA/μm)	-0.18	-0.02	-0.47	-0.34
$Slope_{min}$ of log(I_d)-V_g [mV/dec]	96	91	94	81

Table 4.1 Summary of sub-threshold characteristics extracted from GF devices showing the effect of annealing the gate metal.

With regard to gate-last devices, a decrease in output current was also observed for MOSFETs whose gate was subject to thermal annealing, as figure 4.14 shows. This decrease was much more significant than that observed for gate-first devices, with an obtained average reduction in current by approximately 40% for MOSFETs with 1 μm channel between not annealed and annealed devices. The impact of gate annealing on device performance was even more dramatic with scaling of the gate length, causing an average reduction in output current by approximately 60% for a 300 nm channel. There are three main reasons for the worsening of device performance in gate-last devices with gate metal subject to annealing in oxygen: 1) Double anneal of the source/drain metals in gate-last, with the second annealing

in oxygen atmosphere, compared to only one nitrogen anneal in gate-first MOSFETs; 2) The source/drain anneal of gate-last transistors happens before gate metal deposition, i.e. the dielectric is unprotected in the gate region; 3) In the gate-last device, resist residues are present in the gate region after source/drain metal lift-off, and are subject to high temperatures during the source/drain anneal. The last can be excluded, since performance of gate-last and gate-first devices without gate metal oxygen anneal was comparable at all gate lengths, suggesting the presence of resist residues in the gate region during source/drain anneal is not an issue. With regard to thermal treatment of the uncapped dielectric, this took place also for gate-last devices not subject to O_2 anneal of the gate region, thus it does no cause device deterioration. It is therefore thought that the main reason for the great reduction in output current is the second anneal of source/drain contacts, which happens in oxygen atmosphere. It is likely that the contacts became more resistive due to reaction of metal with the annealing gas, thus injecting fewer carriers into the channel for a given V_{ds}.

A comparison of other device parameters for gate-last devices with or without annealing of the gate metal, shown in table 4.2, exhibits an improvement in threshold voltage values, which shift towards positive values, similarly to what observed for gate-first devices. Once again smaller channel lengths presented a higher shift in threshold voltage. The sub-threshold slope either showed no change for 1 μm gate lengths or displayed a reduction for more scaled MOSFETs. The reduction in sub-threshold slope and the positive shift of the threshold voltage are both indicators of a better gate control. This suggests the oxygen annealing was beneficial to the gate stack, and further supports the theory that deterioration of source/drain contacts caused the reduction in output current, given other device parameters not dependent on source/drain resistance showed an improvement. Summarising, in gate-last devices the thermal anneal of the gate in oxygen atmosphere has a beneficial effect on the gate control function similar to what observed for gate-first devices, but the degradation in source/drain contact resistance due to their exposure to oxygen gas at high temperature causes a reduction in output current.

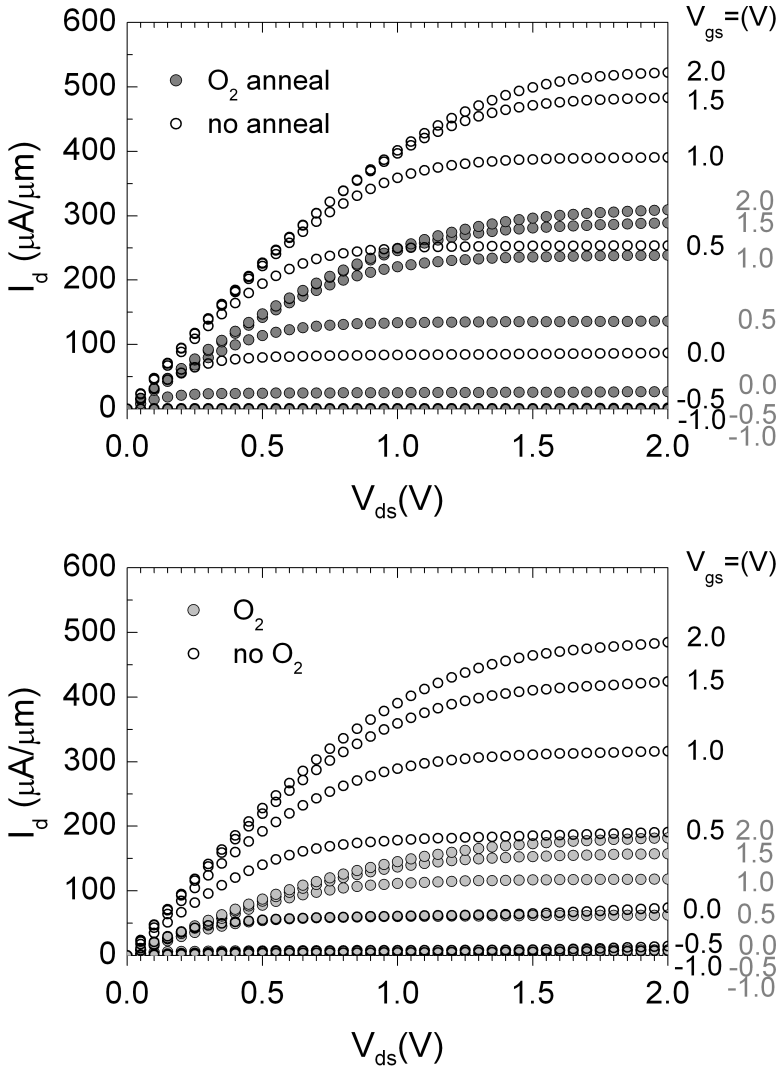


Figure 4.14 I_d vs V_{ds} curves for various V_{gs}, comparison of gate-last MOSFETs with and without O_2 anneal step after gate metallisation for a 1 μm gate length (a) and a 300 nm gate length (b). All traces are average of multiple devices.

Gate-last	**L_g = 1 μm**		**L_g = 300 nm**	
Post gate metallisation O_2 anneal	No	Yes	No	Yes
V_{th} [V] (V_g at I_d = 1 μA/μm)	-0.43	-0.27	-0.45	-0.16
$Slope_{min}$ of log(I_d)-V_g [mV/dec]	82	82	140	87

Table 4.2 Summary of sub-threshold characteristics extracted from GL devices summarising the effect of annealing the gate metal.

This is consistent with gate-last and gate-first devices showing comparable performance at all gate lengths if gate annealing was not part of the fabrication flow, and instead presenting major differences in the way oxygen anneal of the gate affected output current. Given the results obtained, it can be concluded that annealing of the gate metal in oxygen atmosphere improves the gate function. However, if a reduction in output current is to be avoided a gate-first fabrication flow must be chosen so that the oxygen anneal does not hinder the source/drain contact resistance.

4.4 Chapter conclusions

In this chapter several fabrication aspects that could affect MOSFET performance were analysed, and their impact assessed.

It was found that some resist residue is always present in the gate region regardless of development time, exposure dose and type of surface (i.e. dielectric or semiconductor). Oxygen plasma is thus always needed to achieve a metal oxide interface free of organic contamination. A method to determine the thickness of the resist residue was developed, which could be of help in deciding the minimum RF power and ashing time required to clear the residue. Variation in resist thickness was found to have a negligible effect on the gate metal linewidth achieved by lift-off, therefore the exposure dose to clear open areas can be used. However, an overexposure amounting to a 50-90 % increase from clearing dose, depending on the molecular weight of the resist and the substrate type, is recommended to minimise the granularity of resist residues after the development and to produce a smooth line without metal 'flagging' after the lift-off.

When the effect of resist residues on MOSFETs performance was assessed, the devices not subject to oxygen ashing after the gate lithography step showed a great variation in output characteristics. Transistors with an ashed gate region, thus free of resist residues, presented an improved uniformity, however their device parameters indicated plasma damage. A trade-off must therefore be sought between

damage to the gate dielectric and effective removal of resist residue. The effect of an oxygen annealing step applied to the gate contact was also studied for two different fabrication flows, either gate-first or gate-last. It was found that the annealing of the gate improved MOSFET output characteristics for gate-first devices, but negatively impacted gate-last ones. This was related to degradation of source/drain contacts, and it can be concluded that an annealing step in oxygen atmosphere is beneficial to device performance as long as the ohmic contacts have not been patterned yet.

5 Development of specialised MOS capacitor structures for assessment of parasitic elements

5.1 Threshold voltage issues

The III-V heterostructure MOSFET described in the previous chapter was intended for normally-off operation ($V_{ON} > 0$ V) at threshold voltage $V_{TH} \sim 0.3$ V. This was achieved by adjusting the doping levels of the two silicon layers above and below the channel according to equations 5.1, which describes the relationship between the doping levels and the work function of the metal gate layer, in this case, platinum $\Phi_m = 5.7$ eV. The equations are taken from [73]. These devices achieved high output current and good subthreshold characteristics at 1 μm gate length [9, 10, 11].

$$\rho_s = \frac{1}{Q_s \mu q}, \quad where \quad Q_s = \frac{\varepsilon_{ox}(\Phi_m - \Phi(0))}{t_{ox} q} \tag{5.1}$$

- ρ_s is the sheet resistivity (Ω/square) of the source and drain access regions
- μ is the channel mobility
- q is the electronic charge (1.6×10^{-19} As)
- Q_s is the sheet charge (cm^{-2}) of doped layers measured at source and drain access regions
- ε_{ox} is the dielectric constant of gate oxide layer
- t_{ox} is the thickness of gate oxide layer
- Φ_m is the work function of metal gate electrode
- $\Phi(0)$ is the reference work function of metal gate electrode for a specific threshold voltage and no doping layers obtained by two-dimensional device simulation

However, as shown in figure 5.1 that summarises the threshold voltage values extracted from all measured devices in this project, only a few exhibited normally-off operation. For 1 μm gate length devices, the measured V_{TH} ranged between -0.25 V and +0.3 V. Such large variations in device characteristics could be related to variations in the GGO layer growth process. It is expected to have changes in cell temperatures between layers, as cell properties changed with time. Even though a stable growth rate and composition were maintained by a constant flux, GGO reproducibility was still an issue compared to III-Vs, and GGO composition varied from 19 to 23% varying the thickness by 1-2 nm for a 10 nm layer. This was not limited to Glasgow MBE tool and was also observed by Freescale.

It should be noted that 1 μm V_{TH} values for gate-first devices are higher than those for gate-last devices. According to the findings reported in chapter 4 that looked at gate-metallisation damage, this could be due to a reduction of fixed positive charge in the dielectric when the gate metallisation-induced damage was recovered with the 430°C N_2 ohmic contacts anneal, while gate-last devices did not have a thermal treatment after gate deposition.

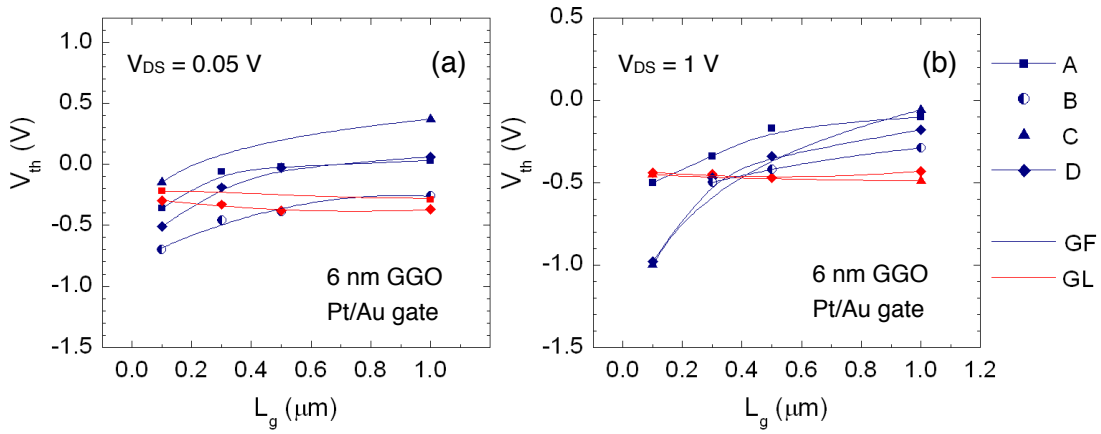


Figure 5.1 $In_{0.3}Ga_{0.7}As$-channel GGO-dielectric Pt/Au-gate heterostructure MOSFET. Threshold voltage spread extracted from Log $I_{DS}(V_{GS})$ in linear (a) and saturation (b) regime, for a variety of wafers grown with identical nominal semiconductor and dielectric layers thicknesses.

Apart from the long-channel device V_{TH} variations, gate-length dependent variations are evident from the data spread of figure 5.1. When the gate length was decreased from 1 μm to 100 nm, a shift of threshold voltage towards smaller gate voltages was observed [74]. This threshold voltage roll-off effect is undesirable as it limits the transistor scaling potential. However, the effect was likely to be process-dependent as only gate-first devices suffered from it, most likely related to the ohmic anneal process step. It was discussed in chapters 2 and 4 that during thermal annealing, atomic out-diffusion of elements could take place. Whether it is arsenic or indium in the semiconductor layers or oxygen loosely bound in the dielectric layer, migration of atoms could be taking place at 430°C. If it is As or In atoms, it is possible that when the gate metal was in place during the ohmic anneal in the gate-first process, accumulation at the metal surface underneath the gate took place, whilst the atomic As and In in the region between the contacts out-diffused completely. This way, the part of the dielectric film under the gate would have different properties to the part that is not covered with the gate. Since the device is designed in such a way that the gate-ohmic separation is fixed for all gate lengths, the volume ratio of the gate region to the region between the gate and the ohmic would vary with the gate length, thus causing variations in the electrical properties of the whole system. In the devices fabricated gate-last, the ohmic anneal takes place before the gate contact is deposited so the atomic out-diffusion is uniform across all of the dielectric area, and there is no gate-length dependent region with properties differing from the rest of the dielectric.

In a planar MOS transistor with lithographically-aligned contacts and a wire-like gate contact, a non-ideality phenomenon exists: in addition to the gate parallel-plate capacitance, which is the main capacitive element in the MOS system, there are three more capacitances, namely fringe capacitance, gate-ohmic capacitance, and air (or passivation) capacitance. Figure 5.2 shows a schematic diagram of the field lines between the gate and the source when the transistor is biased, indicating each capacitive element. For a fixed gate dielectric thickness, the fringe capacitance C_{fringe} is dependent on the periphery of the gate contact, whilst gate-ohmic capacitance $C_{g/o}$ and air capacitances C_{air} depend on the distance between the source/drain and the gate contacts. The air capacitance is insignificant in the high-κ dielectric

containing MOS system, as the dielectric constant of air is 20 times smaller than the dielectric constant of GGO. The C_{fringe} and $C_{g/o}$ can be considered parasitic capacitances as both the value of those capacitances and their effect on the performance of small gate length devices are unknown.

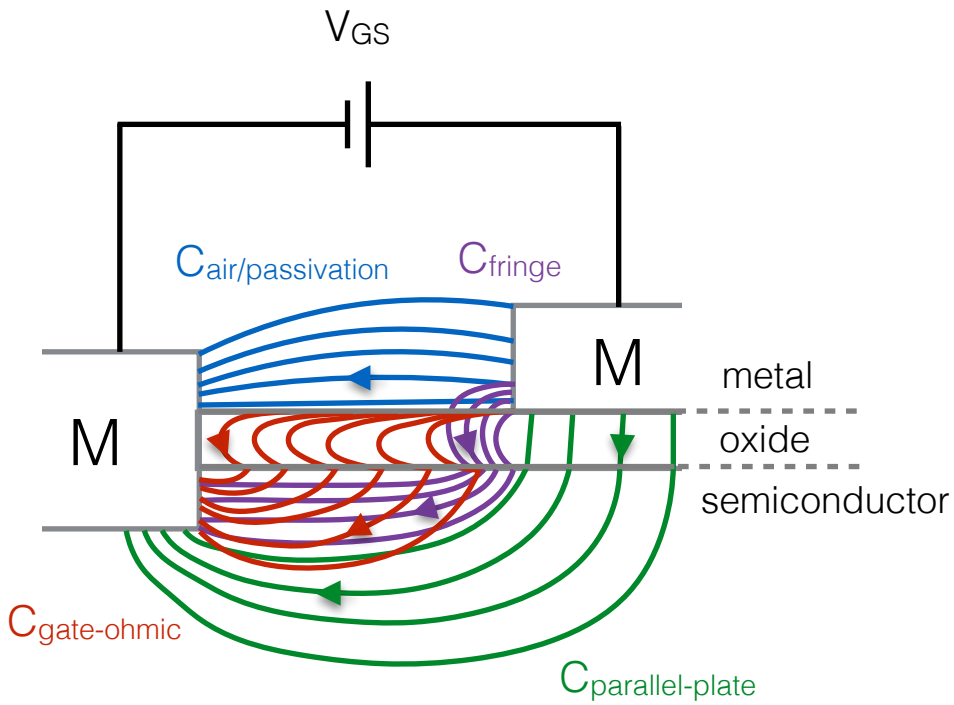


Figure 5.2 Field lines and associated capacitances in a planar MOSFET with lithographically-aligned contacts, showing the region between the gate and the source contacts only.

It is important to investigate the contribution of the two parasitic capacitances to the total gate capacitance for two reasons. First, the majority of the devices measured during the course of this project did have threshold voltages below 0 V, albeit the materials and the architecture of the device designed for $V_{TH} > 0$ V. The reason for this shift in the threshold voltage could be expressed through equation 5.2 [75] that suggests a possible role of parasitic capacitances on the threshold voltage. If the fringing capacitance is significant even at large gate lengths, it translates into a reduced dielectric thickness parameter d_i, and a subsequently reduced threshold voltage.

$$V_{th} = \Phi_B - \frac{\Delta E_C}{q} - \frac{qN_D(d-d_i)^2}{2\varepsilon_0\varepsilon_1} - \frac{\sigma_{pol}}{\varepsilon_0\varepsilon_1}d \tag{5.2}$$

- Φ_m is the barrier height between the metal and the first epitaxial layer
- ΔEc is the discontinuity at the heterojunction
- N_D is the doping layer concentration
- d is the dielectric thickness
- d_i is the reduction of the dielectric thickness
- σ_{pol} is the effective thickness of the 2D electron gas

The second reason for an investigation into the parasitics is possible issues when the devices are scaled to sub-100 nm gate lengths. The gate-ohmic capacitance could be dominating the total MOS capacitance, and if the gate width to gate length ratio is large, the periphery-dependent fringe capacitance could be an issue also.

For assessment of contribution of the fringe and gate-ohmic capacitances to the total gate capacitance, the standard large-area MOS capacitor is unsuitable due to its parallel-plate capacitance dominating the total capacitance. For this, a specialised capacitor structure was designed, where the geometry of the gate and source/drain contacts was altered in such a way as to both mimic the device gate contact geometries and vary the periphery of the gate. The designed structures were then to be measured alongside a standard circular gate large area capacitor. The large-area structure was used to study the dependence of the *total* capacitance on the periphery of the contacts and on the gate-ohmic separation. Since the reduced equivalent source-gate-drain dimensions is the most significant feature of the specialised structure with respect to the standard MOSCAP, throughout this chapter it shall be dubbed “scaled capacitor” or “multi-finger” structure. The reasons for the latter denomination will become clear shortly.

Next, the layout of the scaled capacitor structure will be schematically described and compared to the large area capacitor structure. A description of the fabrication challenges faced during its process development will follow, then C-V measurements carried out will be presented and discussed.

5.2 Scaled capacitor structure design

The most obvious way to increase the total area of the gate/channel whilst keeping the dimensions small is folding the gate contact in a snake-like fashion (figure 5.3). This, however, will have a large gate series resistance. A multi-finger layout is a common design solution to this problem, and it was used for the scaled capacitor structure as well.

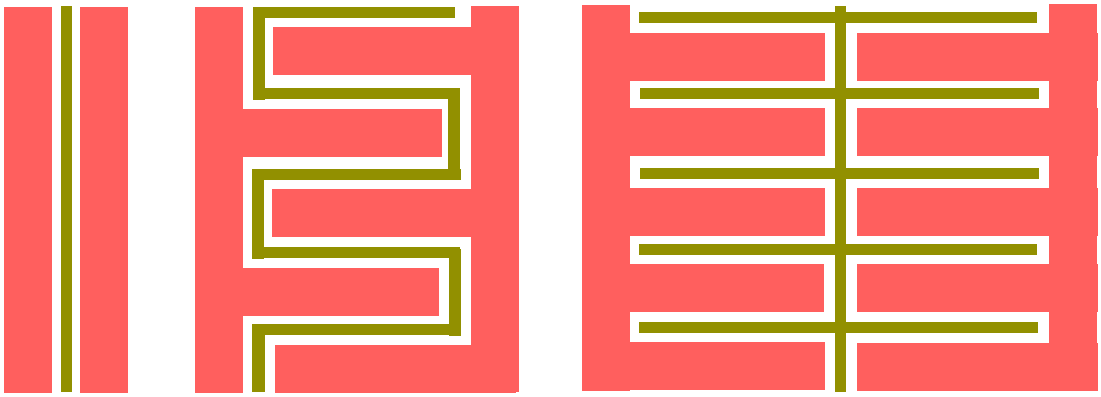


Figure 5.3 Possible ways to increase the gate region area without altering the gate length. Left: standard design for a test MOS transistor; Centre: gate folded into a snake-like shape; Right: multi-finger gate layout.

In order to determine the minimum gate area requirements for the multi-finger MOSCAP, capacitances of standard circular large-area MOSCAPs were measured and plotted against their corresponding top contact area, as the table in figure 5.4(a) summarises. The three measured points are plotted against the effective area (derived from $A = \pi r^2$) together with a linear fit in figure 5.4(b). Figure 5.4(b) also shows the intersection of the linear fit with the horizontal line representing the minimum capacitance measurable by the semiconductor parameter analyser used in this work. This is 3.5×10^{-17} F which corresponds to 124,000 nm^2 area. It was decided that a gate area 1000 times greater than the measurement limit equivalent should be the smallest gate of the scaled capacitor structure to achieve a valid capacitance measurement.

Radius (m)	Effective area (m^2)	C_{acc} (pF)	C_{acc} / area (F/m^2)
$3.5\ 10^{-5}$	$3.85\ 10^{-9}$	10.7	$2.78\ 10^{-3}$
$2.5\ 10^{-5}$	$1.96\ 10^{-9}$	5.5	$2.81\ 10^{-3}$
$2.0\ 10^{-5}$	$1.26\ 10^{-9}$	3.5	$2.78\ 10^{-3}$

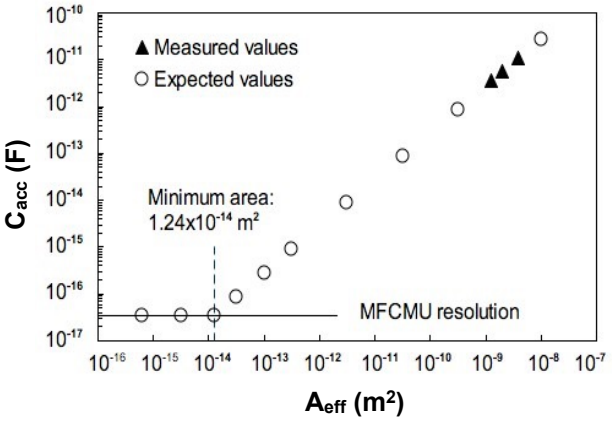


Figure 5.4 Determining the gate area requirements of the multi-finger MOSCAP to produce a valid capacitance measurement. Right: Capacitance per MOSCAP plotted against Gate Area of 40, 50 and 70 μm diameter circular MOSCAPs, continued until intersection with the LCR meter capacitance measurement limit. Left: Key values in a table format.

Assessment of parasitic element contribution to the operation of a MOS capacitor requires varying three design parameters: the *gate length* L_g (the green area in figure 5.5) taken from a cross-section through the centre of the circle/finger, the *gate-ohmic gap* $L_{g/o}$ (the gate is the green area, the ohmic is the red area in figure 5.5), and the number of times the cross-section is repeated, i.e. the *periphery*. The multi-finger design enables both a significant increase in the periphery and a significant decrease in the gate length, when compared to the standard circular structure.

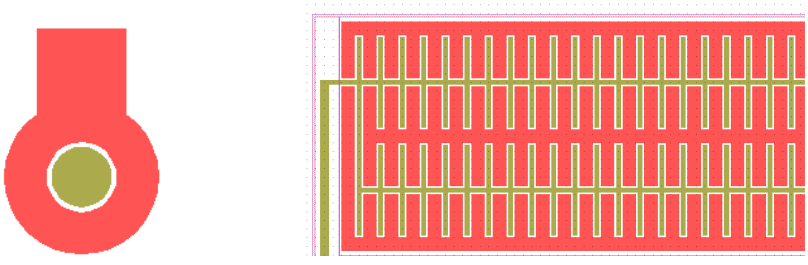


Figure 5.5 Large L_g circular (left) and small L_g multi-finger (right) MOSCAP layouts.

5.3 Fabrication process development

It was challenging to fabricate the scaled capacitor structure due to the dimensions of the gate metal fingers and the proximity of the gate and ohmic contacts. The issues and the solutions adopted to implement the final MOSCAP structures are described in the following.

5.3.1 Flagging of the gate metal

For the devices fabricated in this work, a gate metal stack of 20nm of Pt and 100 nm of Au was used, while the gate metal linewidth went from 90 nm for the smallest MOSCAPs up to 1 μm for the largest ones. As the fabrication process was being developed, flagging of the gate metal for the smallest gate length was observed. This phenomenon occurs when the profile of the resist used for the metal lift-off process does not present sufficient undercut, resulting in some metal being deposited on the resist sidewalls and not being lifted off properly, as the top-view SEM image in figure 5.6(a) shows. Metal flagging is more likely to occur if a thick resist (several hundreds of nm) is used, and at low doses, where the undercut is smaller. Dose tests were carried out and it was found that for PMMA exposed to its clearing dose the lifted-off metal lines were severely affected by metal flagging (figure 5.6(a)). Overexposure of resist pattern to as much as 70% above the clearing dose was necessary in order to have clean metal lines (figure 5.6(b)), therefore when fabricating the specialised MOSCAP structures small gate lengths were always overexposed to guarantee a clean lift-off.

5.3.2 Fabrication of closely spaced S/D contacts

5.3.2.1 Lift-off challenges

The MOSFET devices analysed in the previous chapter were designed with 1 μm spacing between the ohmic contacts and the gate, as shown in figure 5.7(a). For the

scaled MOSCAP structures it was important to reduce that spacing as much as possible to have a wide range of $L_{g/o}$ values for $C_{g/o}$ studies and to minimise the impact of $C_{g/o}$ in periphery studies. As shown in figure 5.7(b), the smallest $L_{g/o}$ was 90 nm, the same as the gate length.

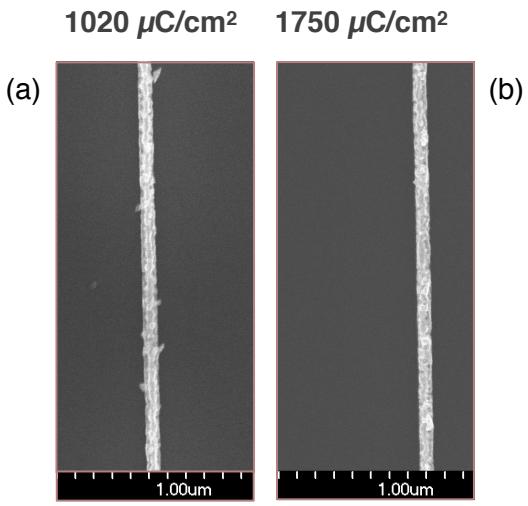


Figure 5.6 Overexposing the gate pattern for a clean gate metal line after lift-off. Line exposed with the clearing dose (left) and overexposed (right).

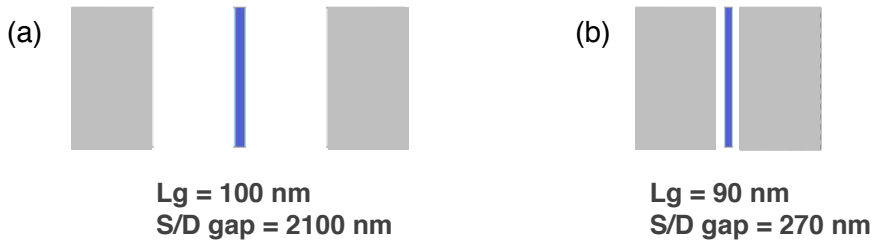


Figure 5.7 MOSFET with large gate-to-ohmic distance (a) and optimised scaled design (b).

The small dimension of the source-drain gap, which was eventually to be scaled down to 270 nm for 90 nm gates, created an issue if it was to be defined using standard e-beam lithography, metallisation and lift-off techniques, since the gap was to be achieved between two 100 μm^2 contact pads. During electron beam exposure large areas suffer *proximity effect*, due to electron back-scattering through the resist

resulting in certain areas receiving extra dose of electrons. In a narrow gap, it is 50% of the applied dose. When trying to define the 270 nm spaced large rectangles constituting source and drain contacts, after exposure and development it was observed that the linewidth of the thin line of PMMA between them was significantly lower than the designed value, with this effect being at times so severe that the resist line disappeared completely and the rectangles were joined together. For this experiment, a bilayer PMMA was used, with a total thickness of 350 nm. After proximity-effect corrected e-beam exposure, the resist was developed in IPA/MIBK 2.5:1 solution for 60 s. Furthermore, AFM surface scans in the source/drain gap region highlighted also that the thickness of the thin PMMA line decreased when going from the rectangle corners towards the middle of the source/drain gap. Figure 5.8(a) shows this loss in resist thickness for several source/drain gaps. Resist thickness loss at the centre of the gaps before oxide etching was measured to be between ~80 and ~180 nm, with the highest value measured for a 270 nm source/drain gap. Such a thin resist between the two large contact areas jeopardised metal lift-off, which had a very low yield and was successful in few isolated cases, such as the optical micrograph shown in figure 5.8(b).

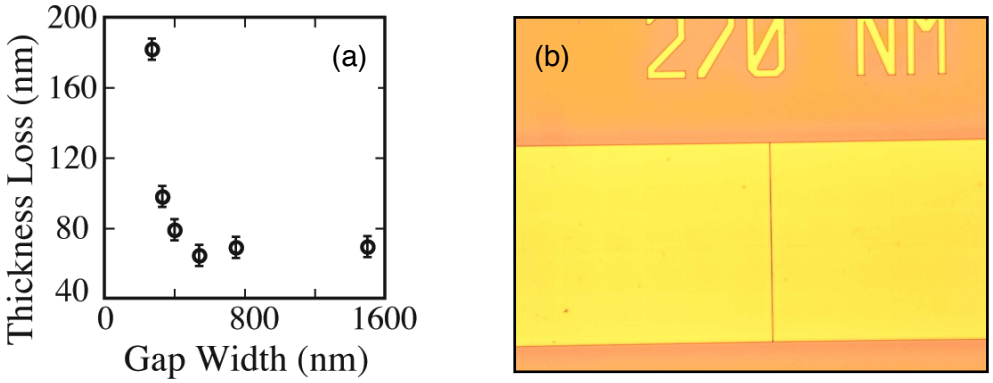


Figure 5.8 (a) Resist thickness loss of the thin PMMA line separating the source and drain region plotted against S/D gap width. Successfully lifted-off S/D contacts are shown in (b).

Many tests were carried out to improve this aspect, e.g. various proximity correction schemes, 'cold' PMMA development at 4°C for a higher contrast, and possible alternatives to PMMA resist such as the higher contrast UVIII resist. The alternative resist reduced the thinning to 15 nm yielding satisfactory lift-off but also decreased

the quality of the ohmic contacts, most likely because of resist residue. Good process reliability and repeatability of lift-off for the source/drain contacts could not be obtained below a 540 nm gap. It can be concluded the source/drain gaps smaller than 500 nm are beyond the means of standard e-beam lithography techniques if a double layer of PMMA is to be used. More complicated methods, such as lift-off by means of a thin HSQ line could overcome this limit, but they were not investigated as part of this work.

Another issue which the use of sub-micrometer source/drain gaps created regarded dielectric removal prior to metal deposition. Removing the oxide from the source and drain contact regions using standard wet etch techniques was no longer viable below a certain source/drain gap because the isotropic nature typical of those processes etched the oxide also laterally and removed it from the gate region. The need for an anisotropic dry etch process became apparent to define closely-spaced structures. Since the same PMMA pattern was to be used both as an oxide etch mask and as a metal lift-off mask, its poor dry etch resistance had to be taken into account considering that as much as 200 nm of resist would be eroded during the dry etching of the oxide. This extra thickness was therefore added to the typical PMMA thickness of 350 nm, necessary to lift off 150 nm thick source/drain contacts, thus the specialised MOSCAP structures were fabricated using a total PMMA thickness of 550 nm.

5.3.2.2 Thermal annealing challenges

The first attempt at fabricating the multi-finger MOSCAP structures with a sub-micron source/drain gap revealed a problem with the annealing of the source and drain contact. Annealing took place in nitrogen atmosphere for 30 seconds at 430°C. The metals composing the stack laterally diffused into the semiconductor and altered the characteristics of the channel region. Fabricated MOSCAP structures which suffered from this phenomenon are shown in figure 5.9, where a lateral diffusion of approximately 330 nm into the semiconductor was observed. Lateral diffusion of ohmic metal did take place also for the devices analysed in the previous

chapter, but it was not detrimental to their operation since they had a source/drain separation of 2.1 μm, well in excess of the metal lateral diffusion length.

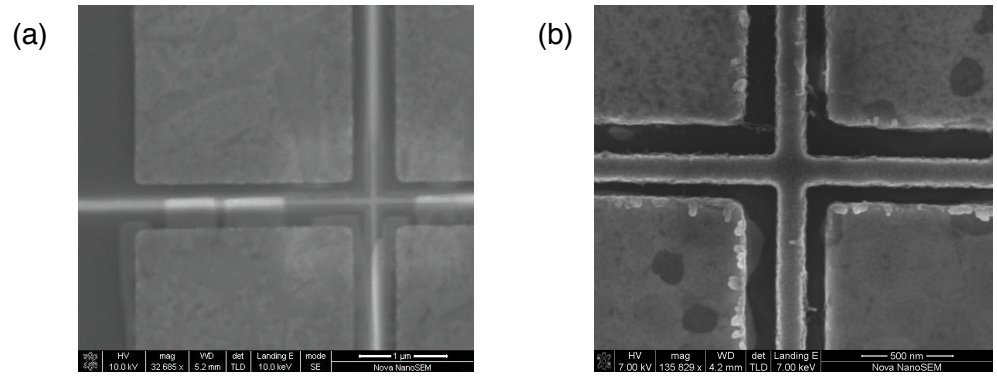


Figure 5.9 (a) Ni(10nm)/Ge(10nm)/Au(100nm) ohmic contacts with 540 nm gap: a) showing gate PMMA pattern before metallisation, and b) after gate metal lift-off. The maximum length of the annealing-induced lateral diffusion of the ohmic metal in the gate region is approximately 330 nm.

Several tests were carried out in order to gain a better understanding of the lateral diffusion of the ohmic contact metal and its dependence on various parameters such as temperature and composition of the metal stack. The original metal stack, in order from the semiconductor surface up, was composed of 10 nm of nickel, 10 nm of germanium, and topped by 100 nm of gold. Figure 5.10 shows the results of annealing this metal stack in nitrogen atmosphere for a range of durations at 430°C: it can be seen that the lateral diffusion length proportionally increases with the annealing time. Two reasons were thought most likely to cause the lateral diffusion: first, the reactivity of nickel; second, the germanium and gold layers not constituting a eutectic alloy, i.e. an alloy having a single and well defined melting point. In order to identify the main cause of lateral diffusion, two alternative ohmic contact metal stacks to the original one were tested: in both the germanium-gold ratio was changed to Au:Ge 88%:12% by weight to achieve a eutectic composition, but in one of them platinum replaced the more-reactive nickel. The corresponding metal compositions were Ge (20 nm)/ Pt (15 nm)/ Au (40 nm) for the eutectic case without nickel and Ge (20 nm)/ Ni (15 nm)/ Au (40 nm) for the second tested composition, making the total thickness of the contact 70 nm in both cases. In [76] it

was stated that Ge/Ni/Au contacts are practically insensitive to changes in annealing temperature, while recommending 450°C for Ge/Pt/Au, thus it was decided to anneal the non-standard metal stacks at 450°C. For the non-standard metals a series of annealing duration tests were also carried out, with 10, 15, and 30 seconds long anneal cycles.

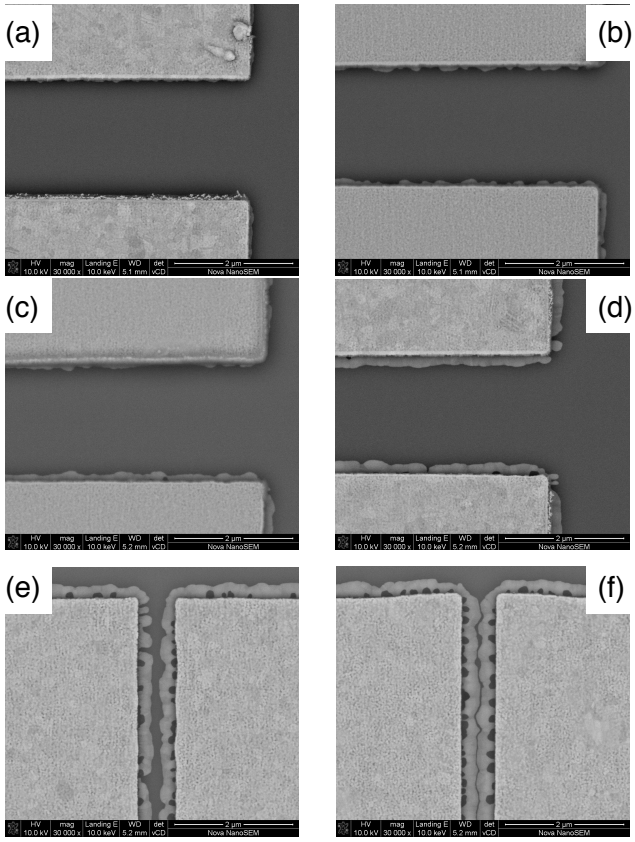


Figure 5.10 Standard source and drain metal stack (Ni(10)/Ge(10)/Au(100)) annealed in N_2 at 430°C for 10 s (a), 30 s (b), 45 s (c) and 60 s (d). The two lower images (e) and (f) illustrate the lateral diffusion problem when the source/drain gap is scaled.

Figure 5.11 shows the results of the annealing tests on the non-standard metals and compares them with the original metal composition. It can be seen that the least lateral diffusion resulted from the platinum containing stack, suggesting nickel played a major role in this phenomenon. However, when the contact resistances of these contacts were compared, it became clear that nickel was required for a good ohmic contact on the adopted epitaxial structure. In fact the nickel-free contacts

exhibited the worst ohmic behaviour, in spite of, and maybe because of their lack of metal diffusion. Further investigation was needed in order to determine an ohmic metal stack with good ohmic behaviour and unaffected by lateral metal diffusion. Given the time constraints associated with the project, it was decided to keep the original ohmic metal composition (Ni(10)/Ge(10)/Au(100)) while relaxing the gate to ohmic spacing to 1 μm.

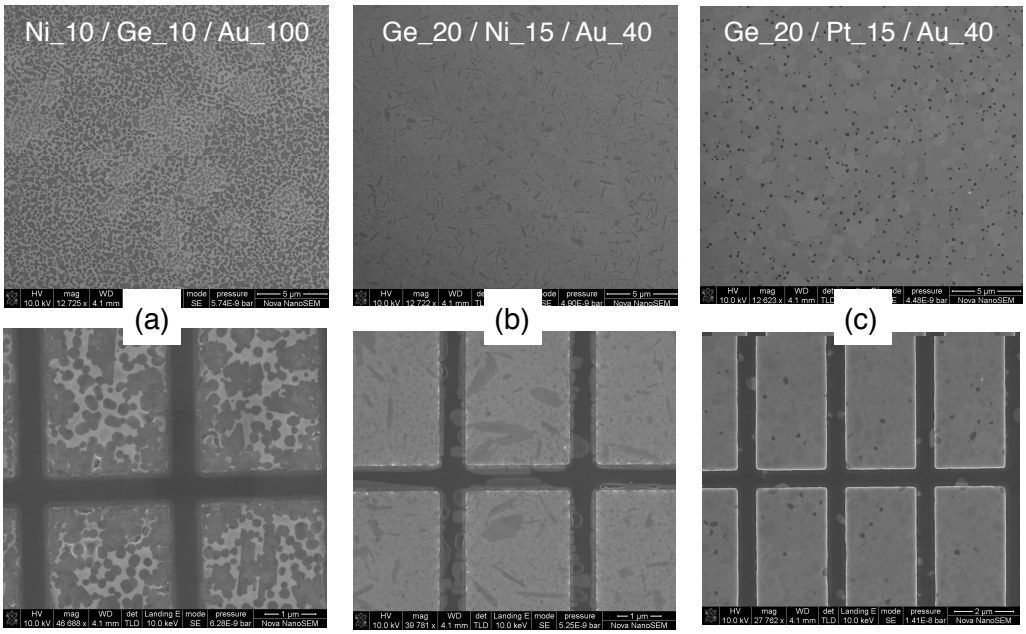


Figure 5.11 Attempt at reducing ohmic contact lateral diffusion. The large area (top row) and the gate region (bottom row) of (a) standard Ni/Ge/Au, (b) eutectic Ni/Ge/Au, c) Pt/Ge/Au. The ohmic-gate gaps are 300 nm.

5.3.3 Final process

The final fabrication process flow was a gate-last process. After surface preparation with a standard organic solvent clean, the NiGeAu ohmic contacts were formed by lift-off ensuring a clean surface prior to dielectric removal in hydrochloric acid with an oxygen ash and finishing the contact formation with a 430°C anneal in nitrogen. Similarly, the PtAu gate was patterned by lift-off with its pattern lithographically aligned to the ohmics using four corner markers. For the experiments involving oxygen annealing of the gate, a 350°C O_2 annealing step was inserted straight after gate metallisation. The final step of the process was forming a narrow trench (300

nm) around the device to separate the active area from the probing pad area. This was done by a two-step wet-etch process, involving an HCl acid step to remove the dielectric, followed by a peroxide/orthophosphoric acid etch to etch 50 nm into the semiconductor layer structure to reach below the active layers, finishing with a resist strip in warm acetone. The complete structure is shown in figure 5.12, detailing various regions.

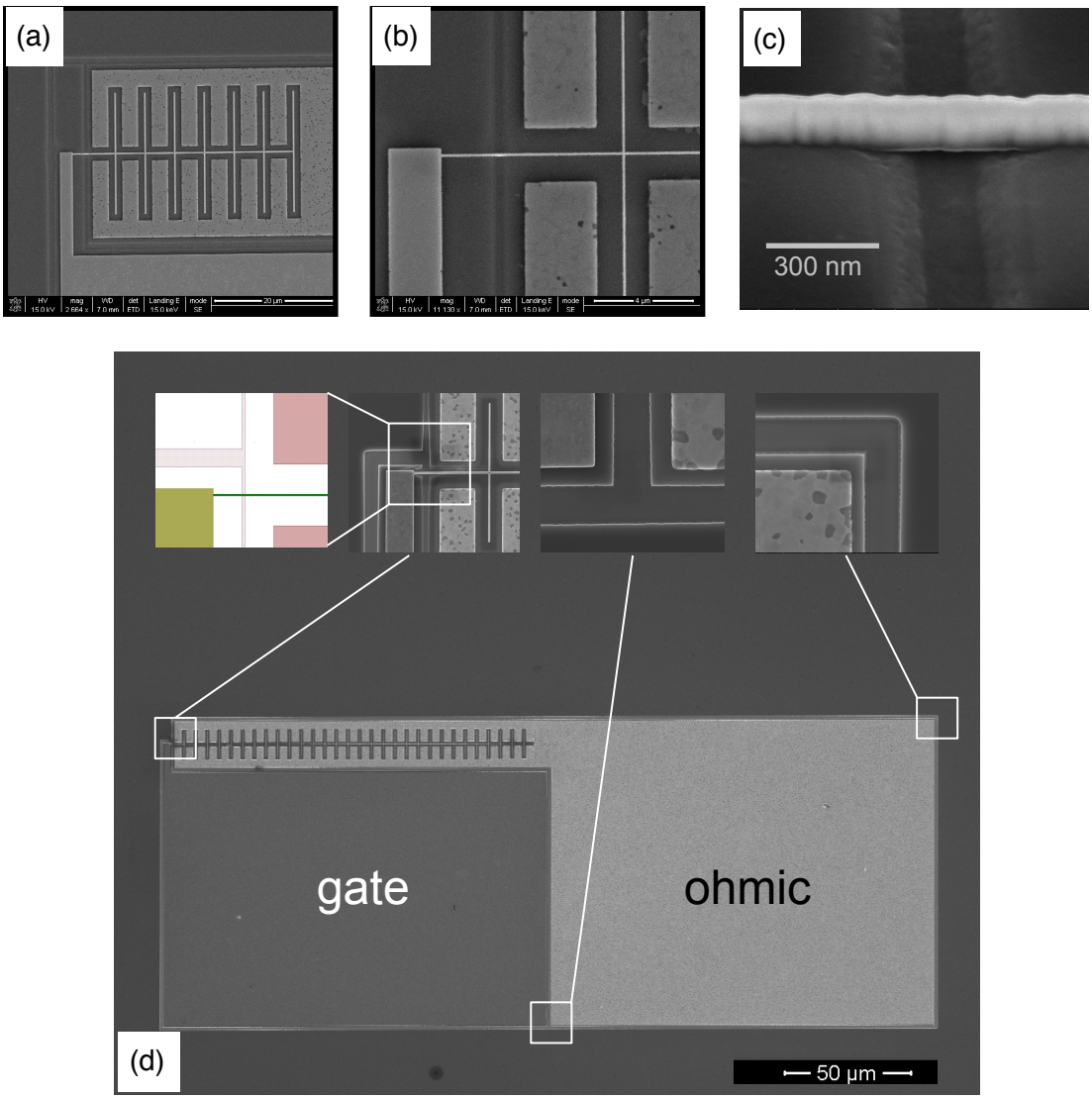


Figure 5.12 The complete scaled MOS capacitor structure: (a) the MOS layers layout shown for a 330 nm finger width structure, (b) a medium zoom into the isolation area shown for a 125 nm finger width structure, (c) a high zoom into the isolation trench showing the metal gate line forming a bridge over the trench, and (d) the full device layout, including inlets of mask layout design and isolation in various areas of the device. The ohmic metal is Ni(10nm)/Ge(10nm)/Au(100nm) annealed by N_2 RTA at 430°C for 60s. The gate metal is Pt(10nm)/Au(90nm).

5.4 Measurements

5.4.1 Device layout

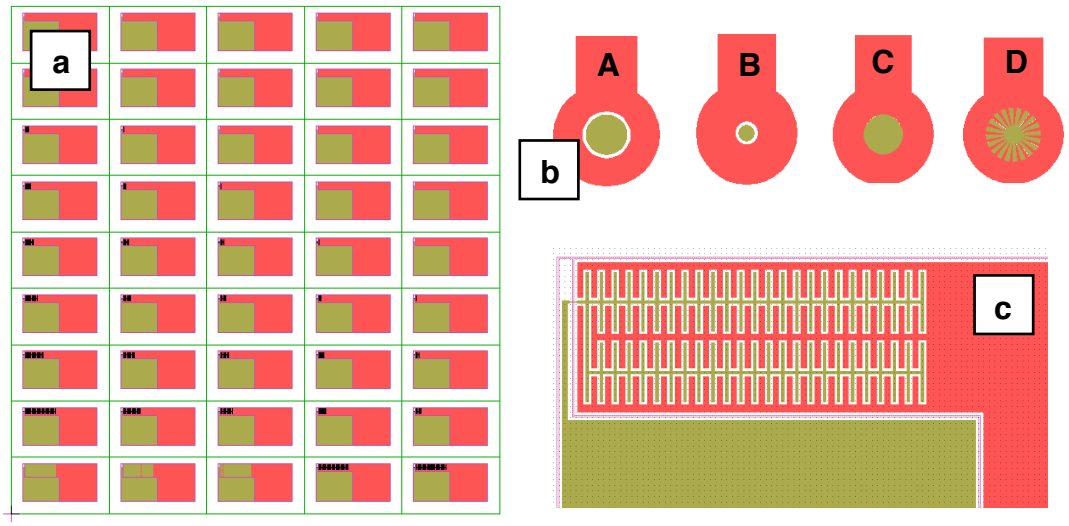
In order to assess the effects of parasitic capacitances on the gate control of the MOSFET, the capacitors were designed with varying finger widths of 90 nm, 125 nm, 330 nm, 510 nm, 660 nm, and 1 μm. At each finger width, the length of the finger was fixed, but the number of fingers were varied in such a way as to cover a range of total gate area values with the minimum value equal to 39.7 μm^2, corresponding to an equivalent circular area with a radius of 3.55 μm, as defined by the capacitance measurement limit of the LCR meter, described earlier in the design part of the chapter. Capacitors with a total gate area equivalent to the standard circular gate MOS capacitor with a diameter of 40 μm were achieved by adding a second row of fingers to some of the 1 μm finger-width capacitors, as shown in figure 5.13(a) and (c).

Due to the small dimensions and many variations in the layout of the gate, the structures were all examined in an SEM and gate contact geometries were measured so that an accurate number of the effective gate width W_g and the gate length L_g could be fed into total gate area A_g calculation. This was necessary because the capacitance is measured per unit area and it is calculated automatically by the measurement tool software where the radius of the equivalent circular area is specified for each C-V measurement.

5.4.2 Periphery effect

In order to understand the periphery capacitance, it is useful to look back at its different components. The fact that the dielectric is etched prior to source and drain deposition, and the following deposition of ohmic metal, give rise to an electrode whose size will be the thickness of the dielectric times the length of the source and drain contact along the gate direction. The gate metal constitutes the second

electrode, and the distance between them to first approximation is the gap between the gate and the ohmics. This means that the periphery capacitance should be affected by both changes in periphery, that correspond to changes in its total area, and changes in gate-ohmic gap, equivalent to changes in dielectric thickness instead. Various multi-finger structures were thus designed, either keeping the gate-ohmic gap the same and changing the periphery or vice-versa. Whenever possible the total area was kept the same, even though it is not essential to a meaningful comparison.



Device	Area (μm^2)	Periphery (μm)	Gate-ohmic gap (μm)	Peripheral area (μm^2)
Dot “A”	7854	314	10	3140
Dot “B”	1250	126	10	1260
Dot “C"	8495	327	1	327
Dot “D”	8495	2210	1.5	3315
1 µm SCAP	1250	2500	1	2500

Figure 5.13 (a) An array of scaled capacitor structures with varying gate area, (b) standard large-area capacitor designs aimed at changing the gate area (“A” and “B”), the edge of the gate contact (“C” and “D”), and the gate-ohmic gap size (“A” and “C”), (c) the 1 μm finger width structure with an additional row of fingers to achieve the gate area equal to the circular capacitor “B”, (d) table detailing the geometry of the MOSCAPs from (b).

First, C-V measurements were carried out on multi-finger capacitors having same gate length and same total area but different periphery. Figure 5.14 displays the results for all the different gate lengths measured. The measured capacitance only showed a plateau at high positive voltages for a gate length of 90 nm, the smallest implemented gate length, whereas for larger gates the capacitance tended not to plateau. With regard to the effect of changing the periphery within a single gate length, as it can be seen from the black dotted curves the results are contradictory. For a gate length of 90 nm a smaller periphery led to an increased stretch-out and to a shift of the C-V curve towards more negative voltages, whereas the opposite was observed for a gate length of 125 nm, with the rest of the measured gate lengths showing a number of behaviours in between. This inconsistency could be due to oxygen vacancies in the dielectric film following epitaxial growth, affecting differently various parts of the wafer.

In order to determine if this was the case, the C-V measurements were repeated after annealing the sample in oxygen atmosphere for 30 seconds at 360°C. The data obtained, displayed by the coloured empty-dotted curves, show that annealing cancels out the differences in C-V curves due to changing periphery that were previously described. This confirms that the changes in capacitances were due to process variability, since as discussed in chapter 3 oxygen annealing improves device uniformity. The lack of observable changes due to periphery in the post-annealing C-V curves suggests that for a given gate length and capacitor area periphery variations have little or no effect on the amount of charges stored in the gate stack at equilibrium. In order to validate this result, C-V measurements were carried out on standard circular capacitors with various peripheries and gate-ohmic spacing and were then compared with the multi-finger structures.

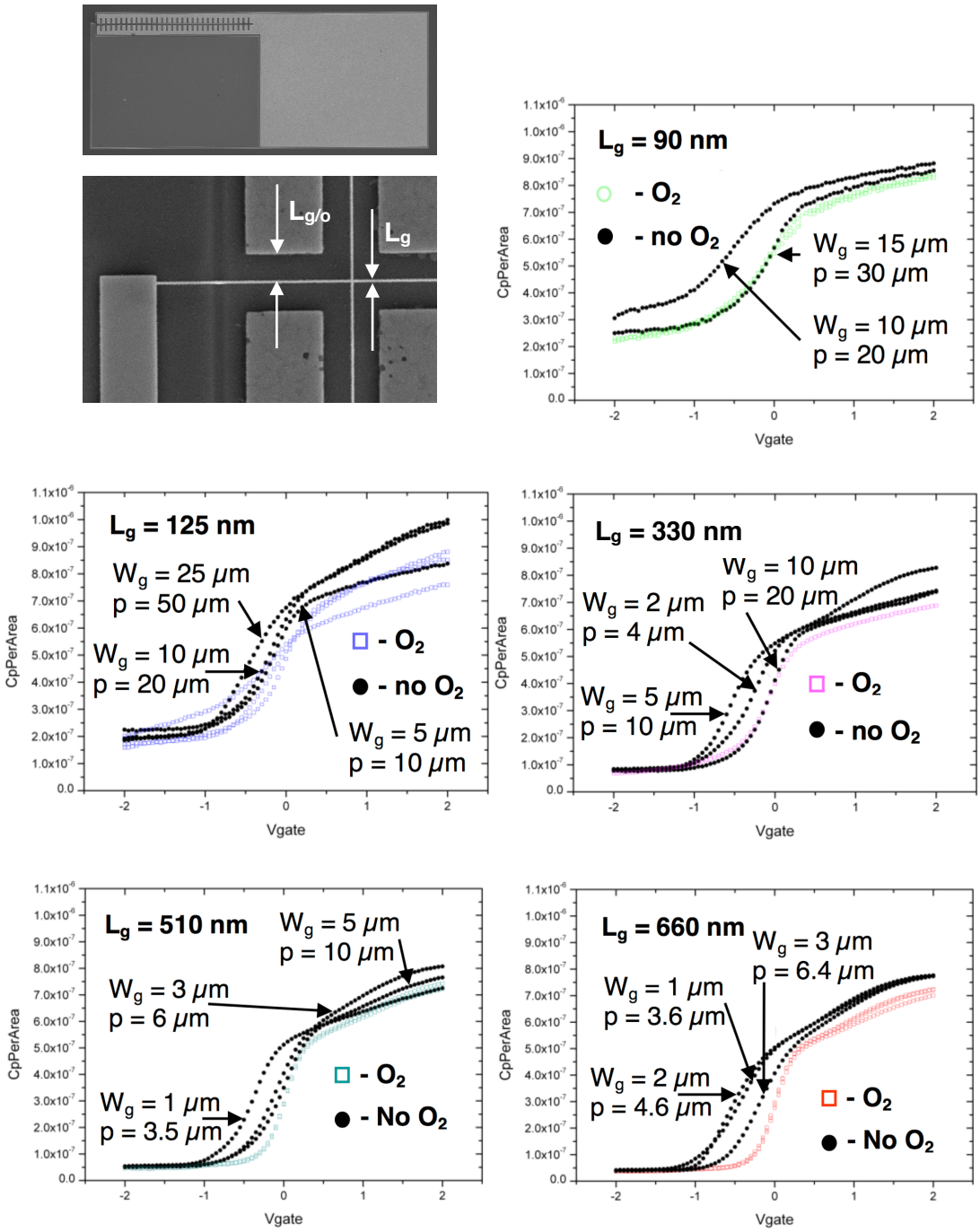


Figure 5.14 High-frequency (1 MHz) C-V measurement of scaled capacitor structures with various gate lengths, as specified in graphs. The gate is swept once from negative to positive bias (inversion to accumulation). The area was kept constant for each L_g group. The $L_{g/o}$ is fixed at 1 μm for all structures.

Figure 5.15 compares results of C-V measurements from two capacitors with total area of 1250 μm^2, one of them being a standard circular dot and the other a multi-finger structure having a periphery approximately 20 times larger. The gate-ohmic gap of the non-standard structure was also ten times smaller than for the circular dot. The different characteristics of the two capacitors should maximise the effect of the periphery capacitance for the multi-finger structure. The results are similar in terms of quality of the interface, by showing a similar slope and stretch out. The main change between the C-V curves regards the minimum and maximum capacitance for the non-standard MOSCAP, which increase by approximately 25 nF/cm^2. The shift of the curve towards higher capacitance values for the increased periphery case does suggest an increase in parasitic capacitance, however small. This is because an increased periphery translates into more dielectric area between the capacitor plates (the gate and the ohmic contacts), which is equivalent to the introduction of parallel RC branches into the system. In addition, due to the increased interface in the high periphery structure, this will have a higher density of interface traps, causing a lower return slope observed in fig. 5.15(b) and a subsequent slight change in hysteresis because of the increased release time of the charge traps.

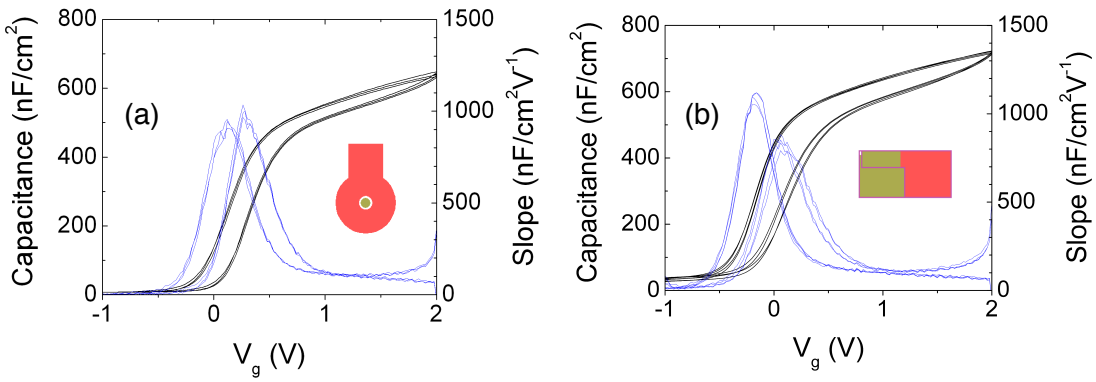


Figure 5.15 High-frequency (1 MHz) double-sweep (negative-to-positive and back) C-V measurement of scaled capacitor structure with 1 μm gate length (b) and standard circular capacitor structure (a). The design specifications can be viewed in the table of 5.13, "Dot B" and "1 μm SCAP".

The effect of gate to ohmic gap on the total capacitance was assessed by comparing two circular capacitors differing only in the gate to ohmic distance, namely 10 μm and 1 μm. Multi-frequency capacitance data are shown in figure 5.16: the compared capacitors show nearly identical slope and stretch-out, but vary in terms of frequency dispersion and threshold voltage. The latter was lower for the 10 μm gate-to-ohmic gap case, consistent with a lower periphery capacitance and thus a lower electric field required to sway carriers; an increase in threshold voltage of 0.2 V was instead observed for a gate-to-ohmic gap of 1 μm. With regard to the variation in frequency dispersion between the compared cases, it is related to an increase in the total density of trapped charges, which generally results in a lower maximum capacitance for high frequency and higher for low frequencies. The plots in figure 5.16 also display a feature known as the D_{it} 'bump', more prominent for the larger gate-ohmic gap, that indicates uneven distribution of interface states in the bandgap. Such feature is associated with the presence of mid-bandgap charge traps. The plotted results thus seem to suggest that for larger gate-ohmic gap the total density of trapped charge is lower (because of small frequency dispersion) but the trapped charge states are mid-bandgap (more obvious 'bump'). On the other hand, a decrease in the gate-ohmic gap associates with an increase in the total trapped charge density (greater frequency dispersion) but the trap states are positioned near the conduction and valence band. The greater edge contribution to the periphery capacitance affects the surface potential and increases the band bending thus allowing charge trapping at increased levels of energy - reflected in the positive shift of the C-V curve in fig. 5.16(b).

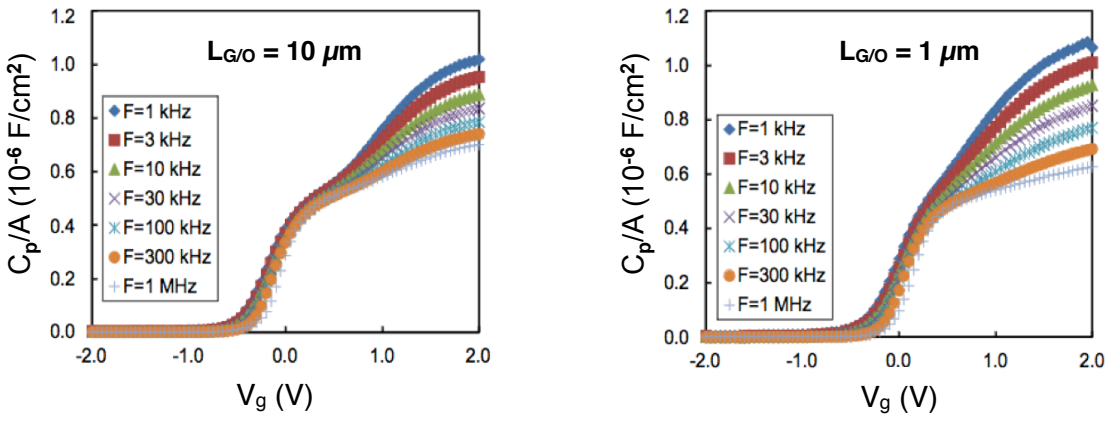


Figure 5.16 Multi-frequency C-V measurement of circular capacitor structure with gate-ohmic separation distance of 10 μm (left) and 1 μm (right).

5.4.3 Area effect

Other than the periphery capacitance, as previously stated also the contribution of fringing capacitance increases for scaled structures. The fringing capacitance is not related to the gate-to-ohmic gap or to the circumference of the contact, but only to the width of the wire, i.e. the gate length. Figure 5.17 summarises the dC/dV of multi-finger capacitors with finger width from 90 nm to 660 nm. The solid-line curves show the data for as-fabricated capacitors, while the scattered points display the data for capacitors where an oxygen anneal was performed after the gate metal deposition. The peak of the curve indicates the threshold voltage of the device. It can be seen that there is a great dispersion in threshold voltage values prior to annealing, while all curves overlap after thermal treatment independently of the gate length of the multi-finger capacitor. The most viable explanation for this is passivation of the oxygen vacancies in bulk of the dielectric and at the interface carried out during the thermal treatment in oxygen ambient. It therefore seems that the effect of the change in fringing capacitance is not measurable, at least for the range of sub-micron gate lengths considered.

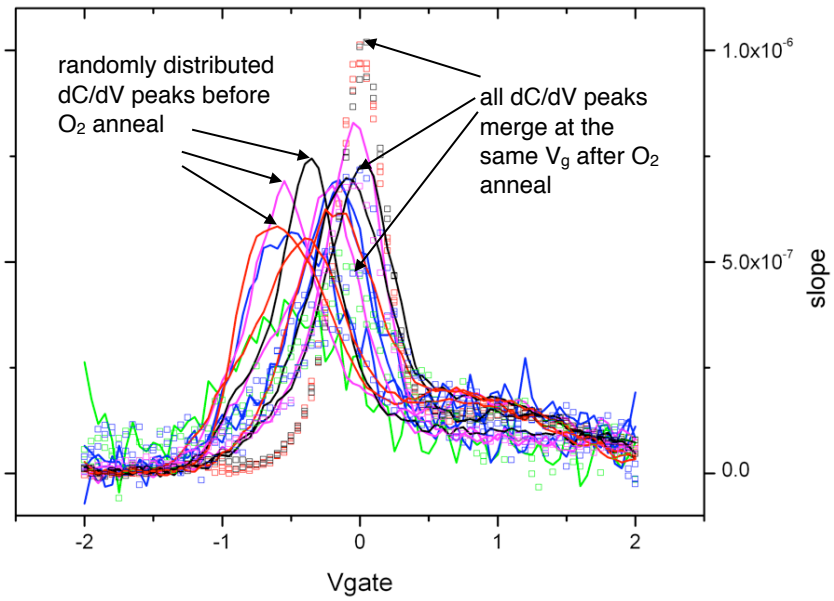


Figure 5.17 C-V slope (dC/dV) data from C-V graphs of multi-finger capacitors with gate lengths from 90 nm to 660 nm shown in fig. 5.14, highlighting the spread of data before O_2 anneal (shown as solid lines) and after the anneal (shown as scattered data points).

Further experiments were carried out exploring a comparison of a multi-finger scaled capacitor with gate length of 1 μm with a circular capacitor with 104 μm diameter. The gate-to-ohmic distance was the same between the two cases, at 1 μm. The two structures, whose C-V curves are compared in figure 5.18, show similar slope and a negative shift in threshold voltage for the scaled device, compatible with what was observed in MOSFETs when scaling gate length (figure 5.1). However, their behaviour differed for high positive voltages applied. Capacitance attained a maximum value of approximately 700 nF/cm^2 in both cases, but for the larger structure an overshoot was present in the curve for high voltage, suggesting extra charge measured, created by current tunnelling mechanism through the dielectric. Another noticeable difference was that the multi-finger C-V curve was shifted upwards by approximately 30 nF/cm^2 at all measured voltages. This suggests the presence of a considerable amount of charge which is not controlled by variations in the electric field applied to the gate. During the measurement the charge traps in the dielectric release carriers at a slower rate than that of the measurement. In equivalent circuit terms, a large parasitic capacitance is in parallel with the parallel-plate capacitance, and the parasitic response time is too long for the channel to become fully depleted during the measurement acquisition phase.

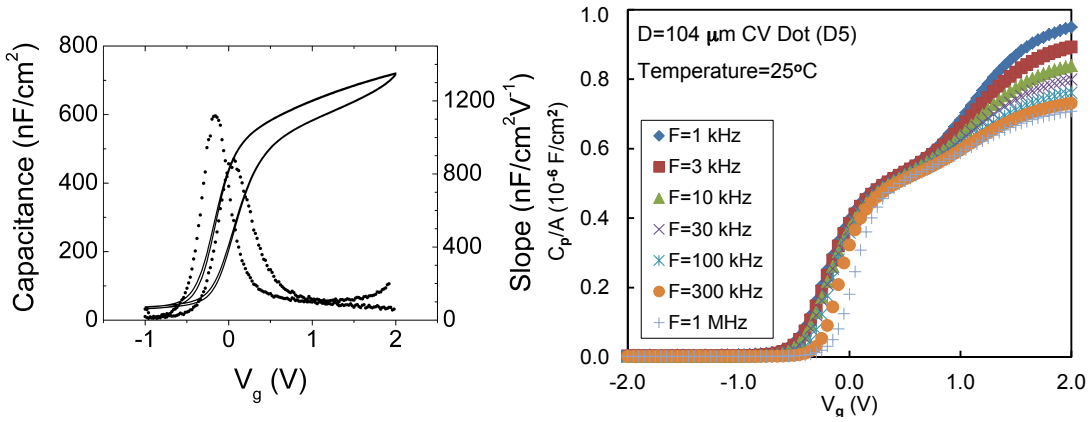


Figure 5.18 Double-sweep C-V response of multi-finger capacitors with L_g = 1 μm (left) and single sweep response of circular capacitors with L_g = 104 μm (right). The gate-ohmic distance is 1 μm for both.

It can therefore be concluded that there is indeed an increase in parasitic capacitance when the gate length is scaled from large values down to micrometer and sub-micrometer values. As figure 5.19 shows, when the diameter of the circular capacitor is decreased from 104 μm to 40 μm (a), no upward shift of the C-V curve for the smaller capacitor was observed. However, when scaling from 40 μm to 1 μm (b), similar results to those of figure 5.18 were obtained. With regard to the differences in threshold voltage between the C-V curves in figure 5.19, these are consistent with the decrease in gate length. The results in figure 5.19 once again confirm what previously stated, i.e. that an increase in parasitic capacitance exists for scaled structures. However, as shown by the similar values of slope and curve stretch-out, the effect is not detrimental to the gate stack operation.

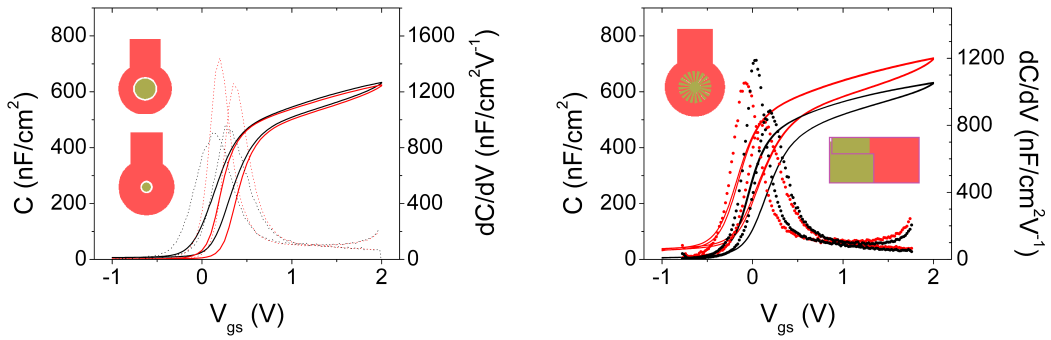


Figure 5.19 (a) High-frequency C-V response of circular capacitors with $L_{g/o}$ = 10 μm and L_g = 40 μm (black) and L_g = 100 μm (red). (b) Varying gate area whilst keeping periphery and gate-ohmic distance fixed: circular capacitor of design "Dot D" , $L_{g/o}$ = 1 μm, circumference = 2310 μm (black line); multi-finger capacitor with large area equivalent d = 40 μm, $L_{g/o}$ = 1 μm, circumference = 2310 μm (red line).

5.5 Chapter summary

This chapter described the work carried out in order to understand the threshold voltage roll-off observed in sub-micron gate length MOSFETs. A variety of capacitor designs were chosen in order to gain a better understanding of the different types of parasitic capacitance affecting device performance. Of these, one is related to the

parallel-plate capacitor periphery while the other is affected by scaling the gate length. The designs were thus chosen so that it was possible to independently vary two components of the lateral MOS structure: the geometries of the single cross-section of the device and the number of the repeats of the cross-section to complete the device. In terms of cross-section geometries, the parameters were: the length of the gate L_g and the distance between the gate and the ohmic contacts $L_{g/o}$. Varying L_g allowed to probe the contribution of the fringing capacitance to the parallel-plate capacitance due to the effect of the contact edge, whilst varying the $L_{g/o}$ saw the peripheral capacitance rise. In order to vary the number of the repeats of the MOS device cross-sections, structures with increased periphery of the gate contact were designed. An initial attempt at fabricating these structures highlighted several fabrication issues, such as gate flagging, lateral annealing of ohmic metal and problematic lift-off; these were solved by either changing the fabrication process or relaxing the design. The finalised designs were fabricated and electrical C-V measurements were carried out, allowing to assess the effect of parasitic capacitance on the gate stack performance. Great dispersion was observed in C-V data measured for capacitors not subject to oxygen anneal, which highlighted the issue of oxygen vacancies in the dielectric film having a random post-growth spatial distribution. The non-uniform spatial density of oxygen vacancies affected greatly multi-finger capacitors with sub-micrometer gate lengths, and their presence was shown to lower the value of threshold voltage. It was also shown how a thermal step in oxygen atmosphere could passivate those vacancies. Structures with different number of repeats of the same cross-section were also compared, and it was found that increasing the number of repeats does introduce an additional component to the total capacitance. The slope and stretch-out of the C-V curves, however, are not affected, suggesting this effect is not detrimental to device performance. With regard to changes in the geometry of the single cross-section, the reduction of $L_{g/o}$ yielded a higher threshold voltage, but once again slope and stretch-out of the curves were not affected. Changing gate lengths while keeping the other parameters of the cross-section geometry constant instead showed a reduction of threshold voltage with decreasing L_g (consistently with what observed in the MOSFETs) and also the introduction of an additional parasitic component to the total capacitance. This additional component prevented full depletion of carriers

from the channel by trapping charges and releasing them at a very slow rate. Once again, no changes in slope or stretch-out were observed, suggesting the quality of the interface is maintained also for scaled structures.

5.6 Chapter conclusions

The results from this chapter provided a useful insight into some of the issues encountered when scaling the MOSFETs gate length. One important observation regards the oxygen vacancies present in the dielectric film, whose spatial density varies across the wafer. This shows that dielectric films grown by MBE introduce defects that are unevenly distributed across areas as small as one square micrometer and affect the uniformity of the transistors performance. A thermal treatment of the dielectric in oxygen atmosphere was shown to help with passivating the vacancies and yielding a better uniformity.

It was also found that a reduction in $L_{g/o}$ increases the periphery capacitance and shifts the threshold voltage towards more positive values. This could be useful when fabricating scaled devices with lithographically aligned contacts on wafers providing a negative threshold voltage. However, in order to bring $L_{g/o}$ to sub-micrometer length, the issue of post-anneal lateral metal diffusion of the ohmic contacts would need to be resolved.

Due to the imperfections in the grown material stack and to the limited number of devices available for measurements, the multi-finger capacitors did not enable a full analysis of the phenomena underlying threshold voltage variation for sub-micrometer gate lengths. However, these structures proved effective at decoupling the effects of changing periphery, gate-ohmic distance and area to a limited extent. Thus it is thought that on materials less affected by epitaxial growth problems, the multi-finger capacitors could be exploited at their full potential and enable a complete analysis of the effect of scaling electronic devices.

6 Conclusions

6.1 Conclusive remarks

This thesis presented an investigation into a III-V heterostructure MOSFET with a high-mobility InGaAs channel and Ga_2O_3/GaGdO high-dielectric-constant gate oxide, designed as an n-channel candidate for high-speed low-voltage CMOS logic circuits. The aim of this work was to improve the gate field control by: a) optimisation of fabrication processes associated with the gate stack of the device, and b) investigation into parasitic elements associated with its gate and source/drain contacts layout. The former was achieved through observation of changes in capacitance-voltage measurements of MOS capacitors as the gate processing parameters were varied, with some of the processes tested on devices. To achieve the latter aim, a specialist MOS capacitor structure was designed, modifying the standard MOSCAP towards increased periphery and much smaller dimensions of its contacts and spacing between them.

Process optimisation work was focused on the gate region and involved investigation of e-beam lithography process parameters, metallisation parameters, and post-metallisation thermal treatment. During the course of the e-beam lithography process studies, it was found that some resist residue was always present in the gate region regardless of development time, exposure dose and type of surface (i.e. dielectric or semiconductor). Oxygen plasma was found to be the most effective method for achieving a metal-oxide interface free of organic contamination. A method to determine the thickness of the resist residue was developed, which could be of help in deciding the minimum RF power and ashing time required to clear residues, thus minimising induced damage on delicate III-V substrates. Variation in resist thickness was found to have a negligible effect on the gate metal linewidth achieved by lift-off, therefore the exposure dose to clear open areas can be used. However, a slight overexposure is recommended to minimise the

granularity of resist residues after the development and to produce a smooth line without metal 'flagging' after the lift-off. Next, the effect of electron irradiation was assessed through replacing electron-beam lithography patterning of the gates with shadow mask metallisation. For these, the radiation damage during the e-beam exposure was found to induce interface states. When resist-free surfaces were compared with surfaces that had been through the lithography process with an oxygen plasma ashing step, nearly identical accumulation capacitances were obtained from ashed and unprocessed resist-free surfaces, indicating that the gate oxide did not react with oxygen during the ashing process. The experiments also showed that removing the barrel oxygen plasma ashing step caused a large drop in capacitance, due to a residual PMMA layer present on the dielectric surface, whose thickness and surface roughness were reduced with increasing exposure dose. When the effect of resist residues on MOSFETs performance was assessed, the devices not subject to oxygen ashing after the gate lithography step showed a great variation in output characteristics. Transistors with an ashed gate region, thus free of resist residues, presented an improved uniformity, however their device parameters indicated plasma damage. It was thus concluded that a trade-off must therefore be sought between damage to the gate dielectric and effective removal of resist residue.

Study of the metallisation parameters made use of shadow mask deposition method as this ensured that the substrate surface condition remained constant. A variety of layer thickness ratios and metals were compared. The common feature of the experimental outcomes was a positive shift of the flat-band voltage, always accompanied by an increase in stretch-out observed in the C-V curves. When a hysteresis measurement was performed, a decrease in hysteresis was observed in conjunction with the positive flatband voltage shift, while the reverse sweep parts of C-V curves overlapped. This suggested that the changes in the forward sweep could not be ascribed to charge trapping, but only to changes in the effective workfunction, the only other variable in the processes compared. It was also observed that a negative flat-band voltage shift occurred every time Au was added to the gate metal, or the thickness of Au in the Pt/Au stack was increased. A plausible explanation to this could be the difference in the Pauling electronegativity between Pt and Au. Thus, an addition of a more electro-positive layer in the stack

strengthened the local dipoles generated on the GGO surface, reduced the effective workfunction of the gate, and caused a reduction of the flatband voltage. A reduction of the deposition rate of the bottom metal in the gate proved to achieve the most significant positive threshold voltage shift. In this case, nominally identical gate stacks were compared, thus the results were not related to changes in the effective workfunction. The improvement was ascribed to an improved uniformity of the deposited layer, as well as the slower metallisation process being less damaging.

When the influence of gate post-metallisation thermal treatment in oxygen and nitrogen on Pt/Au gates was studied, it was first of all concluded that a temperature of at least 430°C was required to affect the substrate in any way. When the sample was heated to this temperature the workfunctions of the different metal stacks all collapsed at a similar value, thus yielding a conduction band offset in the semiconductor and C-V results with little variation, yet before the annealing this was not the case. It was concluded that platinum did not act as a good diffusion barrier thicknesses less than 10 nm. Due to the high diffusion coefficient of gold, layer intermixing took place, making the contact a more ideal conductor. In this case, the initial differences in pre-anneal measurements were caused by the gate effective workfunction being tuned by the electronegativity properties of the metals. Another reason for the observed phenomenon could be the difference in the duration of the deposition of each metal layer. This is because the extent of damage caused to the substrate varied with layer thicknesses as the currents required for the deposition differed, and was reflected in the C-V response of as-deposited capacitors. During the heating, the metallisation-induced defects were annealed out, and the C-V curves merged. In the dielectric layers, the damage recovery could be due to self-rearranging of atoms taking place as a result of thermal energy supplied by the annealing process. The effect of an oxygen annealing step applied to the gate contact was also studied for two different fabrication flows, either gate-first or gate-last. It was found that the annealing of the gate improved MOSFET output characteristics for gate-first devices, but negatively impacted gate-last ones. This was related to degradation of the source/drain contacts, and it was concluded that an annealing step in oxygen atmosphere was beneficial to device performance as long as it takes place pre-ohmic contacts patterning.

The results from fabrication and measurement of scaled multi-finger capacitor structures provided a useful insight into some of the issues that could be encountered when scaling the MOSFET gate length. One important observation regarded the oxygen vacancies present in the dielectric film, whose spatial density varied across the wafer. It was revealed that dielectric films grown by MBE presented defects that were unevenly distributed across areas as small as one square micrometer and affected the uniformity of the transistors performance. A thermal treatment of the dielectric in oxygen atmosphere was shown to help with passivating the vacancies and yielded a better uniformity. The parasitic capacitances in the system, namely the contact periphery dependent gate-ohmic capacitance, played a significant role in the total gate capacitance to such an extent that reducing the distance between the gate and the source/drain ohmic contacts in the device would help with shifting the threshold voltages closely towards the designed values. This could be useful when fabricating scaled devices with lithographically aligned contacts on wafers providing a negative threshold voltage. However, in order to bring $L_{g/o}$ to sub-micrometer length, the issue of post-anneal lateral metal diffusion of the ohmic contacts would need to be resolved. Due to the imperfections in the grown material stack and to the limited number of devices available for measurements, the multi-finger capacitors did not enable a full analysis of the phenomena underlying threshold voltage variation for sub-micrometer gate lengths. However, these structures proved effective at decoupling the effects of changing periphery, gate-ohmic distance and area to a limited extent. Thus it is thought that on materials less affected by epitaxial growth problems, the multi-finger capacitors could be exploited at their full potential and enable a complete analysis of the effect of scaling electronic devices.

In summary, the results obtained from this work allowed gaining an understanding of the negative shift in transistor threshold voltage observed when decreasing the gate length, and of its causes, which prevent normally-off device operation. It was found that the main culprits for the negative threshold voltage shift were the gate metal deposition process and the choice of metal layers, namely the electronegativity properties of metals that alter the effective workfunction of the gate. However, the radiation damage induced by metallisation by e-beam evaporation could be prevented by slowing down the deposition rate, and effective

workfunction shifts could be prevented by either using a platinum-only gate, or matching the layer structure design and the actual gate metal used for the future devices. Moreover, post-metallisation thermal anneal was shown to mitigate both these effects. When the gate contact was deposited before the source/drain contacts, it caused a shift in threshold voltage towards negative values as the gate length decreased, because the ohmic contact anneal process affected the properties of the underlying material differently depending on whether it was covered with the gate metal or not. Opting for a gate-last process would thus be recommended.

6.2 Future work

The key findings of this work were supposed to constitute the base for an improved fabrication process in attempt to achieve the 0.3 V threshold voltage the III-V heterostructure MOSFET was designed for. Time constraints did not allow fabrication of the final devices, and it is therefore proposed as continuation of this work to fabricate and measure gate-last devices with Pt only gate, non-diffusing ohmic contacts, and scaled gate-ohmic gap with a 350°C thermal anneal step after the gate metallisation.

Fabrication of the specialised MOS capacitors developed as part of this work on layer structures with more uniformly distributed defects and more uniform layer thicknesses, such as inversion-channel bulk GaAs / Al_2O_3 material, would provide a further verification of their effectiveness.

The oxide-semiconductor interface quality assessment in this work has been mainly qualitative and limited to majority carrier response. For future devices, fine tuning of the fabrication process would benefit from multi-frequency capacitance and conductance measurements performed on device layers.

References

[1] G. Moore, "Cramming more components onto integrated circuits", *Electronics*, vol. 38, no. 8, p. 114, **1965**.

[2] G. Baccarani, M. Wordeman, R. Dennard, "Generalized scaling theory and its application to a 1/4 micrometer MOSFET design", *IEEE Transactions on Electron Devices*, vol. ED-31, no. 4, p. 452, **1984**.

[3] M. Schulz, "The end of the road for silicon?", *Nature*, vol. 399, p. 729, **1999**.

[4] R. Dingle, H. Stormer, A. Gossard, and W. Wiegmann, "Electron mobilities in modulation-doped semiconductor heterojunction superlattices", *Applied Physics Letters*, vol. 33, p. 665, **1978**.

[5] P. Solomon and H. Morkoc, "Modulation-doped GaAs/AlGaAs heterojunction filed-effect transistors, ultrahigh-speed device for supercomputers", *IEEE Transactions on Electron Devices*, vol. ED-31, no. 8, p. 1015, **1984**.

[6] N. Cirillo, M. Shur, and J. Abrokwah, "Inverted GaAs/AlGaAs modulation-doped field-effect transistors with extremely high transconductances", *IEEE Electron Device Letters*, vol. EDL-7, no.2, p. 71, **1986**.

[7] T. Mimura, S. Hiyamizu, T. Fuji, and K. Nanbu, "A new field-effect transistor with selectively doped GaAs/n-AlGaAs heterojunctions", *Japanese Journal of Applied Physics*, vol. 19, no. 5, p. L225, **1980**.

[8] P. Ruden, M. Shur, A. Akinwande, J. Nohava, D. Grider, and J. Baek, "AlGaAs/InGaAs/GaAs quantum well doped channel heterostructure field effect transistors", *IEEE Transactions on Electron Devices*, vol. 37, no. 10, p. 2171, **1990**.

[9] M. Passlack, R. Droopad, K. Rajagopalan, J. Abrokwah, R. Gregory, and D. Nguyen, "High mobility NMOSFET structure with high-κ dielectric", *IEEE Electron Device Letters*, vol. 26, no. 10, p. 713, **2005**.

[10] R. Hill, D. Moran, X. Li, H. Zhou, D. Macintyre, S. Thoms, A. Asenov, P. Zurcher, K. Rajagopalan, J. Abrokwah, R. Droopad, M. Passlack, and I. Thayne, "Enhancement-mode GaAs MOSFETs with an InGaAs channel, a mobility of over 5000 cm^2/Vs, and transconductance of over 475 $\mu S/\mu m$", *IEEE Electron Device Letters*, vol. 28, no. 12, p. 1080, **2007**.

[11] R. Hill, R. Droopad, D. Moran, X. Li, H. Zhou, D. Macintyre, S. Thoms, O. Ignatova, A. Asenov, K. Rajagopalan, P. Fejes, I. Thayne, and M. Passlack "1 μm gate length, $In_{0.75}Ga_{0.25}As$ channel, thin body n-MOSFET on InP substrate with transconductance of 737 $\mu S/\mu m$", *Electronics Letters*, vol. 44, p. 498, **2008**.

[12] G. Wilk, R. Wallace, and J. Anthony, "High-κ gate dielectrics: Current status and material properties considerations", *Journal of Applied Physics*, vol. 89, no. 10, p. 5243, **2001**.

[13] J. Ajayan and D. Nirmal, "A review of InP/InAlAs/InGaAs based transistors for high frequency applications", *Superlattices and Microstructures*, vol. 86, p. 1, **2015**.

[14] J. Kwo, M. Hong, A. Kortan, K. Queeney, Y. Chabal, R. Opila, D. Muller, S. Chu, B. Sapjeta, T. Lay, J. Mannaerts, T. Boone, H. Krautter, J. Krajewski, A. Sergnt, and J. Rosamilia, "Properties of high-κ gate dielectrics Gd_2O_3 and Y_2O_3 for Si", *Journal of Applied Physics*, vol. 89, no. 7, p. 3920, **2001**.

[15] P. Rouffignac, J-S. Park, and R. Gordon, "Atomic layer deposition of Y_2O_3 thin films from yttrium tris(diisopropylacetamidinate) and water", *Chemistry of Materials*, vol. 17, p. 4808, **2005**.

[16] S. Duenas, H. Castan, H. Garcia, A. Gomez, L. Bailon, K. Kukli, T. Hatanpaa, J. Lu, M. Ritala, and M. Leskela, "Electrical properties of atomic-layer-deposited thin gadolinium oxide high-κ gate dielectrics", *Journal of The Electrochemical Society*, vol 154, no. 10, p. G207, **2007**.

[17] R. Droopad, K. Rajagopalan, J. Abrokwah, L. Adams, N. England, D. Uebelhoer, P. Fejes, P. Zurcher, and M. Passlack, "Development of GaAs-based MOSFET using molecular beam epitaxy", *Journal of Crystal Growth*, vol. 301-302, p. 139, **2007**.

[18] B. Gila, K. Lee, W. Johnson, F. Ren, C. Abernathy, S. Pearton, M. Hong, J. Kwo, J. Mannaerts, and K. Anselm, "A comparison of gallium gadolinium oxide and gadolinium oxide for use as dielectrics in GaN MOSFETs", *Proceedings of IEEE/ Cornell Conference on High Performance Devices*, p. 182, **2000**.

[19] M. Hong, J. Mannaerts, J. Bower, J. Kwo, M. Passlack, W. Hwang, L. Tu, "Novel $Ga_2O_3(Gd_2O_3)$ passivation techniques to produce low D_{it} oxide-GaAs interfaces", *Journal of Crystal Growth*, vol. 175-176, p. 422, **1997**.

[20] P. Ye, G. Wilk, B. Yang, J. Kwo, S. Chu, S. Nakahara, H. Gossmann, J. Mannaerts, M. Hong, K. Ng, and J. Bude, "GaAs metal-oxide-semiconductor field-effect transistor with nanometer-thin dielectric grown by atomic layer deposition", *Applied Physics Letters*, vol. 83, no. 1, p. 180, **2003**.

[21] V. Djara, K. Cherkaoui, M. Schmidt, S. Monaghan, E. O'Connor, I. Povey, D. O'Connell, M. Pemble, and P. Hurley, "Impact of forming gas annealing on the performance of surface-channel InGaAs MOSFETs with an ALD Al_2O_3 gate dielectric", *IEEE Transactions on Electron Devices*, vol. 59, no. 4, p. 1084, **2012**.

[22] C. Kim, S. Cho, M. Cho, K. Chung, D. Suh, D. Ko, C. An, H. Kim, and J. Lee, "Changes in the structure of an atomic-layer-deposited HfO_2 film on a GaAs (100) substrate as a function of postannealing temperature", *Applied Physics Letters*, vol. 95, p. 042903-1, **2009**.

[23] R. Pandey, H. Madan, H. Liu, V. Chobpattana, M. Barth, B. Rajamohanan, M. Hollander, T. Clark, K. Wang, J. Kim, D. Gundlach, K. Cheung, J. Suehle, R. Engel-Herbert, S. Stemmer, and S. Datta, "Demonstration of p-type InGaAs/GaAsSb and n-type GaAsSb/InGaAs complimentary heterojunction vertical tunnel FETs for ultra-low power logic", *Symposium on VLSI Technology Digest of Technical Papers*, p. T206, **2015**.

[24] M. Holland, C. Stanley, W. Reid, I. Thayne, G. Paterson, A. Long, P. Longo, J. Scott, A. Craven, and R. Gregory, "GdGaO: A gate dielectric for GaAs metal-oxide-semiconductor field-effect transistors", *Journal of Vacuum Science & Technology B*, vol. 25, p. 1024, **2007**.

[25] M. Passlack, J. Abrokwah, R. Droopad, Z. Yu, C. Overgaard, S. Yi, M. Hale, J. Sexton, and A. Kummel, "Self-aligned GaAs p-channel enhancement mode MOS heterostructure field-effect transistor", *IEEE Electron Device Letters*, vol. 23, no. 9, p. 508, **2002**.

[26] G. Paterson, J. Wilson, D. Moran, R. Hill, A. Long, I. Thayne, M. Passlack, and R. Droopad, "Gallium oxide (Ga_2O_3) on gallium arsenide - a low defect, high-κ system for future devices", *Materials Science and Engineering B*, vol. 135, p. 277, **2006**.

[27] M. Passlack, "Methodology for development of high-κ stacked gate dielectrics on III-V semiconductors", *Chapter 12 in book 'Materials Fundamentals of Gate Dielectrics'*, p. 403, **2005**.

[28] J. Swanson, "Compound semiconductor/insulator device interfaces - present status and requirements", *Microelectronic Engineering*, vol. 22, p. 111, **1993**.

[29] D. Fleetwood, P. Winokur, R. Reber, T. Meisenheimer, J. Schwank, M. Shaneyfelt, and L. Riewe, "Effects of oxide traps, interface traps, and "border traps" on metal-oxide-semiconductor devices", *Journal of Applied Physics*, vol. 73, no. 10, p. 5058, **1993**.

[30] A. Revesz and R. Evans, "Kinetics and mechanism of thermal oxidation of silicon with special emphasis on impurity effects", *Journal of Physics and Chemistry of Solids*, vol. 30, p. 551, **1969**.

[31] Y. Nishi, "Study of silicon-silicon dioxide structure by electron spin resonance", *Japanese Journal of Applied Physics*, vol. 10, no. 1, p. 52, **1970**.

[32] W. He and T. Ma, "Inelastic electron tunneling spectroscopy study of ultrathin HfO_2 and HfAlO", *Applied Physics Letters*, vol. 83, no. 13, p. 2605, **2003**.

[33] J. Hackley, J. Demaree, and T. Gougousi, "Interface of atomic layer deposited HfO_2 films on GaAs (100) surfaces", *Applied Physics Letters*, vol. 92, p. 162902-1, **2008**.

[34] P. Laukkanen, M. Punkkinen, J. Lang, M. Tuominen, M. Kuzmin, V. Tuominen, J. Dahl, J. Adell, J. Sadowski, J. Kanski, V. Polojarvi, J. Pakarinen, K. Kokko, M. Guina, M. Pessa, and I. Vayrynen, "Ultrathin (1x2)-Sn layer on GaAs(100) and InAs(100) substrates: A catalyst for removal of amorphous surface oxides", *Applied Physics Letters*, vol. 98, p. 231908-1, **2011**.

[35] R. Rahbe, B. Pajot, J. Chevallier, A. Marbeuf, R. Logan, and M. Gavand, "Hydrogen diffusion and acceptor passivation in p-type GaAs", *Journal of Applied Physics*, vol. 73, p. 1723, **1993**.

[36] G. Dalapati, Y. Tong, W. Loh, H. Mun, and B. Cho, "Electrical and interfacial characterisation of atomic layer deposited high-κ gate dielectrics on GaAs for advanced CMOS devices", *IEEE Transactions on Electron Devices*, vol. 54, no. 8, p. 1831, **2007**.

[37] K. Queeney, N. Herbots, J. Shaw, V. Atluri, and Y. Chabal, "Infrared spectroscopic analysis of an ordered Si/SiO2 interface", *Applied Physics Letters*, vol. 84, no. 4, p. 493, **2004**.

[38] W. Songprakob, R. Zallen, W. Liu, and K. Bacher, "Infrared studies of hole-plasmon excitations in heavily-doped p-type MBE-grown GaAs:C", *Physical Review B*, vol. 62, no. 7, p. 4501, **2000**.

[39] E. Hellman and J. Harris, "Infra-red transmission spectroscopy of GaAs during molecular beam epitaxy", *Journal of Crystal Growth*, vol. 81, p. 38, **1987**.

[40] T. Eickhoff, S. Morley, D. Zahn, W. Richter, D. Woolf, D. Westwood, and R. Williams, "MBE grown GaAs on Si(100) studied by infrared spectroscopy", *Materials Research Society Proceedings*, vol. 160, **1989**.

[41] G. Gledhill, R. Newman, and J. Collins, "High resolution FTIR measurements of LVM-structure in GaAs", *Mikrochimica Acta*, vol. I, p. 411, **1988**.

[42] F. Li, B. Bayer, S. Hofmann, J. Dutson, S. Wakeham, M. Thwaites, W. Milne, and A. Flewitt, "High-κ (κ = 30) amorphous hafnium oxide films from high rate room temperature deposition", *Applied Physics Letters*, vol. 98, p. 252903, **2011**.

[43] M. Passlack, M. Hong, and J. Mannaerts, "Quasistatic and high frequency capacitance-voltage characterisation of Ga_2O_3-GaAs structures fabricated by in situ molecular beam epitaxy", *Applied Physics Letters*, vol. 68, p. 1099, **1996**.

[44] E. O'Connor, S. Monaghan, K. Cherkaoui, I. Povey, and P. Hurley, "Analysis of the minority carrier response of n-type and p-type $Au/Ni/Al_2O_3/In_{0.53}Ga_{0.47}As/InP$ capacitors following an optimised $(NH_4)_2S$ treatment", *Applied Physics Letters*, vol. 99, p. 212901-1, **2011**.

[45] A. Penumatcha, S. Swandono, and J. Cooper, "Limitations of the high-low C-V technique for MOS interfaces with large time constant dispersion", *IEEE Transactions on Electron Devices*, vol. 60, no. 3, p. 923, **2013**.

[46] B. O'Sullivan, P. Hurley, F. Cubaynes, P. Stolk, and F. Widdershoven, "Flat band voltage shift and oxide properties after rapid thermal annealing", *Microelectronic Reliability*, vol. 41, p. 1053, **2001**.

[47] E. Nicollian and A. Goetzberger, "The Si-SiO_2 interface - electrical properties as determined by the metal-insulator-silicon conductance technique", *The Bell System Technical Journal*, vol. 46, no. 6, p. 1055, **1967**.

[48] J. Piscator, B. Raeissi, and O. Engstrom, "The conductance method in a bottom-up approach applied on hafnium oxide / silicon interfaces", *Applied Physics Letters*, vol. 94, p. 213507-1, **2009**.

[49] T. Ma and P. Dressendorfer, "Ionizing radiation effects in MOS devices and circuits", Chapter 7, John Wiley & Sons, **1989**.

[50] C. Chen and E. Hu, "Metallisation-induced damage in III-V semiconductors", *Journal of Vacuum Science & Technology B*, vol. 16, p. 3354, **1998**.

[51] G. Burek, Y. Hwang, A. Carter, V. Chobpattana, J. Law, W. Mitchell, B. Thibeault, S. Stemmer, and M. Rodwell, "Influence of gate metallisation processes on the electrical characteristics of high-κ/InGaAs interfaces", *Journal of Vacuum Science and Technology B*, vol. 29, no. 4, p. 040603-1, **2011**.

[52] F. Auret and P. Mooney, "Deep levels introduced during electron-beam deposition of metals on n-type silicon", *Journal of Applied Physics*, vol. 55, p. 988, **1984**.

[53] W. Cabrera, M. Halls, and Y. Chabal, "Chemical nature and control of high-κ dielectric / III-V interfaces", *ECS Transactions*, vol. 66, no. 6, p. 65, **2015**.

[54] C. Weiland, P. Lysaght, J. Price, J. Huang, and J. Woicik, "Hard x-ray photoelectron spectroscopy study of As and Ga out-diffusion in InGaAs/Al_2O_3 film systems", *Applied Physics Letters*, vol. 101, p. 061602-1, **2012**.

[55] H. Dong, W. Cabrera, R. Galatage, KC. Santosh, B. Brennan, X. Qin, S. McDonnell, D. Zhernokletov, C. Hinkle, K. Cho, Y. Chabal, and R. Wallace, "Indium diffusion through high-κ dielectrics in high-k/InP stacks", *Applied Physics Letters*, vol. 103, p. 061601, **2013**.

[56] B. Brennan, M. Milojevic, C. Hinkle, F. Aguirre-Tostado, G. Hughes, R. Wallace, "Optimisation of the ammonium sulphide $(NH_4)_2S$ passivation process on InGaAs", *Applied Surface Science*, vol. 257, no. 9, p. 4082, **2011**.

[57] J. Chung, J. Roberts, E. Piner, and T. Palacios, "Effect of gate leakage in the subthreshold characteristics of AlGaN/GaN HEMTs", *IEEE Electron Device Letters*, vol. 29, no. 11, p. 1196, **2008**.

[58] Y. Hori, Z. Yatabe, T. Hashizume, "Characterisation of interface states in Al_2O_3/AlGaN/GaN structures for improved performance of high-electron-mobility transistors", *Journal of Applied Physics*, vol. 114, p. 244503-1, **2013**.

[59] H. Trinh, E. Chang, P. Wu, Y. Wong, C. Chang, Y. Hsieh, C. Yu, H. Nguyen, Y. Lin, K. Lin, and M. Hudait, "The influences of surface treatment and gas annealing conditions on the inversion behaviours of the atomic-layer-deposition Al_2O_3/n-InGaAs metal-oxide-semiconductor capacitor", *Applied Physics Letters*, vol. 97, p. 042902-1, **2010**.

[60] A. Jaksic, G. Ristic, M. Pejovic, A. Mohammadzadeh, C. Sudre, and W. Lane, "Gamma-ray irradiation and post-irradiation responses of high dose range RADFETs", *6th European Conference on Radiation and Its Effects on Components and Systems*, p. 57, **2001**.

[61] T. Ma, "Interface trap transformation in radiation or hot-electron damaged MOS structures", *Semiconductor Science and Technology*, vol. 4, p. 1061, **1989**.

[62] D. Shahrjerdi, M. Oye, A. Holmes, and S. Banerjee, "Unpinned metal gate/high-κ GaAs capacitors: Fabrication and characterisation", *Applied Physics Letters*, vol. 89, no. 4, p. 043501-1, **2006**.

[63] D. Kennedy and R. O'Brien, "On the measurement of impurity atom distributions by the differential capacitance technique" (Letter to the Editor), *IBM Journal of Research and Development*, vol. 13, no. 2, p. 212, **1969**.

[64] M. Hauptmann, H. Struyf, S. Gendt, C. Glorieux, and S. Brems, "Importance of bubble size control in ultrasonic surface cleaning by pulsed high-frequency sound fields", *ECS Journal of Solid State Science and Technology*, vol. 3, no. 1, p. N3032, **2014**.

[65] R. Mints, V. Kortov, V. Melekhin, E. Kislitsin, E. Plekhanova, and G. Peshchin, "Effect of deformation on the work function and exoemission from the surface of a noble metal", *Soviet Physics Journal*, vol. 13, issue 7, p. 866, **1970**.

[66] C. Quinn and M. Roberts, "Chemisorption of oxygen and subsequent processes on metal films: work function measurements", *Transactions of the Faraday Society*, vol. 60, p. 899, **1964**.

[67] H. Jagannathan, V. Narayanan, and S. Brown, "Engineering high dielectric constant materials for band-edge CMOS applications", *ECS Transactions*, vol. 16, no. 5, p. 19, **2008**.

[68] K. Schaeffer, D. Gilmer, C. Capasso, S. Kalpat, B. Taylor, M. Raymond, D. Triyoso, R. Hegde, S. Samavedam, and B. White Jr, "Application of group electronegativity concepts to the effective work functions of metal gate electrodes on high-κ gate oxides", *Microelectronic Engineering*, vol. 84, p. 2196, **2007**.

[69] I. Maximov, A. Zakharov, T. Holmqvist, L. Montelius, and I. Lindau, "Investigation of polymethylmethacrylate resist residues using photoelectron microscopy", *Journal of Vacuum Science and Technology B*, vol. 20, no. 3, p. 1139, **2002**.

[70] S. Yasin, D. Hasko, and H. Ahmed, "Fabrication of < 5 nm width lines in poly(methylmethacrylate) resist using a water:isopropyl alcohol developer and ultrasonically-assisted development", *Applied Physics Letters*, vol. 78, p. 2760, **2001**.

[71] S. Chou, P. Krauss, and P. Renstrom, "Imprint of sub-25 nm vias and trenches in polymers", *Applied Physics Letters*, vol. 67, p. 3114, **1995**.

[72] A. Chen, S. Chua, P. Chen, X. Chen, and L. Jian, "Fabrication of sub-100 nm patterns in SiO_2 templates by electron-beam lithography for the growth of periodic III-V semiconductor nanostructures", *Nanotechnology*, vol. 17, p. 3903, **2006**.

[73] M. Passlack, O. Hartin, M. Ray, N. Medendorp, "Enhancement mode metal-oxide-semiconductor field effect transistor", *United States Patent*, No: US 6,963,090 B2, Nov. 8, **2005**.

[74] D. Moran, R. Hill, X. Li, H. Zhou, D. Mcintyre, S. Thoms, R. Droopad, P. Zurcher, K. Rajagopalan, J. Abrokwah, M. Passlack, and I. Thayne, "Sub-micron, metal gate, high-κ dielectric, implant-free, enhancement-mode III-V MOSFETs", *ESSDERC 2007*, p. 466, **2007**.

[75] W. Chong, M. Xiaohua, F. Qian, H. Yue, Z. Jincheng, and M. Wei, "Development and characteristics analysis of recessed-gate MOS HEMT", *Journal of Semiconductors*, vol. 30, no. 5, p. 054002-1, **2009**.

[76] C. Lin and C. Lee, "Comparison of Au/Ni/Ge, Au/Pd/Ge, and Au/Pt/Ge ohmic contacts to n-type GaAs", *Journal of Applied Physics*, vol. 67, no. 1, p. 260, **1990**.