



CHALMERS
UNIVERSITY OF TECHNOLOGY

Numerical Simulations of Device Scaling of a Pseudomorphic InP HEMT

Master's thesis in wireless, photonics and space engineering

MARCUS AHLSTRAND

THESIS FOR THE DEGREE OF MASTER OF SCIENCE IN
WIRELESS, PHOTONICS AND SPACE ENGINEERING

Numerical Simulations of Device Scaling of a Pseudomorphic InP
HEMT

MARCUS AHLSTRAND



Examiner
JAN GRAHN

Microwave Electronics Laboratory
Department of Microtechnology and Nanoscience – MC2
Chalmers University of Technology
Gothenburg, Sweden 2014

**Numerical Simulations of Device Scaling of a Pseudomorphic InP
HEMT**

MARCUS AHLSTRAND

© Marcus Ahlstrand, 2014

Chalmers University of Technology
Department of Microtechnology and Nanoscience – MC2
Microwave Electronics Laboratory
SE-412 96 Gothenburg, Sweden
+ 46 (0)31-772 1000

Printed by Chalmers Reproservice
Gothenburg, Sweden 2014

Abstract

A Technology CAD (TCAD) model of a pseudomorphic indium phosphide high electron mobility transistor (HEMT) optimized for cryogenic operation has been developed. The model has been used to investigate the prospects of improving the radio frequency performance of the device by means of device scaling, specifically through the scaling of the gate length and the distance between the gate electrode and the channel.

The model describes the movement of charge carriers within the HEMT structure using an isothermal transport model, where carrier velocities are limited by phonon scattering and high-field velocity saturation. The undesired transport of electrons from the gate electrode into the channel is accounted for through the modeling of thermionic emission and quantum tunneling.

The lateral scaling of the device model suggests that advantages such as a 92 % increase of the unity current gain cut-off frequency f_T can be achieved by reducing the gate length from its original 130 nm to 10 nm. Scaling of the gate-channel separation from 11 nm to 8 nm was attempted as a measure to reduce the impact of various short-channel effects (SCEs), but was found to be insufficient.

A drastically lowered peak transconductance and threshold voltage at gate lengths below 70 nm suggests that further engineering of the HEMT structure is necessary in order to mitigate the SCEs and to improve the small-signal power gain. Proper dimensions for future transistor designs are ultimately decided by requirements in terms of the optimal low-noise bias point and the suppression of leakage currents.

Contents

1	Introduction	1
2	Pseudomorphic InP HEMTs	3
2.1	Principles of operation	5
2.2	Current transport	5
3	Numerical Device Modeling	9
3.1	Reference HEMT	9
3.2	Numerical device simulations	10
3.3	Virtual device definition	12
3.3.1	Cross-sectional geometry	12
3.3.2	Doping profiles	12
3.3.3	Contact definitions	14
3.3.4	Activated device physics equations	15
3.4	Drain current calibration at DC	15
3.4.1	Drain current at weak lateral electric fields	16
3.4.2	Drain current at strong lateral electric fields	21
3.4.3	Gate leakage currents	24
3.5	Microwave characterization	27
4	Device scaling results	31
4.1	Scaling of the gate length	31
4.1.1	Current-voltage characteristics	31
4.1.2	Subthreshold leakage characteristics	34
4.1.3	Small-signal characteristics	36
4.2	Scaling of barrier thickness	38
4.2.1	Current-voltage characteristics	38
4.2.2	Subthreshold leakage	41
4.2.3	Small-signal characteristics	41
5	Conclusion and future work	47
A	Gate Experiments	55
A.1	Fabrication	55
A.1.1	Preprocessing of the epi-wafer	55
A.1.2	Formation of mask	56
A.1.3	Development of e-beam resist	56
A.1.4	Recess cavity etching	56

A.1.5	Metal deposition	57
A.1.6	Lift-off	57
A.2	Verification and inspection	57

Chapter 1

Introduction

In space communication and radio astronomy, the information of interest is often received as extremely weak electromagnetic signals. In order to make the nature of these signals distinguishable from other disturbances, the noise temperature of a receiver on earth needs to be kept at an absolute minimum. Even in the complete absence of electromagnetic interference from, for instance, celestial objects such as the sun (not to mention earth's own atmosphere), the noise temperature of a receiving antenna will at the very least be determined by the cosmic microwave background radiation [1]. However, its contribution to the equivalent noise temperature of the receiver is rather insignificant as compared to that of the radiofrequency pre-amplifier, which will constitute a major bottleneck in a low-noise receiver. As the emitted radiation from several natural phenomena of scientific interest, such as black-body radiation, are of sub-millimeter wavelengths there is a demand for low-noise amplifiers (LNAs) operable at the intermediate frequencies of these types of receivers.

An active microwave device that has received significant attention in this respect is the indium phosphide (InP) high electron mobility transistor (HEMT). This type of field-effect transistor (FET), which is pseudomorphic and contains an InAlAs/InGaAs/InP heterostructure was demonstrated as early as in 1987 [2] and has demonstrated excellent microwave properties. Further improvements to the original design, such as numerous modifications to the epitaxial structure [3] and the addition of a low-resistance T-gate, have led it to become an established low-noise active component.

The active components in low-noise receivers have historically been cooled down to cryogenic temperatures in order to improve current transport and to reduce the impact of various parasitics [1]. In 2013, Chalmers University of Technology was able to demonstrate a custom in-house process optimized for cryogenic operation, resulting in an InP HEMT possessing the lowest internal noise figure recorded at the date of publication [4]. However, these ultra-low noise properties have currently only been demonstrated at frequencies below 20 GHz [4] but recent work on device scalability [5, 6] suggests that the upper frequency limit can be increased through careful gate scaling.

The purpose of this work is to develop a physical device model of the Chalmers HEMT which models the deteriorating mechanisms that are expected as the gate length is reduced. Specifically, measurements on two physical InP HEMTs, both utilizing two gate fingers with a length of 130 nm each, will act as references [4].

The resulting model, as well as an analysis of its inaccuracies and drawbacks, will then be used to estimate the impact of the geometrical scaling of the device on key properties such as the current pinch-off conditions, subthreshold leakage and small-signal power gain. A secondary objective is to investigate the possibilities of manufacturing the required gates using e-beam lithography. An overview of the latter procedure is presented in appendix A.

While the original device was specifically designed for cryogenic operating conditions, the model presented in this work will be based around a HEMT operating at room temperature. Operation at lower temperatures would call for the use of more sophisticated model to describe charge conservation and current transport, models which aim to take the quantized nature of the charge carriers into account [7]. The lack of reliable information regarding the physical parameters required for these model together with the introduction of a great number of unknown variables will increase the complexity of the problem at hand, but will not certainly improve the model's ability to accurately predict its electrical behavior.

In order to preserve the electrical integrity of the HEMT, a reduction of the gate length needs to be followed by a reduction of the separation between gate and channel. Further scaling will require a corresponding scaling of other vertical features of the epistructure [5]. However, the latter will not be considered in this work as it calls for the redesign of the epistructure (which is produced by an outside manufacturer) as opposed to the scaling of the gate-channel separation which can be controlled through in-house processing.

The thesis begins with a chapter concerning the operational principles behind pseudomorphic InP HEMTs and the underlying theory of its current transport. The procedure of developing a numerical representation of the physical device from this theory and associated physical data is outlined in chapter 3. The subsequent device scaling study and the results acquired from it is presented in chapter 4, followed by a summary in chapter 5 of the main conclusions which are drawn.

Chapter 2

Pseudomorphic InP HEMTs

A conceptual schematic of a pseudomorphic InP HEMT is shown in figure 2.1. The feature that primarily distinguishes the HEMT from other types of FETs is the use of a heterostructure consisting of a wide bandgap and a narrow bandgap semiconductor. The former of the two is known as a barrier due to its role in electrically isolating the gate terminal from the channel, and is in the case of the InP HEMT composed of indium aluminum arsenide (InAlAs). Currents between drain and source are conducted through a channel made out of an indium-rich alloy of indium gallium arsenide (InGaAs) where contact with the channel is made via highly n -doped, low-resistance drain and source contacts.

In order to allow current conduction, the barrier is doped with donor impurities which donate their electrons to the surrounding material. A number of advantages, such as reduced short-channel effects (SCEs) and the prospect to use a thinner barrier, can be achieved through the addition of a thin, planar donor sheet known as a δ -doping [8]. The donor sheet is separated from the channel by a spacer which prevents the wave functions of the electrons in the channel to interact with the donor impurities [9].

Due to differences in electron affinity between the barrier and the channel material, a quantum well is formed allowing mobile electrons to migrate into the channel. There they reside as a thin, high density sheet of electrons known as a two-dimensional electron gas (2DEG). As the concentration of free electrons occur over a narrow region close to the heterointerface, it is generally quantified in terms of an equivalent sheet charge density n_s , expressed in units of cm^{-2} .

As the channel is undoped, electrons traveling through the 2DEG are spatially separated from the donor impurities and therefore do not suffer from impurity scattering. The formation of a 2DEG causes the channel to experience an overall increase in electron mobility, which characterizes how fast an electron is allowed to drift through the InGaAs lattice under the influence of an electric field. The electron mobility μ within the 2DEG is limited mainly by the scattering of electrons by acoustic phonons, an effect which can be alleviated through a proper choice of channel material.

InGaAs with its indium (In) content exhibits a relatively high electron mobility in this respect as compared to other common semiconductor materials

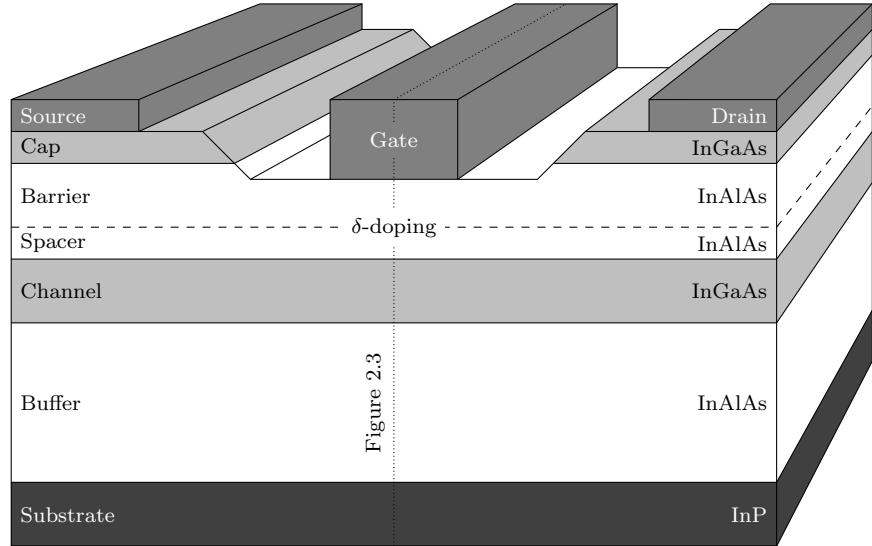


Figure 2.1: Conceptual schematic of the pseudomorphic InP HEMT. The figure indicates the placement of the cross-section from where the band structures and electron distributions featured in figure 2.3 were extracted. The dimensions of the structure (particularly the vertical ones) are exaggerated.

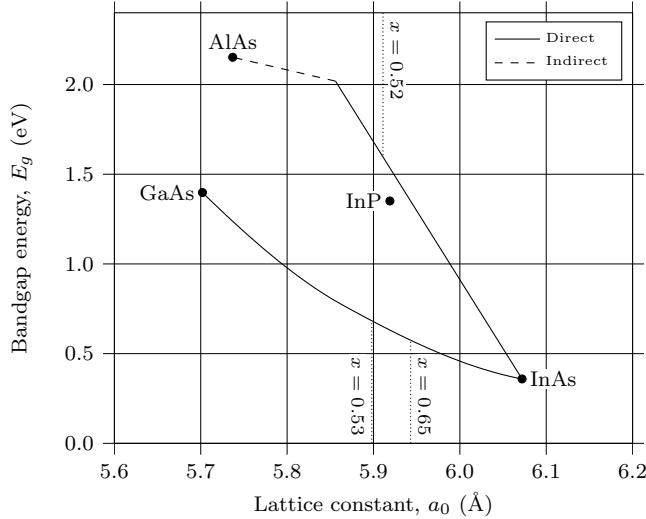


Figure 2.2: Bandgap energy and lattice constants of the III-V semiconductors present in the InP HEMT structure [9]. The values are for room temperature (300 K).

[9]. InGaAs is a ternary compound between indium arsenide (InAs) and gallium arsenide (GaAs) and in order to achieve successful crystal growth without major lattice defects, the channel material should ideally be lattice-matched to the underlying InP substrate. In order to allow a larger percentage of InAs, the pseudomorphic HEMT circumvents this criteria by the use of a buffer layer which gradually overgrows the lattice dislocations. As figure 2.2 further suggests, InGaAs also exhibits a smaller energy band gap for higher levels of InAs which has the additional advantage of increasing the confinement of electrons to the quantum well.

2.1 Principles of operation

Figure 2.3 illustrates the computed energy band structure of the Chalmers HEMT at different gate voltage biases. The operational principle of the HEMT is to control the amount of charges (more specifically, the amount of electrons) in the 2DEG which are able to participate in current transport. The current is driven through the channel by the application of an electric field between drain and source, in other words a drain voltage.

The Chalmers HEMT is a depletion mode device, meaning that the channel is populated by electrons even at zero gate bias (figure 2.3c and 2.3d). However, as a sufficiently large negative bias is applied to the gate (figure 2.3a), the sheet charge density of the electron gas diminishes, leading to reduced channel conductivity. The channel is in this state said to be pinched off.

Correspondingly, a positive bias applied to the gate causes the sheet charge density to increase. However, for very large forward biases (figure 2.3e) a potential well will take shape at the location of the δ -doping which will result in a parasitic parallel conduction path in the barrier layer. This type of conduction is not desirable as the electrons conducted through this path suffer from impurity scattering, resulting in an increased level of noise in the overall drain current. As opposed to other FETs, the ability to control n_s is expected to decrease at large enough positive gate biases.

2.2 Current transport

The general purpose of semiconductor device simulators is to describe the transport of charges from one terminal to another due to various conduction mechanisms. The conservation of charges within a semiconductor material is readily expressed in terms of the electrostatic continuity equations, which are

$$\nabla \cdot \vec{J}_n = q \left(R_{\text{net}} + \frac{\partial n}{\partial t} \right) \quad (2.1)$$

for electrons and

$$-\nabla \cdot \vec{J}_p = q \left(R_{\text{net}} + \frac{\partial p}{\partial t} \right) \quad (2.2)$$

for holes, respectively. In essence, the equations state that the net flow of current out of an infinitesimally small region is either due to generation or recombination

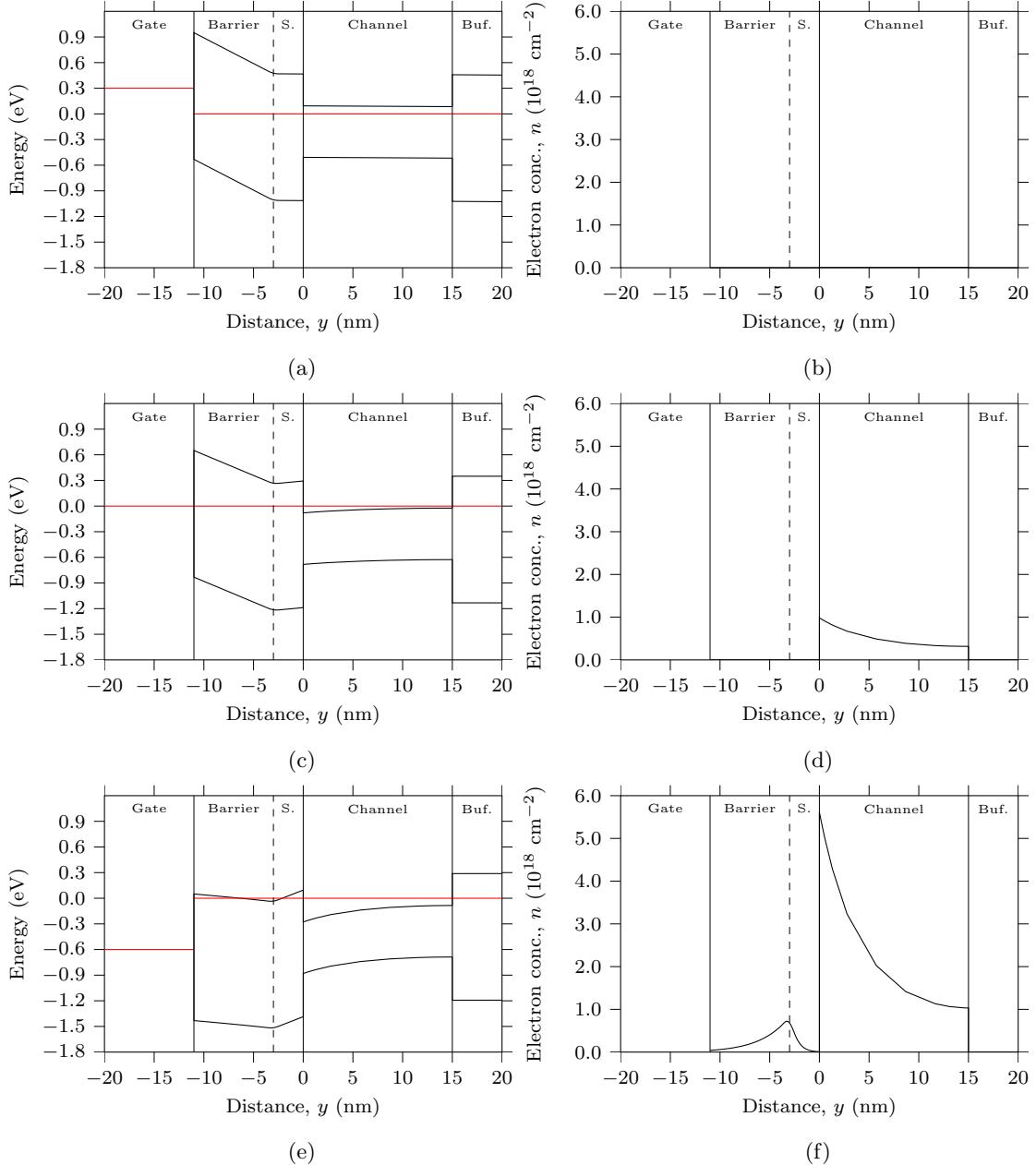


Figure 2.3: Computed energy band structures and electron distributions underneath the gate of the HEMT treated in this thesis. The distance y is measured towards the substrate from the InAlAs/InGaAs heterointerface whose properties are described in section 3.3. Figure (a) and (b) features the device behavior under a reverse gate bias of 0.3 V, figure (c) and (d) under zero bias and figure (e) and (f) under a forward gate bias of 0.5 V.

of electron-hole pairs (described by the carrier generation rate, R_{net}) or by other means (described by the free carrier time-derivatives, $\partial n/\partial t$ and $\partial p/\partial t$).

In classical mechanics, the electron and hole current densities \vec{J}_n and \vec{J}_p of a system in a state of non-equilibrium are commonly described by the Boltzmann transport equations (BTEs), which takes into account the statistical movement of charges due to electrostatic gradients [9]. Due to their importance in science and engineering, the equations have been subject to a number of discrete simplifications tailored to specific problems within the field of semiconductor simulations.

An isothermal transport model is the drift-diffusion model, which computes J_n as

$$\vec{J}_n = \mu_n \left(n \nabla E_C - \frac{3}{2} n k T \nabla \ln m_n \right) + D_n (\nabla n - n \nabla \ln \gamma_n) \quad (2.3)$$

and J_p as

$$\vec{J}_p = \mu_p \left(p \nabla E_V + \frac{3}{2} p k T \nabla \ln m_p \right) - D_p (\nabla p - p \nabla \ln \gamma_p). \quad (2.4)$$

At the right-hand side of (2.3) and (2.4), the first term concerns the contribution to the current due to the drift of charges. In this term, the carrier drift arising from the application of an external electric field is considered as well as drift occurring due to spatial variations in electron affinity and band gap energy, collectively described by the conduction and valence band gradients ∇E_C and ∇E_V . The term contains an additional contribution due to random thermal movement of electrons with masses m_e and holes with masses m_p at a uniform temperature T .

The second term represents the contribution due to the diffusion of charge carriers (characterized by the material's diffusion constant D for each type of carrier), which are both due to spatial variation in carrier concentration and the availability of free energy states at high carrier densities. The latter is characterized by the constant γ which is a proportionality constant between the amount of free energy states as predicted by Boltzmann statistics and those predicted by Fermi statistics [7].

The band structure of the heterojunction causes the quantized nature of the 2DEG to have an impact on device performance. Therefore, for devices with small active regions such as the HEMT, a quantum corrected transport model known as the hydrodynamic transport model is typically used [7]. The hydrodynamic equations are similar to the drift-diffusion equations apart from the addition of terms which takes the diffusion due to carrier temperature gradients into account.

Chapter 3

Numerical Device Modeling

This chapter describes the development of a numerical device model of a physical 130 nm pseudomorphic InP HEMT. The presentation of the modeling process has been divided into five sections. The first three sections concern the reference device and its portrayal to the Technology CAD (TCAD) environment. Section four and five describes the calibration of the model against direct current (DC) and alternating current (AC) reference data.

3.1 Reference HEMT

The device model developed in this thesis was based primarily on data acquired through DC and radio frequency (RF) measurements on two physical InP HEMTs in a Lakeshore model CRX-4K cryogenic probe station [10]. Both reference devices had been manufactured from a single heteroepitaxial structure, which was grown on an InP substrate by means of molecular beam epitaxy (MBE). An outline of the structure is provided in table 3.1.

However, the widths of the reference devices differed, since the RF characterization of the device called for a higher gate impedance as a means to reduce measurement uncertainties. As such, a device with a $2 \times 100 \mu\text{m}$ wide gate was used in the RF characterization of the device and another one with a $2 \times 10 \mu\text{m}$ wide gate was used in the DC characterization.

The devices are pseudomorphic in the sense that an intermediate buffer layer is used to reduce the number of dislocations caused by lattice mismatch between the

Table 3.1: Overview of the heteroepitaxial structure of the Chalmers HEMT.

Name	Material	Doping	Thickness
Cap	$\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$	$5 \times 10^{19} \text{ cm}^{-3}$ (Si)	20 nm
Barrier	$\text{In}_{0.52}\text{Al}_{0.48}\text{As}$		11 nm
δ -doping		$5 \times 10^{12} \text{ cm}^{-2}$ (Si)	
Spacer	$\text{In}_{0.52}\text{Al}_{0.48}\text{As}$		3 nm
Channel	$\text{In}_{0.65}\text{Ga}_{0.35}\text{As}$		15 nm
Buffer	$\text{In}_{0.52}\text{Al}_{0.48}\text{As}$		500 nm
Substrate	InP		75–100 μm

indium-rich channel material ($\text{In}_{0.65}\text{Ga}_{0.35}\text{As}$) and the underlying InP substrate. In order to efficiently overgrow crystal defects that could potentially have negative impact on carrier mobility and, in extension, the noise properties of the HEMT, the $\text{In}_{0.52}\text{Al}_{0.48}\text{As}$ barrier layer was made relatively thick in comparison with other vertical features.

A δ -profile using silicon (Si) as its donor species provided the electrons required for conduction and was inserted following a 3 nm spacer. The uppermost layer of the epistructure consisted of a highly n -doped cap layer, intended to facilitate the formation of ohmic contacts with the drain and source metals.

STEM micrographs of the region surrounding the gate finger are shown in figure 3.1. The transistor is of a symmetrical design, utilizing a low-impedance T-gate which resides at the bottom of a recess cavity. The cavity was formed through etching of the cap layer. As a direct consequence of the processing conditions – which are further outlined in appendix A – an excess removal of the barrier layer occurred, effectively reducing the thickness of the barrier within the recess. The remaining cap layer formed the drain and source regions of the device which are illustrated along with their corresponding metal contacts in figure 3.1a. In order to reduce the number of defects at the crystal boundary of the recess cavity, the device had been passivated with an approximately 40 nm thick film of silicon nitride (Si_3N_4).

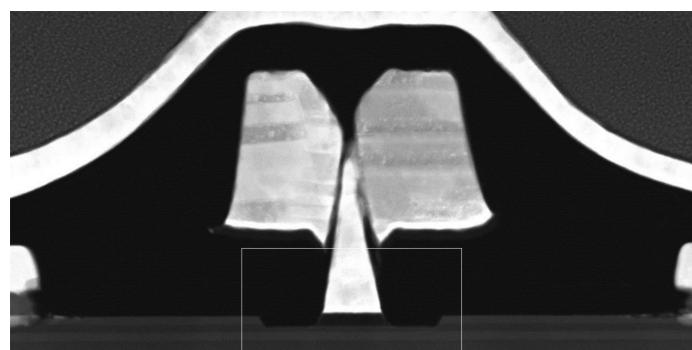
3.2 Numerical device simulations

The modeling of the Chalmers HEMT was carried out using the Sentaurus TCAD software suite, developed and maintained by Synopsys, inc. [7] in Mountain View, California, US.

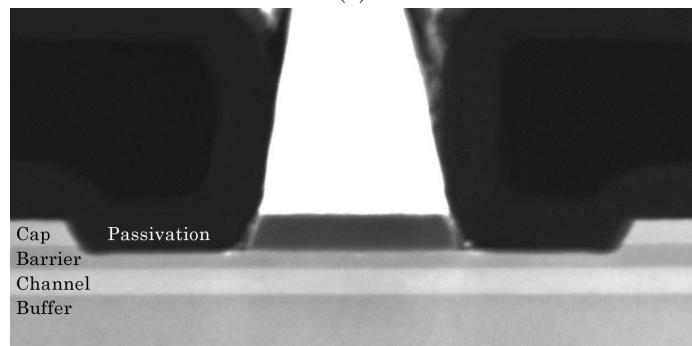
Typically, simulation workflow begins with the definition of the device cross-section in a structure editor, which allows the user to parametrize the device geometry and assign materials and doping profiles. At execution, the program converts the continuous structure into a virtual device whose physical properties are discretized into a nonuniform mesh [7]. Guidelines for the meshing algorithm are set by the user prior to execution, allowing the creation of a denser grid in highly active regions of the device. The guidelines can be utilized to reduce the complexity of the device in low priority regions as a mean to reduce computation times.

The virtual device is then simulated using a special-purpose finite element (FEM) solver, which computes the voltages and terminal currents based on a set of physical device equations. The equations are solved numerically in each grid point (or node as they are conventionally called) and values between nodes are estimated through interpolation. The device equations, which describes the carrier distribution and conduction mechanisms, are normally enabled based on relevance to the behavior of the device under investigation. As the number of equations needed to be solved per iteration greatly reduces the likelihood that the solver will be able to arrive at a solution, the default set of equations is limited and new equations are added incrementally.

The results of a simulation are generally stored for future evaluation and processing. The distinguishing feature and arguably the main benefit of numerical device simulations is the possibility to evaluate the internal physical properties of a hypothetical structure, such as the distribution of carriers within the device



(a)



(b)

Figure 3.1: Scanning transmission electron microscope (STEM) photographies of the region surrounding the gate of the Chalmers HEMT (a) and a magnification around the region at the base of the gate foot (b).

structure under a certain bias. Equivalently, while external input such as bias voltages and ambient temperatures can be applied experimentally in order to investigate the behavior of a device, TCAD environments allows one to estimate how internal properties such as doping levels and geometrical features affect the device behavior.

3.3 Virtual device definition

A virtual device model was constructed based on data available regarding the implementation and fabrication of the reference device. In order to avoid the additional complexity of constructing the multi-gate reference transistors, the numerical model was constructed as a single gate electrode with adjacent source and drain contacts.

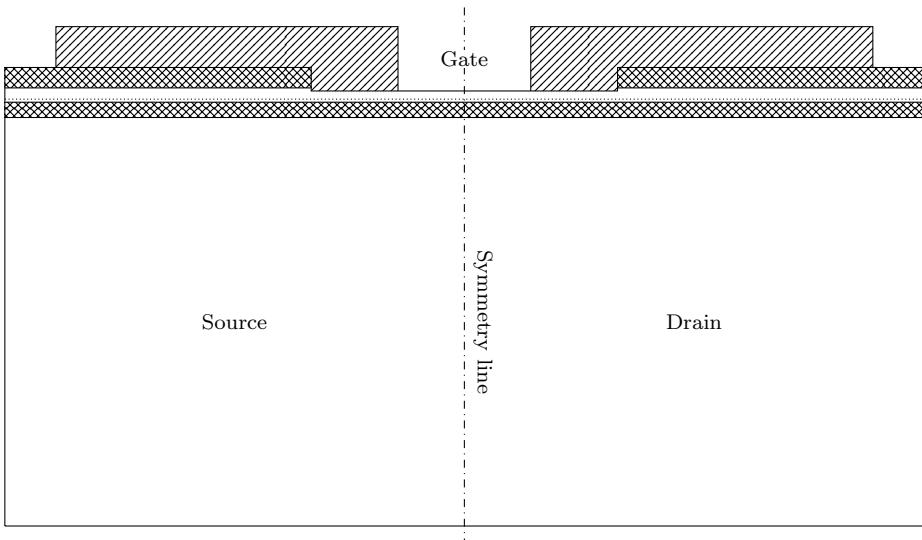
3.3.1 Cross-sectional geometry

The properties of the epitaxial structure, such as layer thicknesses and molar composition of the various alloys, were replicated for each layer in accordance with the data in table 3.1. The InP substrate was omitted entirely from the definition of the virtual device as its presence was found to have negligible impact on the current-voltage characteristics and carrier distributions. Key figures for each layer, such as mole-fraction composition of the various III-V semiconductor compounds were also added to the definition.

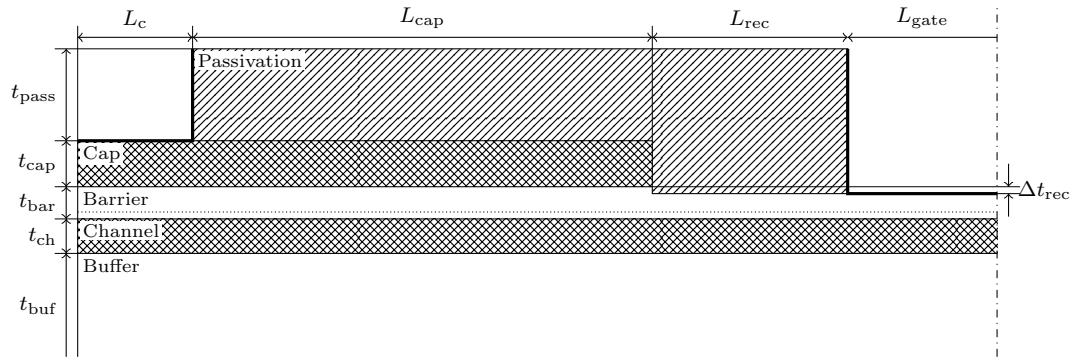
Additional figures related to the dimensions of the recess cavity and the source and drain contacts were obtained from the STEM micrographs in figure 3.1, and were used to replicate the geometry of the two-dimensional cross-section. The distance L_{cap} between the source and drain contact metals and the corresponding edge of the recess cavity were originally measured to be approximately 400 nm. In order to allow the solver to capture and arrive at a solution in these regions, a relatively fine mesh was required which resulted in an unnecessarily high node count and consequently unnecessary high computational workload. As it was noted that the charge distribution and electrostatic potential was rather uniform in these regions, L_{cap} was safely reduced to 250 nm (along with a complementary increase of the contact resistances, which will be discussed below) without impacting noticeably on the simulated current-voltage characteristics. The final continuous geometry is illustrated in figure 3.2, along with the dimensions of its various features which is listed in table 3.2.

3.3.2 Doping profiles

Data regarding the ionization cross-sections and the properties of the donor species Si as implemented in InGaAs was not readily available, and no reliable values were found in literature for the particular alloy composition. As complete ionization of the donor impurities inside the lattice was assumed to occur at the operating temperature (300 K), these properties was considered to have little or no effect. Therefore, the donor species was substituted with arsenic (As) for which data was available [7]. The cap layer were given a uniform doping profile with a constant donor concentration of $5 \times 10^{19} \text{ cm}^{-3}$.



(a)



(b)

Figure 3.2: Cross-sectional schematic of the symmetric virtual device structure (a) and a technical schematic of the region source region (b).

Table 3.2: Geometrical quantities measured from the STEM micrographs in figure 3.1 and the corresponding values of the parameters after the virtual device model was optimized for computational simplicity. Values replaced by dashed lines indicates that the values remained unchanged throughout the optimization.

Symbol	Value (nm)	
	Measured	Optimized
L_c	50	–
L_{cap}	250	400
L_{rec}	85	–
L_{gate}	130	–
t_{pass}	40	–
t_{cap}	20	–
t_{bar}	14	–
t_{ch}	15	–
t_{buf}	400	500

As no dopant profiling data – for instance from secondary ion mass spectroscopy (SIMS) or capacitance-voltage profiling – was available for the δ -doping, it was for sake of simplicity given a completely rectangular profile. The intended surface donor density was $5 \times 10^{19} \text{ cm}^{-2}$. The spatial extent of such doping profiles are generally small and the profile was therefore given a arbitrary thickness of 5 Å. As such, the resulting volume density of donor impurities became 10^{20} cm^{-3} , roughly similar to the donor concentration of the cap layer. As the discretization engine of the software did not allow discontinuous doping profiles to be defined, the profile was implemented as a homojunction between a special-purpose donor layer and the surrounding barrier and spacer layers.

In practical donor injection in III-V compounds, some donors are displaced in the lattice and act as acceptors rather than donors [8]. Therefore, the activation efficiency η_δ of the δ -doping – i.e. the fraction of free electrons provided to the lattice to the the number of donor atoms injected – was therefore specified as a configurable model parameter. The parameter was at the point of definition set to 100 %.

3.3.3 Contact definitions

Due to the high conductivity of the source and drain regions, the source and drain electrodes were defined as ohmic contacts. They were given contact resistances $R_c = 0.030 \Omega\text{mm}$ [4], which had been obtained from previous measurements on the reference device, but the reader should note that these values are particularly low for these types of contacts. In addition, the contact resistances were reduced to $R_c = 0.028 \Omega\text{mm}$ in order to take into account the reduction of L_{cap} .

The gate metal-semiconductor junction was modeled as an ideal Schottky contact with barrier height φ_{Bn} . The barrier height was expected to be affected by more parameters than just the difference in electron affinities between the interfacing materials (platinum (Pt) for the gate metal and InAlAs for the barrier), as defects due to residues of oxide and trapped charges were expected. The defects were in addition likely to make the Schottky contact resistive, but

as the gate currents were small (resulting in a negligible change to the effective gate bias) and no information regarding the magnitude of the series resistance was available, the resistive part was omitted entirely.

3.3.4 Activated device physics equations

Carrier transport was described using the drift-diffusion model outlined in chapter 2. While the quantum-corrected hydrodynamic transport model is generally better at describing the behavior of devices with small active region and high carrier densities, such as the HEMT in this work, complete knowledge of a number of unknown parameters was required in order to utilize the model accurately.

Fermi-Dirac statistics were used to compute the probability of an electron occupying a certain energy state, due to the high carrier densities and discontinuous band structures involved in the HEMTs basic mode of operation. In practice, it means that the computation of γ_n and γ_p in (2.3) and (2.4) had to be computed for each iteration, causing it to contribute notably to the overall computation time.

An inspection of the band structure revealed the unexpected formation of a potential well at the location of the δ -doping (figure 3.3), believed to be caused by BGN. According to former studies on the material properties of InAlAs – or more specifically its constituents InAs [11] and aluminum arsenide (AlAs) [12] – BGN due to the donor levels concerned is in fact noticeable. Yet, as severe band bending due to this type of planar doping has not been documented for similar epi-structures [13] and that it had a detrimental effect on the device behavior, BGN was disabled altogether.

Initially, the electron mobility of InGaAs and InAlAs was computed merely from phonon scattering which required a constant mobility at room temperature to be specified. The electron mobility in the channel material was obtained from Hall measurements on the $2 \times 10 \mu\text{m}$ reference device when the cap layer had been completely etched off [4]. As the majority of current is conducted through the channel layer which in contrast to the barrier layer suffers less from crystal defects and on average enjoys a higher electron mobility, the estimated electron Hall mobility of $10\,000 \text{ cm}^2/\text{Vs}$ was applied directly to InGaAs. Precise knowledge of the electron mobilities in the barrier and the buffer is not critical to device performance as little activity occurs in these regions under standard operating conditions, and a value of $4\,000 \text{ cm}^2/\text{Vs}$ for bulk InAlAs was taken from literature [14].

3.4 Drain current calibration at DC

The current-voltage characteristics measured for the $2 \times 10 \mu\text{m}$ device reference are shown in figure 3.4. The currents are specified per unit length which is common practice in the study of semiconductor components.

The voltage transfer characteristics of FETs on the sub-micrometer scale are affected by an occurrence known as velocity saturation. The drift speed v_d of charge carriers are normally proportional to the electric field they are subject to, but as the field reaches a critical magnitude increased interaction of the carriers with the lattice causes them to lose kinetic energy and their speeds to saturate. As the modeling of velocity saturation introduced additional variables to the

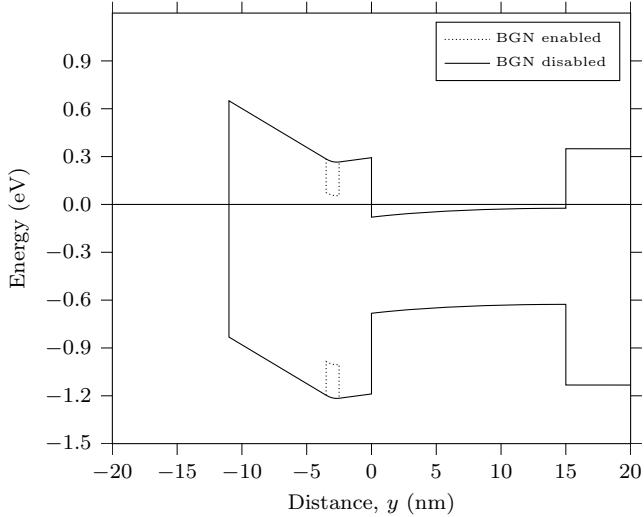


Figure 3.3: Formation of a quantum well within the barrier layer due to band gap narrowing (BGN) caused by the presence of the donor impurities of the δ -doping.

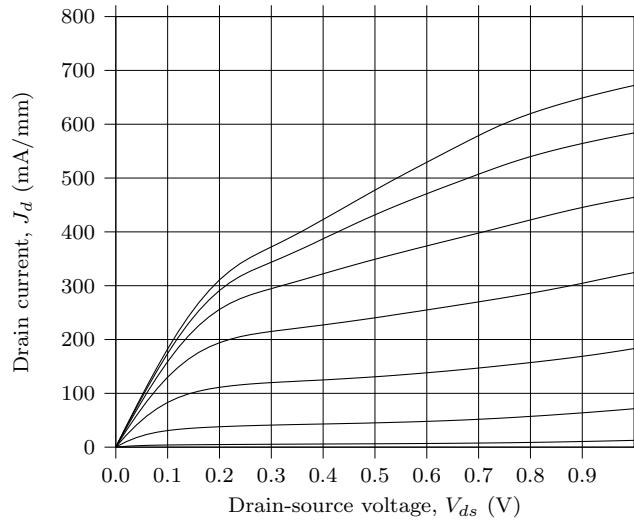
calibration procedure, an initial calibration at small drain voltage biases was carried out in the ohmic region where velocity saturation could safely be omitted.

3.4.1 Drain current at weak lateral electric fields

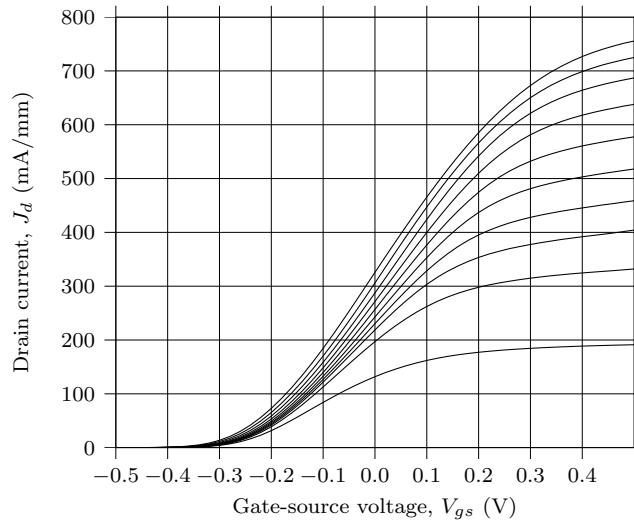
The simulated 2DEG density at zero bias was estimated by integrating the electron density in the channel region along thin vertical segments, and by averaging the contribution from each segment. Using an initial activation efficiency of 100 %, an estimate of 1.35 cm^{-2} was obtained which was in reasonable agreement with the reference value $1.4 \times 10^{12} \text{ cm}^{-2}$ available from previous Hall measurements [4].

Despite the notable agreement at zero bias the device model was in its prevailing state unable to capture the gate bias dependence of n_s . As figure 3.5 indicates, the simulated energy band diagram and particularly the height of the potential barrier are affected by the amount of active donor impurities in the δ -doping. By reducing η_δ from 100 %, the difference between the gate voltage bias at which the channel became pinched off (figure 3.5a and 3.5b) and the bias at which parasitic conduction starts to occur (figure 3.5c and 3.5b) could to some extent be controlled. The modification to the threshold voltage was compensated by altering the Schottky barrier height φ_{Bn} .

By reducing the number of donor atoms that were able to release their valence electron to the lattice, n_s was as a consequence reduced further from its reference value. However, the application of a small drain voltage bias revealed that the magnitudes of the simulated drain currents exceeded those from measurements for all gate biases above pinch-off. As the electron mobility in the channel is largely independent of gate bias and as Hall assessments of charge densities are prone to significant uncertainties [9], it was determined that the prior estimate of n_s was likely too high.



(a)



(b)

Figure 3.4: Current-voltage characteristics of the $2 \times 10 \mu\text{m}$ wide reference device used in the DC characterization of the 130 nm InP HEMT [4]. The gate bias of the output characteristics (a) are stepped from -0.3 V to 0.3 V and the drain bias of the transfer characteristics (b) are stepped from 0.1 V to 1.0 V. Both cases features a step size of 0.1 V.

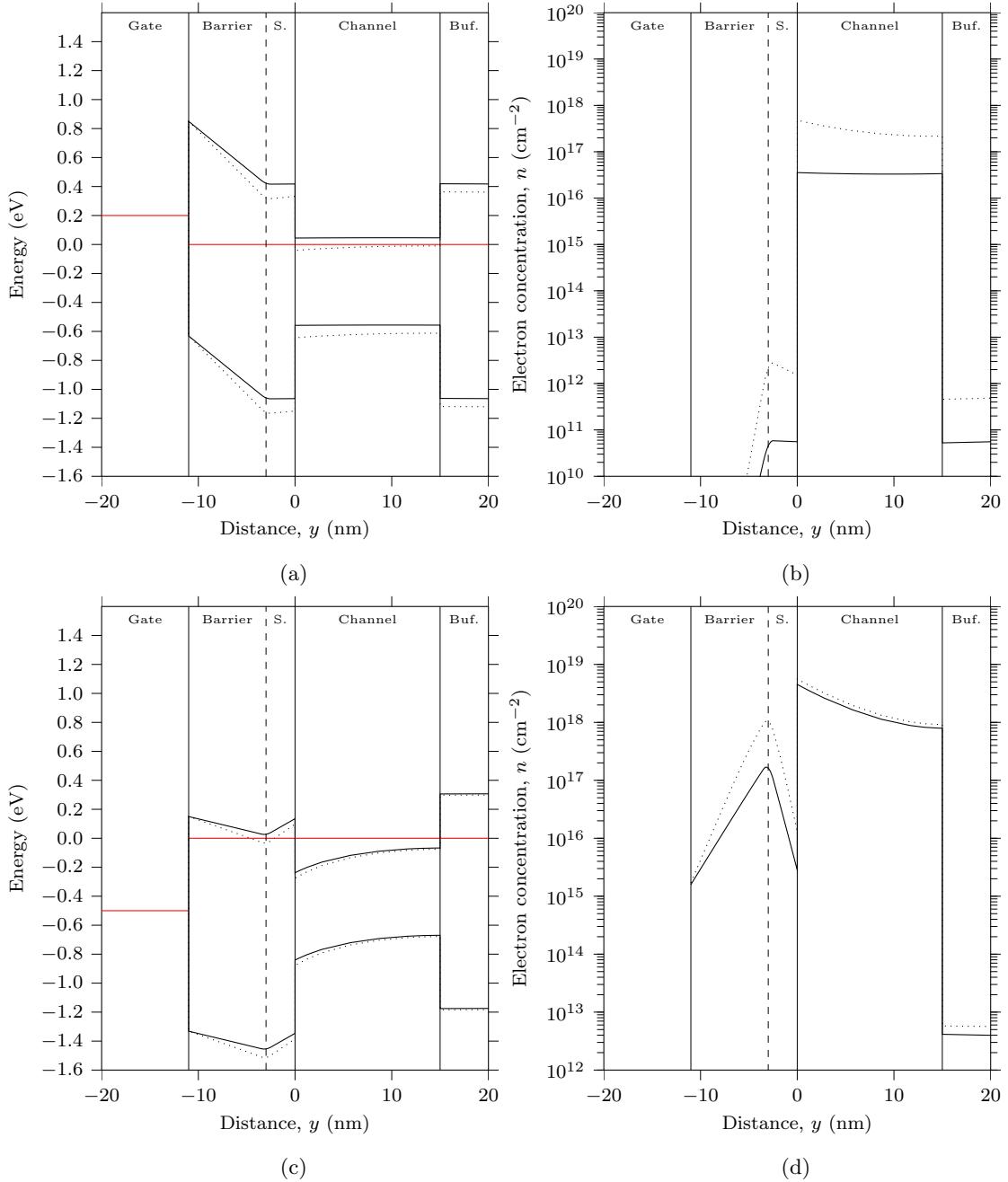


Figure 3.5: Impact on the conduction and valence band profiles underneath the gate due to the level of active donors in the δ -doping. Solid lines indicates an activation efficiency of 75 % and dotted lines an activation efficiency of 100 %.

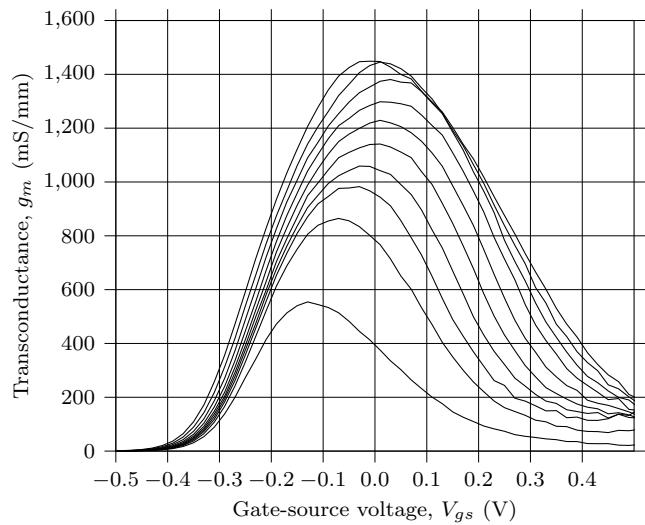


Figure 3.6: Transconductance characteristics of the $2 \times 10 \mu\text{m}$ wide reference device. The drain voltage bias is increased from 0.1 V to 1.0 V in steps of 0.1 V.

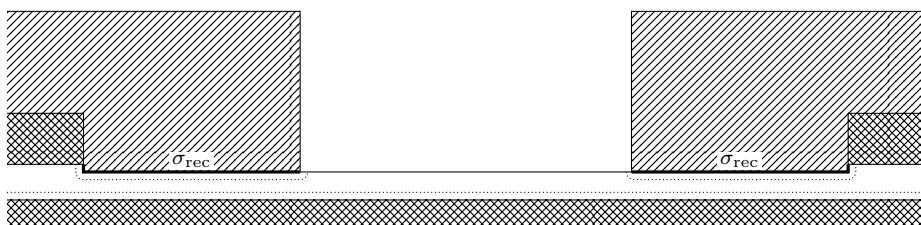


Figure 3.7: Illustration of the placement of surface traps at the $\text{Si}_3\text{N}_4/\text{InAlAs}$ interface.

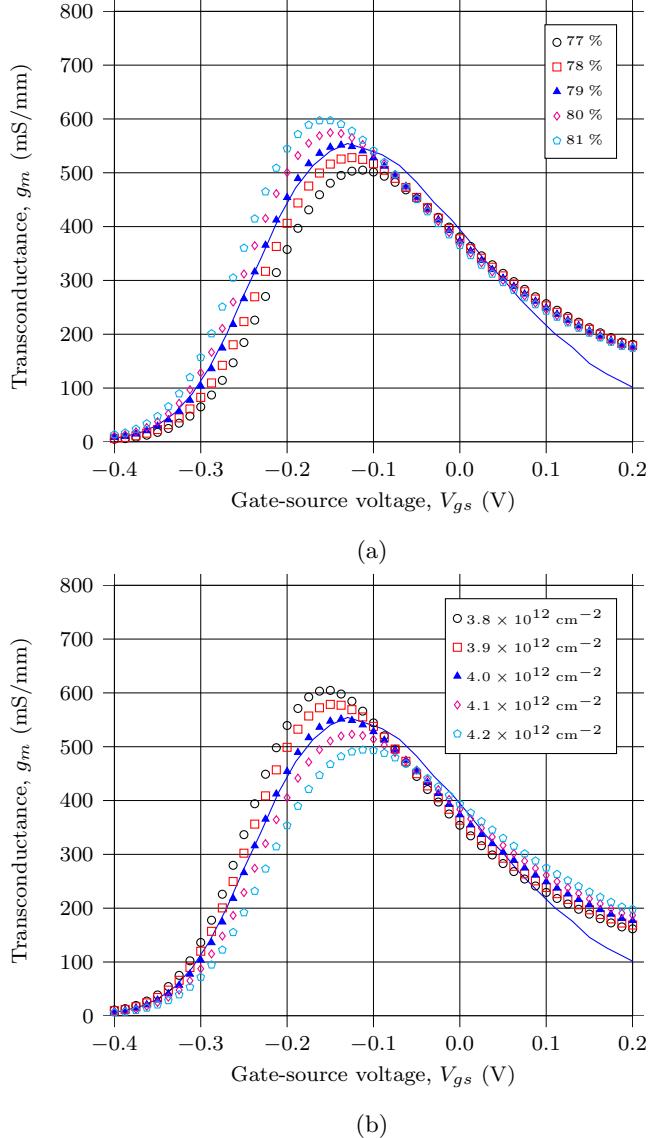


Figure 3.8: Impact on the simulated transconductance characteristics at $V_{ds} = 0.1$ V by the activation efficiency η_δ of the δ -doping (a) and the recess trap density σ_{rec} (b). A solid line indicating the measured transconductance is plotted for comparison.

By reducing η_δ and adjusting φ_{Bn} accordingly, the shape of the transconductance characteristics (figure 3.6) for $V_{ds} = 0.1$ V could be reproduced for gate biases up to 0.1 V, although the magnitude of the simulated transconductance was still greater in magnitude. Therefore, a secondary mechanism limiting the 2DEG density was investigated. From literature [14, 15], it was determined that the presence of acceptor traps at the InAlAs/Si₃N₄ interface (figure 3.7) might cause some electrons to occupy some of these traps and therefore not participate in current conduction. As the mechanism affects only the magnitude of n_s while not noticeably affecting the pinch-off conditions, the surface trap density σ_{rec} could be used to compensate the effects of altering η_δ . The parameters' effect on the simulated transconductance characteristics are illustrated in figure 3.8.

The end result of the initial calibration was the set of model parameters listed in table 3.3. Figure 3.9 features a comparison between the simulated and the measured output characteristics. As velocity saturation was not taken into account, the simulated current does not notably saturate and greatly exceeds the measured drain currents in the saturation region of the HEMT. The range of gate biases at which reasonable agreement could be obtained in the ohmic region was limited to gate voltages up to 0.1 V which is in the order of what has been achieved in previous Monte Carlo (MC) studies of similar devices [15].

3.4.2 Drain current at strong lateral electric fields

As the channel conducts the majority of the current, the calculation of the high-field electron mobility was deliberately limited to the channel layer. As with other III-V compounds, the velocity saturation properties of InGaAs exhibits a region at intermediate field intensities where the transferred-electron effect (figure 3.10) is dominant [16, 17]. However, as the devices considered in this thesis have a gate of 130 nm length (or less, if taking the subsequent down-scaling of the gate length into account), velocity overshoot [9] was considered to cause some electrons to gain velocities exceeding the saturation velocity, which for strained InGaAs has been estimated to 10^7 cm/s [16, 17]. MC simulations on a similar 130 nm InP HEMT predicts charge carrier velocities exceeding 6×10^7 cm/s at $V_{ds} = 0.6$ V [15].

The extended Canali model of high-field mobility modulation [7] was used to describe the electron mobility of InGaAs. The model utilizes an expression to compute the high-field electron mobility μ' based on estimates of the low-field electron mobility, namely

Table 3.3: Collection of model parameters obtained after the calibration procedure outlined in section 3.4.1.

Quantity	Value	Unit
η_δ	79	%
N_δ	2.6×10^{12}	cm ⁻²
σ_{rec}	4.0×10^{12}	cm ⁻²
φ_{Bn}	0.62	eV

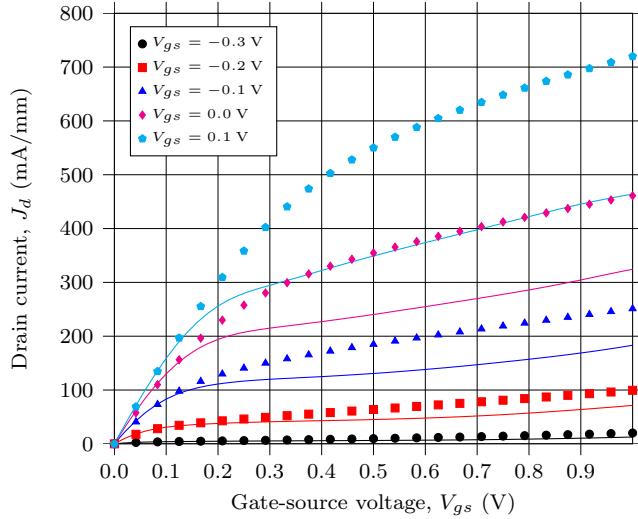


Figure 3.9: Simulated output characteristics with high-field velocity saturation and its impact on electron mobility disabled. The output characteristics of the $2 \times 10 \mu\text{m}$ reference device is plotted as solid lines for reference.

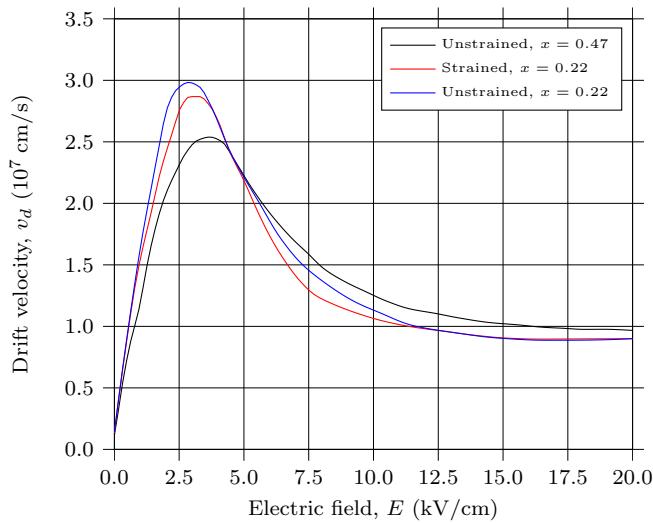


Figure 3.10: Field dependence of the electron drift velocity in different compositions of bulk InGaAs under different degree of lattice strain [16]. The constant x denotes the molar composition ($\text{In}_{1-x}\text{Ga}_x\text{As}$) at 300 K.

$$\mu'(E) = \frac{(\alpha + 1)\mu}{\alpha + \left[1 + \left(\frac{(\alpha+1)\mu E}{v_{\text{sat}}}\right)^{\beta}\right]^{1/\beta}} \quad (3.1)$$

where E is the magnitude of the electric field, v_{sat} is the saturation velocity of the material at 300 K and α and β are fitting parameters. The expression indicates that the electron drift velocity increases non-linearly with the magnitude of the electric field until the carriers eventually reach their saturation velocity, which was defined as a configurable model parameter in the range 1.0×10^7 cm/s to 6.0×10^7 cm/s. The form of the non-linearity was primarily governed by the fitting parameter β . The range of values for β in bulk InGaAs (having a 35 % GaAs content) is 1.2 to 1.5 according to literature [18], but was allowed to be higher as the parameter was supposed to also emulate the effect of velocity overshoot. The additional fitting parameter α is typically used for fine-tuning the behavior in the ohmic region [7] and was set to its default value of zero in order to reduce the model complexity and the number of unknowns.

Using the model parameters obtained in the previous section (table 3.3), as (3.1) was initially activated the simulated drain currents were reduced significantly even at drain bias of 0.1 V. Therefore, an attempt to correct the behavior was made in which N_δ was increased followed by a corresponding compensation of φ_{Bn} to counteract the threshold voltage shift.

The calibration process was reduced into finding a combination of the model parameters (N_δ , σ_{rec} , φ_{Bn} , v_{sat} and β) that caused the virtual device to reproduce the transfer characteristics (figure 3.4b) of the reference device. The number of unknowns made the problem unwieldy, and an iterative optimization procedure was thus implemented using a heuristic optimizer (i.e. based on trial and error) featured in the TCAD software suite. The optimizer operates by designing a number of experiments using different sets of model parameters. The operator then instructs the software to produce a set of scalar responses, such as resulting threshold voltages and peak transconductances, which the optimizer then uses in order to establish a second-order error function. If a local minimum is found, the optimizer reduces the parameter domain and repeats the process until a sufficiently small error is obtained. Since the optimizer is heuristic, the existence of a local minimum does not guarantee that the solution represents a global minimum.

The experiment was designed to generate a set of $J_d(V_{gs})$ curves based on the results of the simulation. The response values returned for each experiment was the threshold voltage V_T and maximum transconductance $g_{m,\text{max}}$ for each curve. As an effort to capture the overall shape of the curves, the residual sum of squares (RSS) between each curve and their reference equivalent was also returned as a response value. The execution of the optimization routine produced the set of parameters listed in table 3.4, which apart from β were all within the acceptable ranges defined earlier in this section. The value of β is considered to be due to a discrepancy between the material properties of bulk InGaAs and InGaAs grown through MBE.

Figure 3.11 illustrates the simulated output characteristics once the calibration had been performed. The discrepancy in the saturation region of the HEMT is considered to be due to a combination of channel length modulation (CLM) and drain-induced barrier lowering (DIBL). CLM denotes the reduction of the

Table 3.4: Collection of model parameter acquired after the heuristic optimization procedure outlined in section 3.4.2.

Quantity	Value	Unit
η_δ	81.5	%
N_δ	4.075×10^{12}	cm^{-2}
σ_{rec}	3.520×10^{12}	cm^{-2}
φ_{Bn}	0.650	eV
β	2.185	
v_{sat}	3.825×10^7	cm/s

effective channel length that occurs as the application of a drain bias causes the 2DEG to extend further towards the source, thus resulting in an increased channel conductance [9]. In addition, as the reduction of the effective channel length also has a negative effect the electrostatic shielding between gate and drain, the drain becomes able to attract electrons to the 2DEG. As such, an applied drain voltage is able to lower the barrier for conduction between source and drain (hence the name drain-induced barrier lowering) or, stated differently, to lower the threshold voltage.

While artificial barrier lowering (i.e. where φ_{Bn} is reduced based on the magnitude of the electric field perpendicular to the gate contact) is a feature of Sentaurus Device, all efforts at implementing a solution was abandoned due to time constraints and limited possibilities in distinguishing the contribution to the 2DEG density from the effects of velocity saturation. Prior studies concerning the conduction mechanisms of the InP HEMT have concluded that at 300 K the increase in output conductance in the saturation region can not be attributed to an increased availability of free electrons due to impact ionization [15, 19].

3.4.3 Gate leakage currents

As excessive gate leakage currents are a limiting factor in the device scaling of short-channel FETs [20, 21], an effort was made at quantitatively describing the gate leakage mechanisms relevant to the InAlAs/InGaAs HEMT structure. Figure 3.12 shows the magnitudes of the gate current measured for the $2 \times 10 \mu\text{m}$ reference at different gate voltage biases. The figure uses a convention in which positive values of the gate current signifies a net flow of current from the gate electrode into the channel. While the hole current has the same direction as the overall current, the reader should be aware that due to convention, a positive electron current denotes a net flow of electrons in the opposite direction.

At reverse gate voltage biases below pinch-off (region 1), a small negative gate leakage current can be observed which is similar in magnitude to the drain leakage current. The current was considered to arise from the drain-gate voltage V_{dg} , which causes the Ti/InAlAs potential barrier to become sufficiently thin in order to allow some electrons to enter the channel from the gate by means of thermionic emission (TE) and tunneling. As the quantities related to TE, such as the energy band structure and the Schottky barrier height was determined beforehand, the simulated tunneling probability was calibrated against its reference by modifying the electron tunneling mass m_{te}^* within the barrier layer.

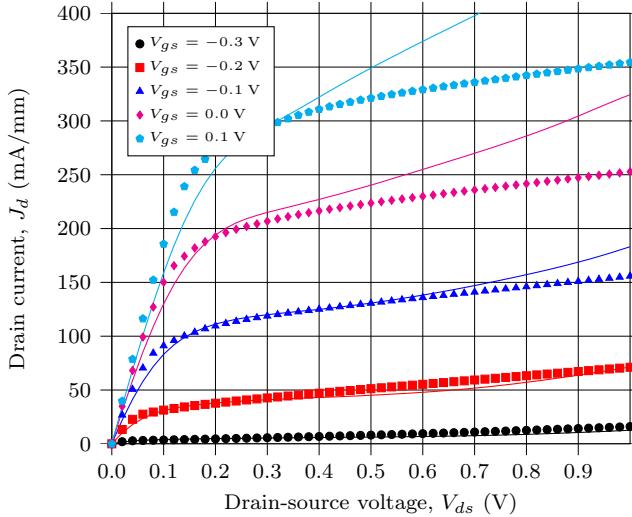


Figure 3.11: Comparison of the output characteristics of the device model as high-field velocity saturation is enabled. The gate voltage bias goes from $V_{gs} = -0.3$ V to $V_{gs} = 0.1$ V in steps of 0.1 V.

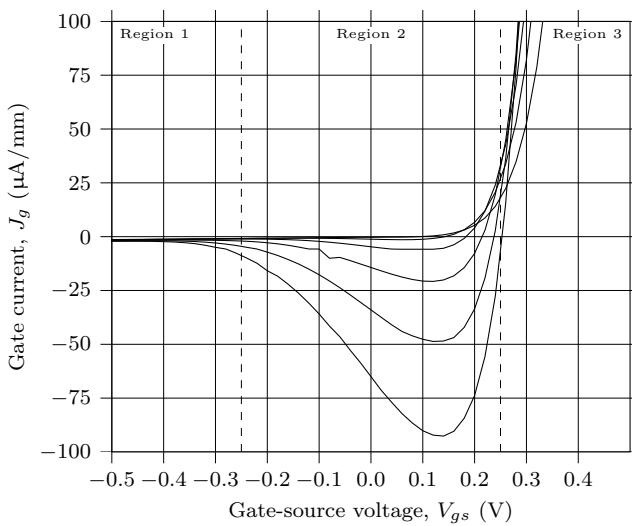


Figure 3.12: Measured gate current characteristics, illustrating the onset of different conduction mechanisms at different gate biases. V_{ds} goes from 0.5 V to 1.0 V in steps of 0.1 V [4].

The bell shaped negative gate current (region 2) occurring for V_{ds} greater than 0.5 V is considered to be due to impact ionization in the depletion region inside the 2DEG. Electrons generated at ionization were expected to drift towards the drain, producing a negligible contribution to the overall drain current. The holes, however, are able to gain enough kinetic energy on their way towards the source as to overcome the potential barrier and be collected by the gate electrode [20, 22].

At large forward biases (region 3) the electron concentration in the channel is increased, resulting in a drastically elevated gate leakage current due to TE and tunneling.

The tunneling of charge carriers was described using a non-local tunneling model [7] for electrons. Since the tunneling current is determined by the conduction and valence band profiles between two points, the model made use of a dense mesh to compute an estimate of the band structure. The tunneling of electrons from the gate electrode into the channel was emulated through the artificial generation of electrons inside the channel based on calculated probability [7].

As the nature of the band structure depends on carrier concentration which needs to be evaluated for each different bias, the method contributed significantly to the computational complexity. The extent of the non-local mesh was limited to an area covering the area surrounding the gate electrode, illustrated more clearly in figure 3.13. Convergence problems in the search for an error minimum were reduced by refining the virtual device mesh in said area.

Figure 3.14 illustrates the match between the measured and simulated gate leakage currents after calibration. Carrier generation due to impact ionization (or avalanche generation) in III-V compounds require the knowledge of a number of physical quantities related to the electric field dependence of the ionization rate. As the rate at which electron-hole pairs are generated is the product of both the ionization rate and the amount of free electrons in the 2DEG, it was not feasible to distinguish each contribution to the hole current from another. In addition, atoms in the lattice are unlikely to be ionized at all under electric fields below a certain magnitude [7, 9] so an additional parameter was needed in order to model the phenomenon. In essence, due to technical difficulties and time constraints the hole current contribution was not modeled and will in the following sections be described qualitatively rather than quantitatively.

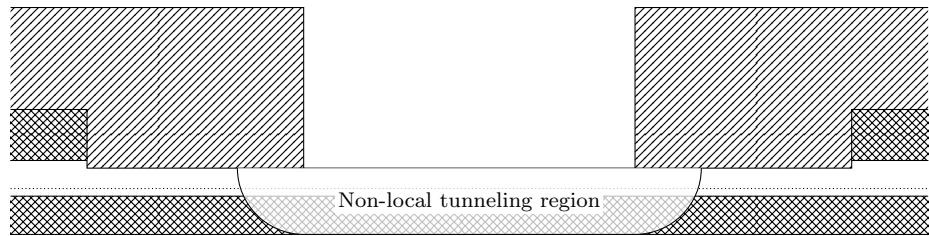


Figure 3.13: Extent of the non-local tunneling region in which the energy band structure is computed and used to compute the probability of tunneling.

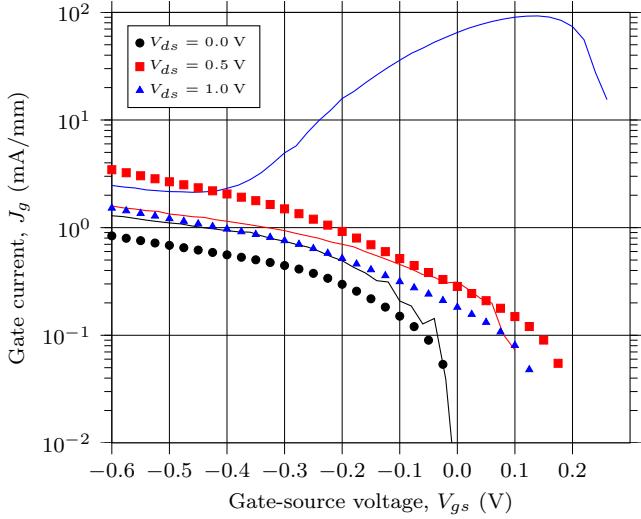


Figure 3.14: Comparison of simulated gate leakage current characteristics and the measured gate current of the $2 \times 10 \mu\text{m}$ reference.

3.5 Microwave characterization

The small-signal performance of the virtual device was compared to the behavior of the $2 \times 100 \mu\text{m}$ HEMT reference. The characterization was done at a low-noise bias point for the reference device at 300 K arrived at by Schleeh [4], namely a V_{gs} of -0.14 V and a V_{ds} of 0.60 V .

An equivalent circuit topology is illustrated in figure 3.15. The network takes into account the parasitics associated with the connector pads necessary for the coplanar microwave characterization of the device, more specifically the capacitances C_{pg} and C_{pd} and the inductances L_g and L_d . The resistances R_d and R_s accounts for the contact resistances of the ohmic contacts, including the bulk resistances leading up to the active region [23]. The gate resistance R_g corresponds to the bulk resistance of the metallic gate. The component values at the given bias point are listed in table 3.5 and were gathered from data regarding the characterization of the HEMT [4]. An unphysical negative source inductance was omitted from the topology as it is believed to have negligible influence on device performance [23].

As the correct magnitude of the input and output impedances of the intrinsic transistor were a prerequisite for characterizing the two-port network, the gate width W_g had to be explicitly specified. The gate resistance was assumed to be approximately uniform along each gate finger, allowing the gate width to be set to $2 \times 100 \mu\text{m} = 200 \mu\text{m}$. The drain and source contact resistances were reduced as to take into account the increased device width, which until that point had defaulted to $1 \mu\text{m}$.

The admittance matrices for the two-port network in figure 3.15 were acquired at a number of logarithmically spaced frequencies between 1 GHz and 1 THz. Figure 3.16 illustrates the transfer characteristics computed from the y -parameters of the simulated device. The figure also features the equivalent quantities for the reference device which were generated from the intrinsic equiv-

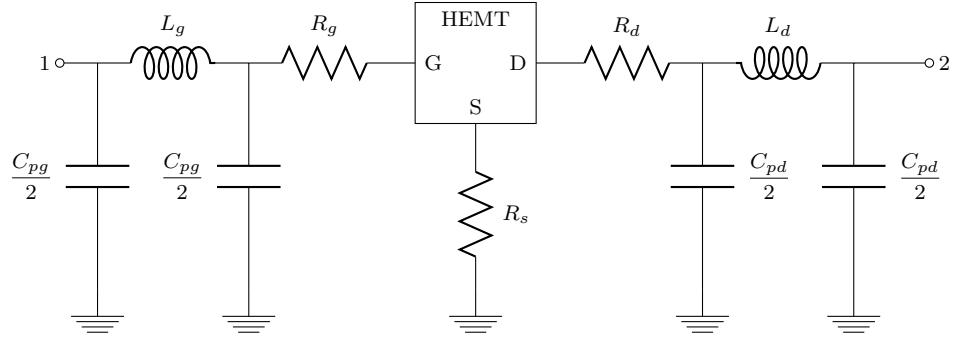


Figure 3.15: Equivalent small-signal circuit topology of the extrinsic parasitic elements associated with the contact pads and bond wires. The contact with the gate, drain and source of the intrinsic components are made at node G, D and S, respectively. The component values are listed in table 3.5.

Table 3.5: Component values extracted for the extrinsic elements of the circuit topology in figure 3.15. The values were obtained at a voltage bias of $V_{gs} = -0.14$ V and $V_{ds} = 0.60$ V [4].

Component	Value	Name
R_g	5.000	Ω
L_g	35.054	pH
C_{pg}	19.300	fF
R_d	1.307	Ω
L_d	36.219	pH
C_{pd}	19.300	fF
R_s	1.155	Ω

alent circuit model [4] of the device based on measurements in the frequency range 2 GHz to 10 GHz.

One of the quantities used to characterize the device model was a figure-of-merit known as the Mason's unilateral gain [24]. The quantity, which is invariant under lossless and reciprocal input and output matching networks, can be described in terms of the y -parameters of the two-port network and is given by

$$U = \frac{|y_{21} - y_{12}|^2}{4(\operatorname{Re}\{y_{11}\}\operatorname{Re}\{y_{22}\} - \operatorname{Re}\{y_{12}\}\operatorname{Re}\{y_{21}\})}. \quad (3.2)$$

Additional figures related to the power gain of the HEMT are the maximum stable gain (MSG) and the maximum available gain (MAG). As their names suggests, they both measure the upper limit of the power gain which can be gained from the transistor while simultaneously suppressing spurious oscillations that might arise due to positive feedback. The MAG is given by

$$\text{MAG} = \frac{|y_{21}|}{|y_{12}|} \left(K - \sqrt{K^2 - 1} \right) \quad (3.3)$$

and indicates the highest amount of power gain that can be achieved while the device is unconditionally unstable. The condition is met only when the Rollet stability factor,

$$K = \frac{2\operatorname{Re}\{y_{11}\}\operatorname{Re}\{y_{22}\} - \operatorname{Re}\{y_{12}y_{21}\}}{|y_{12}y_{21}|} \quad (3.4)$$

is greater or equal to unity. On the other hand, if the device is unconditionally stable,

$$\text{MSG} = \frac{|y_{21}|}{|y_{12}|} \quad (3.5)$$

indicates the maximum amount of gain available from the device.

The frequency at which (3.2) and (3.5) becomes equal to unity was of particular interest as it indicates the upper frequency limit f_{\max} at which the device can experience power gain while maintaining stable oscillations.

The small-signal current gain of the two-port was calculated from

$$|h_{fe}| = \frac{|y_{21}|}{|y_{11}|}. \quad (3.6)$$

The unity current gain cut-off frequency f_T of the transistor – including its parasitic elements – was estimated graphically based on the trend observed for $|h_{fe}|^2$ in figure 3.16. A local peak in the transfer characteristics is evident as the gain approaches unity, thus ruling out the possibility to find the the value by searching directly for the x -intercept. Instead, it was noted that the quantity steadily diminishes by 20 dB per decade frequency and the x -intercept was therefore found through extrapolation. Using the above extraction method, the estimated cut-off frequency of the device model was calculated as 153 GHz as compared to 152 GHz for the reference device, which was considered to be in excellent agreement.

The value of f_{\max} was instead obtained directly from the x -intercept of U . From figure 3.16, an f_{\max} of 196 GHz was obtained for the device model while

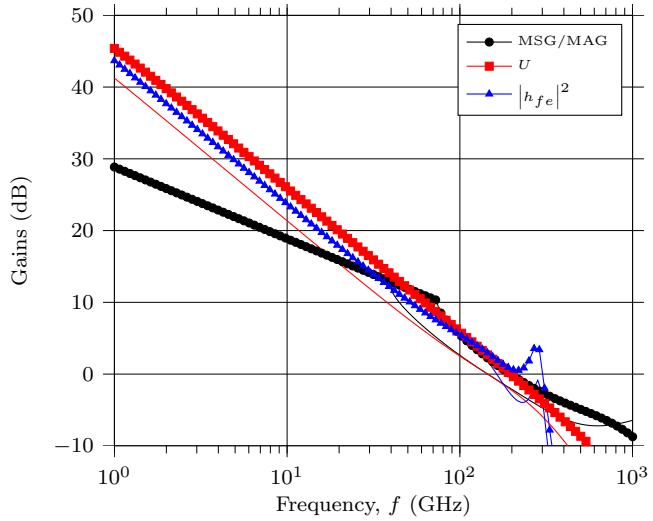


Figure 3.16: Comparison of small-signal transfer characteristics between the device model and a $2 \times 100 \mu\text{m}$ reference.

the corresponding value obtained for the reference HEMT was 145 GHz. The discrepancy between the results can in part be explained by the larger output impedance experienced by the virtual device. As such, the unity power gain of the device model as well as the frequency at which the transistor becomes stable, is drastically increased.

Chapter 4

Device scaling results

This chapter presents the outcome of a subsequent scaling study aimed at establishing the prospects for scaling the Chalmers HEMT towards higher frequency operation. The scaling has primarily been carried out by modifying the values of two model parameters, namely the gate length L_{gate} and the excess depth of the recess cavity Δt_{rec} .

4.1 Scaling of the gate length

Reduced channel lengths due to a shortening of the gate will improve the intrinsic charging times and correspondingly lift the unity gain frequency limits f_T and f_{\max} of the device.

Yet, Yamashita et al. mandates that in order to mitigate the onset of severe SCEs in pseudomorphic HEMTs, the combined thickness of the barrier and the channel layers should under no circumstances exceed the length of the gate [25]. Under this assumption, an effective barrier thickness of 11 nm implies a lower limit of L_{gate} in the vicinity of 26 nm. As no practical issues ruled out further down-sizing of L_{gate} , a series of experiment with values ranging from 130 nm down to 10 nm were constructed.

4.1.1 Current-voltage characteristics

Figure 4.1 shows the simulated output characteristics of the device model for different values of L_{gate} . In accordance with expectations, current saturation starts to degrade as the gate length is reduced, which is further emphasized in figure 4.2. The latter suggests that for L_{gate} less than 70 nm, acute loss of control over the amount of charges in the 2DEG occurs due to increased CLM and DIBL. Figure 4.3 features the energy band diagrams along the channel which were extracted for $L_{\text{gate}} = 130$ nm and $L_{\text{gate}} = 10$ nm, and verifies that lowering of the potential barrier at the source end occurs for both gate lengths, albeit to different extent.

It is essential to reduce the effect of SCEs as much as possible as accurate control over the 2DEG is necessary for successful device operation. Previous studies [5, 21] show that DIBL can be alleviated by a reduction of the distance

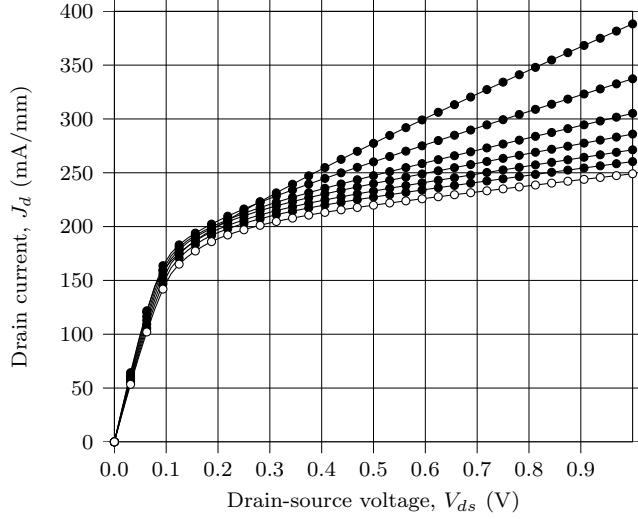


Figure 4.1: Simulated output characteristics at a gate bias of $V_{gs} = 0.0$ V and for gate lengths reduced from 130 nm (denoted by hollow markers) to 10 nm in steps of 10 nm.

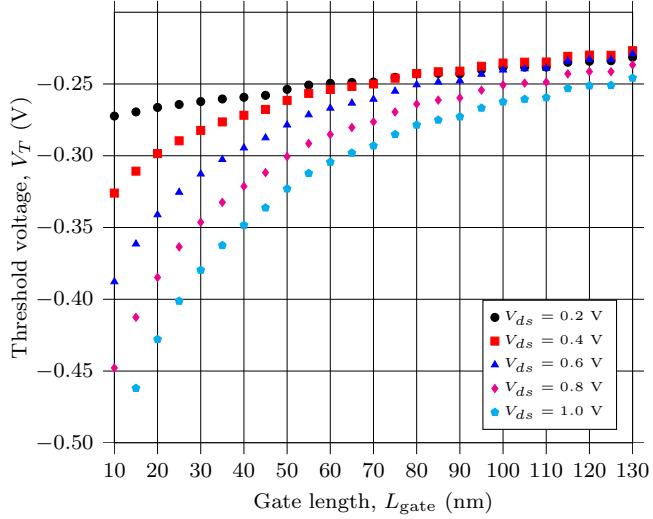


Figure 4.2: Estimated threshold voltage shift at a number of drain biases due to scaling of the gate length.

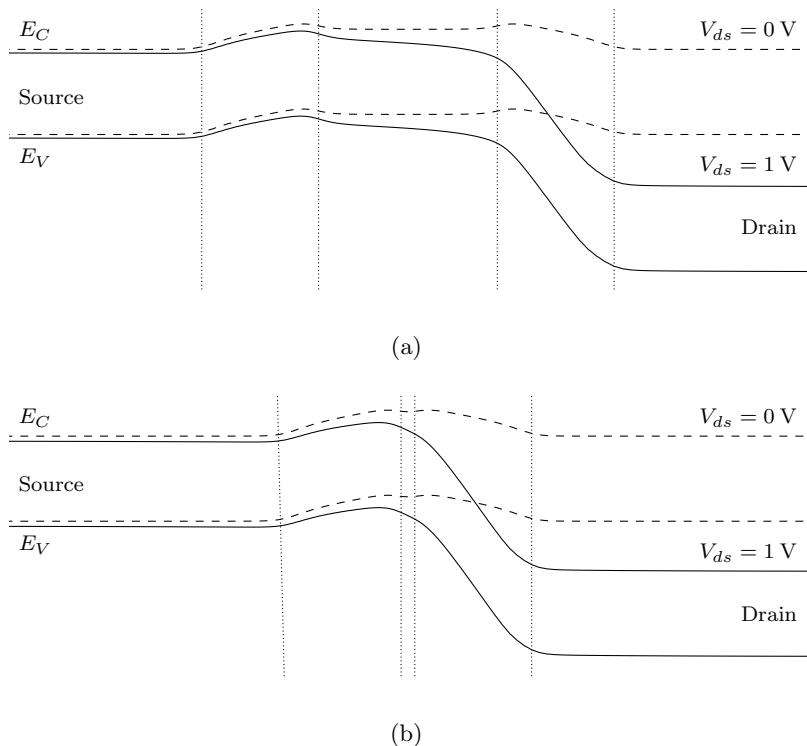


Figure 4.3: Energy band diagram 3 nm inside the channel from the heterointerface for a 130 nm channel (figure a) and a 10 nm channel (figure b) HEMT, showing the different levels of DIBL leading to runaway current at the source. Dashed lines signifies a $V_{ds} = 0.0$ V and solid lines signifies a $V_{ds} = 1.0$ V.

between gate and channel, which will be further investigated in the following section.

In extension, the lack of current saturation also has a negative impact on the output impedance. High output impedance is a desired property of the active components in LNAs due to its positive effect on the available power gain and consequently f_{\max} . However, these benefits are not gained unconditionally as the HEMT needs to be biased at a point at which g_m diminishes as little as possible.

In addition to DIBL, there are a number of causes for the increase in output conductance as the gate length is shortened. Most notably, the magnitude of the lateral electric field in the channel is significantly increased due to a steeper electrostatic potential gradient. This leads to drift velocities in the vicinity of the saturation velocity occurring at lower drain biases. Due to the adverse effect of lowering the potential barrier with low L_{gate} , the 2DEG also experiences a net increase of the sheet charge density, but this effect is negligible in comparison. In addition the shorter physical channel length increases the possibility for velocity overshoot.

Additional information regarding the scaling behavior can be obtained from the transconductance characteristics which are shown figure 4.4. The effects of DIBL are evident here as well, but what can also be seen is that the overall transconductance appears to degrade with reduced L_{gate} , which is expected from InP HEMTs with an 11 nm thick barrier [21]. A more pronounced relationship between the peak transconductance and the size of L_{gate} can be observed in figure 4.5.

4.1.2 Subthreshold leakage characteristics

Due to limitations in the modeling of the gate current, all leakage mechanisms of importance to the InP HEMT can not be reproduced using the proposed model. The fact is most evident in the omission of the gate leakage current occurring at intermediate gate biases.

At relatively large negative gate biases, the magnitude of the drain and gate leakage currents in figure 4.6 increases as L_{gate} is reduced. What can also be observed is that for reduced L_{gate} , an increased discrepancy between the drain and gate current occurs, indicating that a larger fraction of the drain leakage current consists of electrons leaking from the source.

At intermediate gate biases, the gate current of a physical HEMT is expected to experience an increase in magnitude due to the presence of a larger hole current (as is the case in figure 3.12 on page 25). Instead, figure 4.6 indicates a net reduction of the gate leakage at gate voltages above -0.6 V which is due to the reduction of the gate contact area ($W_{\text{gate}}L_{\text{gate}}$) through which a current can flow. Similar studies on gate scaling have shown that increased drift velocities together with lowered conduction thresholds causes an overall increase of the hole current as the gate length is reduced [22]. As barrier lowereing becomes more severe with shorter gates and increased CLM, the mechanism will dominate the gate leakage current at even lower gate biases.

As impact ionization of holes inadvertently results in an equal number of free electrons being generated, it should be noted that practical device might exhibit a minor kink in their output characteristics [13]. However, prior Monte Carlo

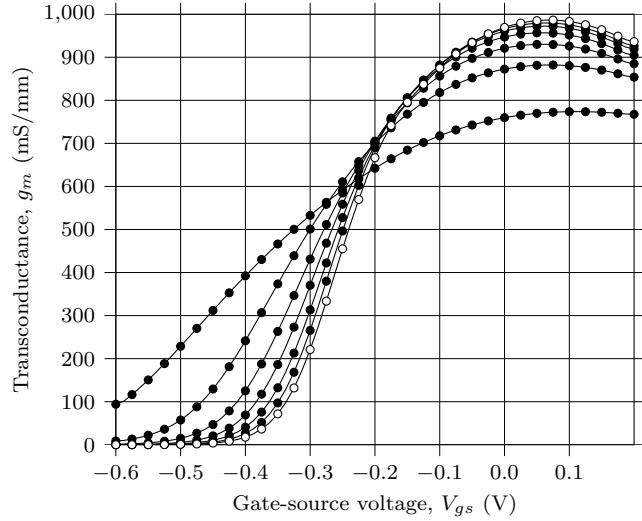


Figure 4.4: Simulated transconductance characteristics for different values of L_{gate} at $V_{ds} = 0.6$ V. The gate length L_{gate} is reduced from 130 nm (denoted by hollow markers) to 10 nm in steps of 10 nm.

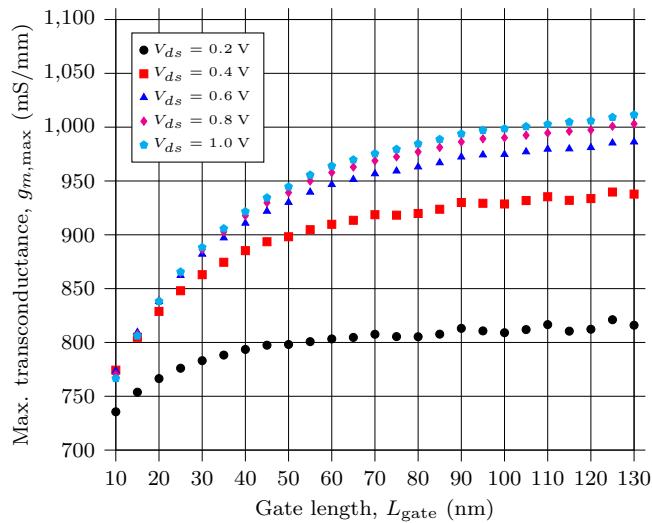


Figure 4.5: Estimated peak transconductance for different gate lengths, obtained from the simulated transfer characteristics (fig. 4.4) at a gate voltage bias of 0.0 V.

studies have indicated that it is hardly visible at the operating temperature concerned (300 K) [15].

4.1.3 Small-signal characteristics

The HEMT was characterized for different values of L_{gate} in a similar fashion as was done for the 130 nm device model in the previous chapter. As the unit gain frequencies f_T and f_{\max} of the *intrinsic* transistor are of particular interest an equivalent circuit topology was created consisting only of the virtual transistor model in series with a gate resistor. The value R_g of the resistor was scaled along with L_{gate} in order to reproduce the impact of a reduced cross-section on its ability to conduct current. While information regarding the cross-sectional area was available from the STEM images of the 130 nm on page 11 and the scanning electron microscope (SEM) images on page 57, respectively, little information was known regarding the effect of L_{gate} on the overall cross-section. A simple scaling rule was implemented where the cross-section of the gate hat was held constant while the cross-section of the gate foot was scaled proportionally with L_{gate} . The scaling of L_{gate} from 130 nm to 10 nm therefore resulted in corresponding scaling of R_g from 5.00 Ω to 6.56 Ω , although the difference was found to have negligible impact on the small-signal characteristics.

The admittance matrices were obtained through small-signal simulations at the previously used low-noise bias point, namely $V_{gs} = -0.14$ V and $V_{ds} = 0.60$ V. The gate bias should ideally be scaled to accommodate the reduction of V_T but as the optimal biasing conditions in terms on noise rejection was not known at the time, the bias was kept constant.

The main improvement on the small-signal properties expected from a down-scaling of L_{gate} was improved responsiveness or, stated differently, shorter transit times τ . These benefits are expected to arise mainly because of a reduced capacitance C_{gs} between gate and channel, the latter which is closely coupled to the source. The expectation is analogous to the behavior expected from a parallel plate capacitor, which experiences a reduction in its ability to contain charges as the area of each surface charge is reduced. The combined charging time of the gate is given by C_{gs}/g_m , and relates to the unity gain cut-off frequency by

$$f_T = \frac{1}{2\pi \sum \tau} = \frac{g_m}{2\pi C_{gs}} \quad (4.1)$$

where g_m denotes the small-signal transconductance at the given bias point which is equal to the real part of y_{21} at DC ($f = 0$ Hz).

The gate-channel capacitance was computed from the y -parameters of the equivalent two-port for different values of L_{gate} and is presented in figure 4.7. Just as for parallel plate capacitors, C_{gs} increases linearly with L_{gate} . For sake of comparison, the diagram also features the corresponding values of the equivalent circuit on page 28 (figure 3.15).

Figure 4.8 shows the f_T and f_{\max} estimated for the transistor (both including and excluding the parasitic components in figure 3.15 on page 28) using the previously described graphical method for different values of L_{gate} . The cut-off frequency of the intrinsic device model experiences a dramatic increase from 202 GHz to 388 GHz as L_{gate} is scaled from 130 nm down to 10 nm.

Figure 4.8 also contains the estimated values of f_{\max} , which in the case of the intrinsic HEMT is shown to be less than f_T for all channel lengths below

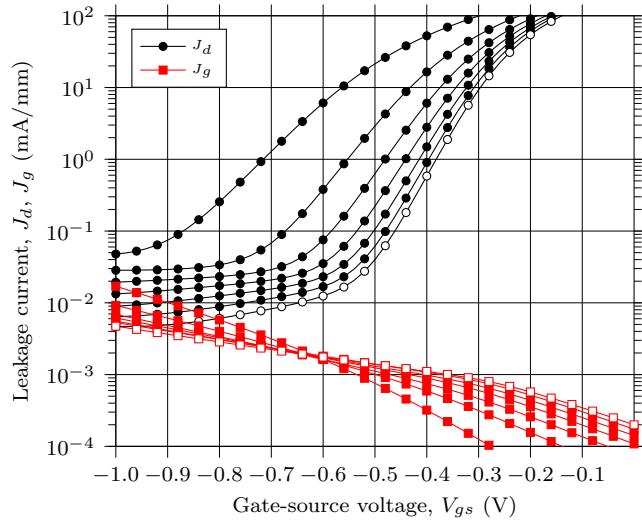


Figure 4.6: Simulated subthreshold characteristics of the drain and gate currents for different values of L_{gate} . The drain voltage bias of 0.6 V, and L_{gate} is reduced from 130 nm (denoted by hollow markers) to 10 nm in steps of 20 nm.

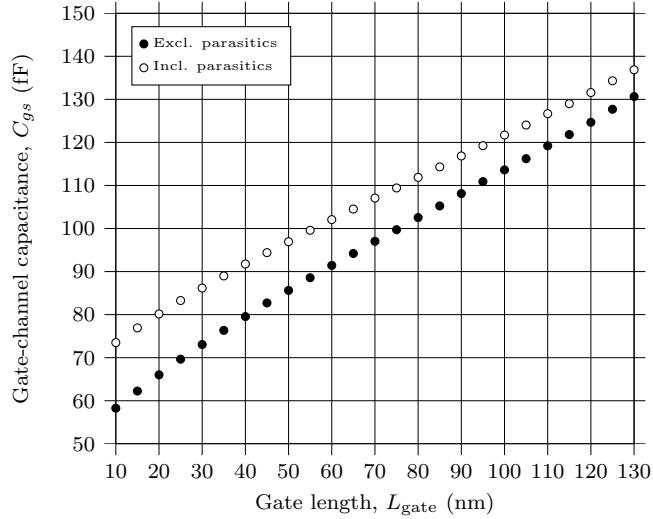


Figure 4.7: Gate-channel capacitance computed for devices with different L_{gate} . Solid markers denotes the values obtained from the y -parameters of the intrinsic circuit topology, consisting only of the virtual device model in series with a gate resistor. Hollow markers indicates the corresponding values as the parasitic elements in figure 3.15 were included. The bias point is $V_{gs} = -0.14$ V and $V_{ds} = 0.60$ V.

100 nm. While both the modeling of the power gain and the extraction procedure are prone to inaccuracies, there is a possibility that f_{\max} rather than f_T will dominate the device behavior for these gate lengths [26]. For the intrinsic device model, the simulated f_{\max} experiences a peak value around 20 nm.

4.2 Scaling of barrier thickness

In order to reduce the predominance of SCEs at gate lengths less than 70 nm, and to preserve electrostatic integrity within the channel, the separation between gate and channel was scaled.

A new quantity t_{ins} was introduced which denotes the insulator thickness or the distance between the gate electrode and the channel. The distance was controlled by altering the excess recess depth Δt_{rec} (figure 4.9) which in practical applications is controlled through the etching times of the gate recess cavity. The simulations described below were then carried out on devices with Δt_{rec} increasing from 3 nm to 6 nm, which in turn meant reducing the insulator thickness from its original 11 nm down to 8 nm.

The practical formation of the gate recess which is outlined in appendix A implies that Δt_{rec} cannot be increased without also increasing the overall length of the recess L_{rec} . However, as there exists other process methods (including, but not limited to heating) to prevent this effect, L_{rec} will be kept constant in order to also keep the drain and source access resistance constant. The reader should further note that the practical reduction of the gate-channel separation can take other forms and have ambiguous effects on device performance. A reduced gate-channel separation might as well be attained through the growth of a thinner barrier layer during epitaxy, but might simultaneously have detrimental effects on the number of bulk traps in the barrier. It has been suggested that reduced gate-channel separation might be achieved through the application of buried gates based on Pt [27].

4.2.1 Current-voltage characteristics

Figure 4.10 illustrates the simulated transconductance characteristics and the impact that a reduction of t_{ins} from 11 nm to 8 nm has on the pinch-off conditions of the virtual device. Due to a lowering of the potential barrier that is close to proportional to the reduction of t_{ins} , the threshold voltage experiences a nearly linear increase which is further illustrated in figure 4.11.

In contrast to studies on the device scaling of similar HEMT structures [5, 21, 28], the scalability of the transconductance was not noticeably improved which is further emphasized in figure 4.12. This might be an indication that further thinning of the insulator is required in order to mitigate the SCEs or that such mitigation can not be achieved through barrier scaling. On the other hand, also adjacent to the results of these studies is that the benefits in terms of control over the 2DEG does not appear to come at the cost of an overall reduced transconductance, which could be explained by a radically lower source access resistance than in the other studies [13].

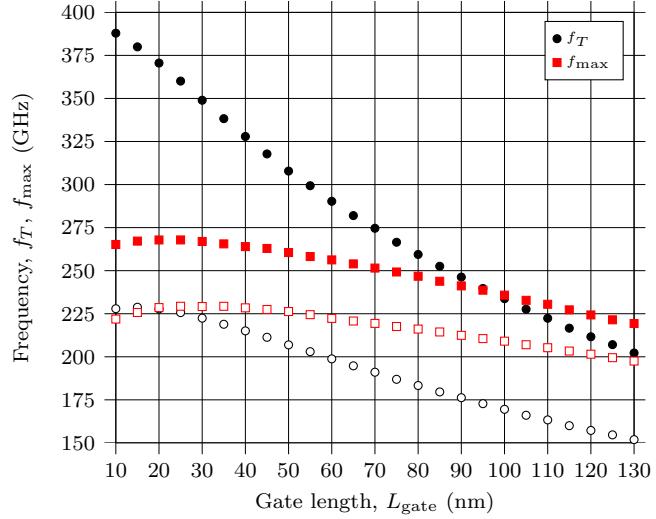


Figure 4.8: Unity gain frequencies f_T and f_{max} as a function of gate length. Filled markers indicate the quantities as extracted from the intrinsic device model and hollow markers those extracted when external parasitics were included.

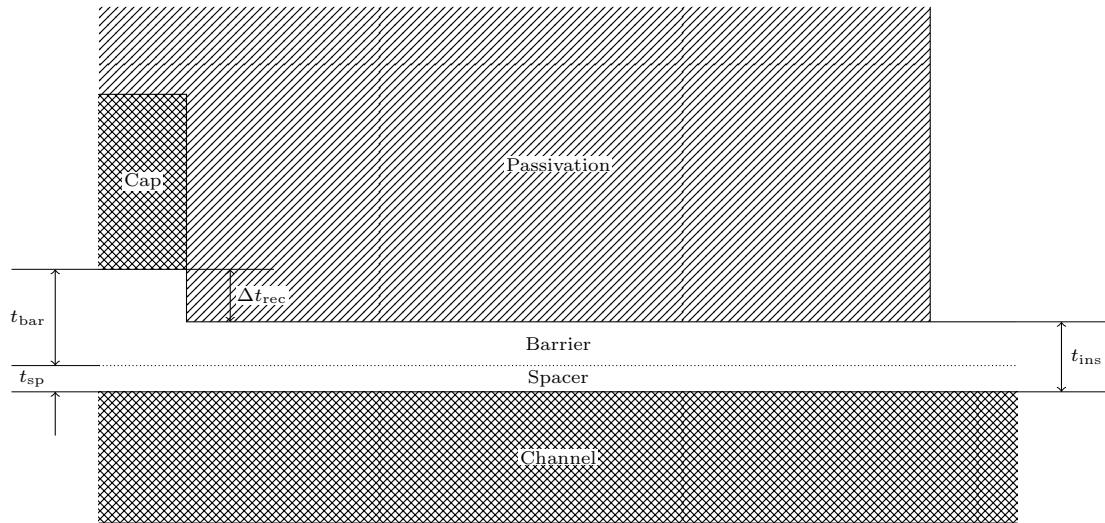


Figure 4.9: Schematic of the source recess region as the excess recess depth Δt_{rec} is 6 nm, producing a gate-channel separation of 8 nm.

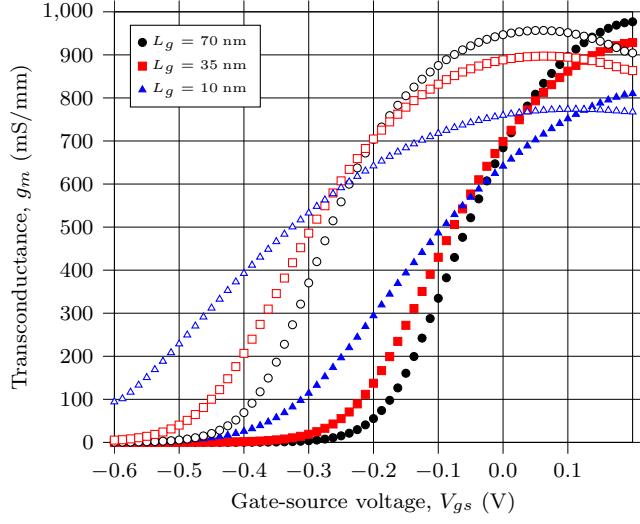


Figure 4.10: Transconductance characteristics for barrier thicknesses 11 nm (hollow marks, \circ) and 8 nm (solid marks, \bullet). The drain voltage bias is $V_{ds} = 0.6$ V.

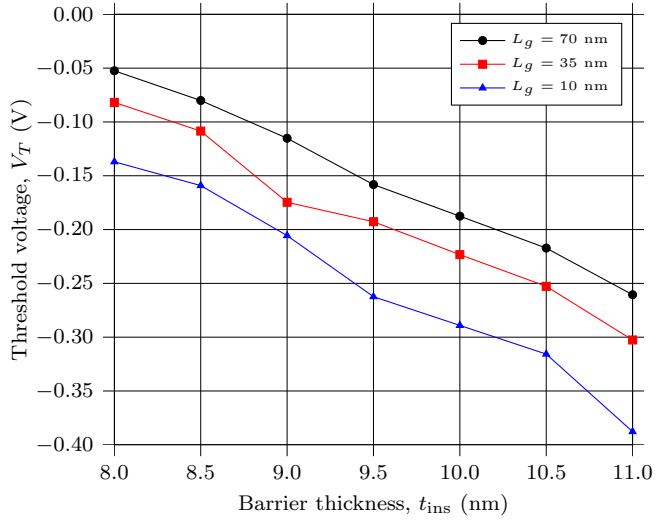


Figure 4.11: Threshold voltage V_T extracted for devices with different L_{gate} as a function of t_{ins} . The drain voltage bias is $V_{ds} = 0.6$ V.

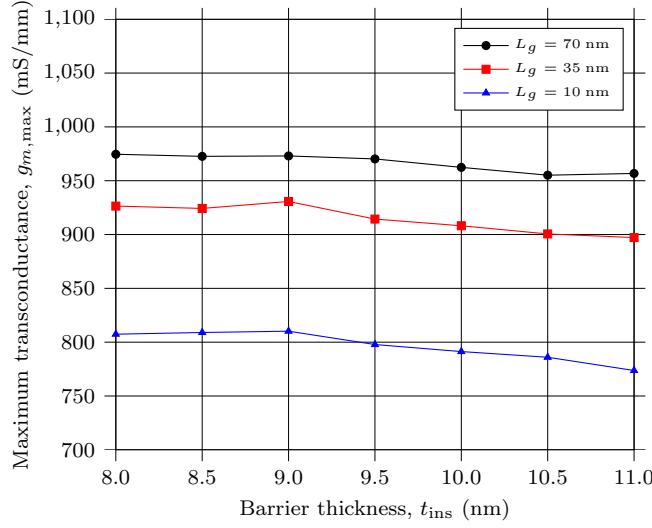


Figure 4.12: Estimated peak transconductance $g_{m,\text{max}}$ for devices with different t_{ins} at a drain voltage bias is $V_{ds} = 0.6 \text{ V}$.

4.2.2 Subthreshold leakage

Figure 4.13 illustrates the simulated leakage currents for virtual device models using different insulator thicknesses and gate lengths. In accordance with expectations, the drain leakage currents experiences a reduction due to improved pinch-off which can be observed as t_{ins} is reduced from 11 nm to 8 nm. As before, the gate current is proportionally reduced along with the gate length as the gate contact area becomes smaller.

However, contra what one might intuitively expect from a reduction of the insulator thickness, the gate leakage currents at large negative gate biases are also diminished. This can be explained by a reduced potential difference between the gate electrode and the channel that occurs as the 2DEG is not fully depleted of electrons. As a less negative gate bias is applied, the potential difference and the number of electrons which are able to overcome the barrier is increased, leading to higher gate leakage currents.

4.2.3 Small-signal characteristics

Using the analogy regarding planar capacitors, C_{gs} was expected to increase as the distance between the 2DEG and the gate contact was reduced. According to (4.1), an increase of C_{gs} would have a detrimental effect on f_T . Generally, proper gate voltage bias can be used to alleviate this effect by increasing the small-signal transconductance, allowing for the mitigation of SCEs without noticeably affecting the power gain properties of the HEMT. For a gate-channel separation of less than approximately 6 nm, the virtual device resembles an enhancement mode device and a positive V_{gs} is required in order to cause conduction in the channel.

The gate-channel capacitance (figure 4.14) was as before calculated from estimates of f_T and g_m . In order to rule out the possibility of large errors in

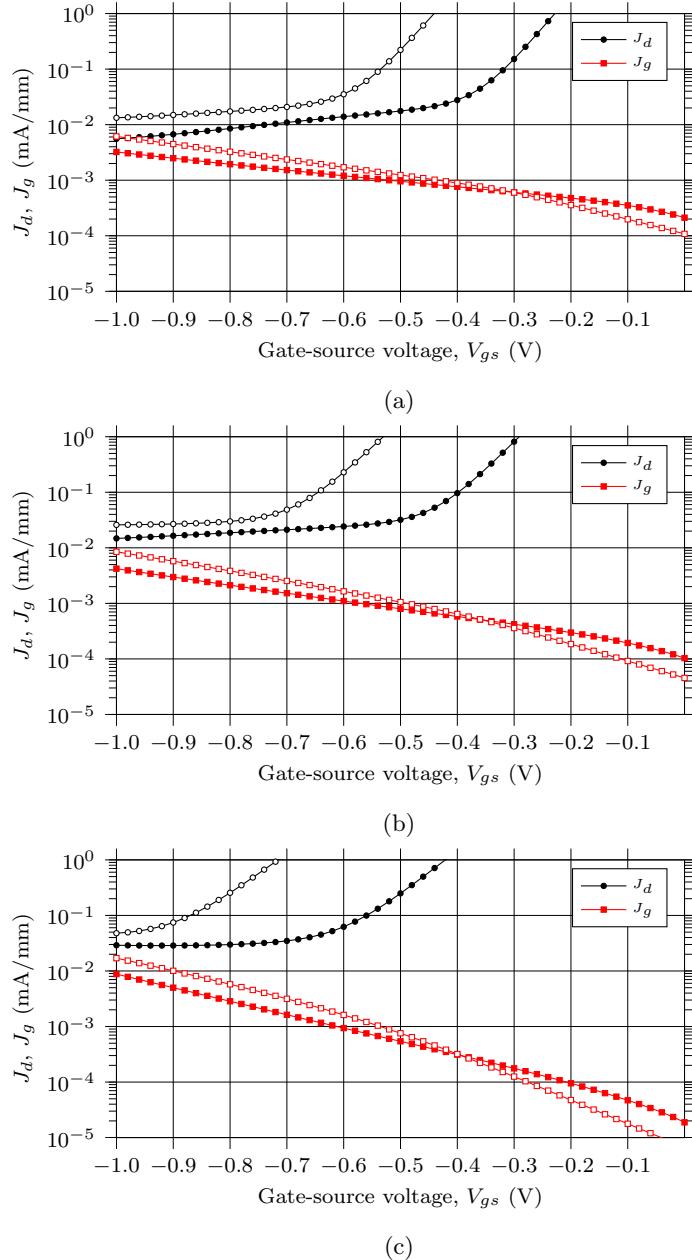


Figure 4.13: Simulated subthreshold characteristics for gate lengths 70 nm (a), 35 nm (b) and 10 nm (c) at a drain voltage bias of 0.6 V using different values of t_{ins} . The insulator thickness is reduced from 11 nm (denoted by hollow markers) to 8 nm in steps of 1 nm.

the graphical extraction procedure used prior, C_{gs} and g_m was also computed in parallel using an lumped element extraction procedure [29]. Contra what one might expect, the trend in figure 4.14 indicates that the modulation of the channel current remains either unaffected or degrades with reduced gate-channel separation.

However, as the current driving capabilities of the virtual device was not noticeably improved (figure 4.12) by barrier scaling, the current and power gain of the device deteriorate rather than improve. As such, a reduction of the gate-channel separation from 11 nm to 8 nm is not able to fully mitigate the SCEs that were observed for gate lengths below 70 nm. As the effect is also not noticeably alleviated by changes to the biasing conditions (4.15), the behavior can potentially be explained by increased access resistances at drain and source [13, 21]. As it stands, figure 4.15b indicates that improvements in terms of f_{\max} are gained through proper scaling of the biasing conditions rather than by barrier scaling.

As there is a discrepancy between the measured and simulated output characteristics at the bias points under investigation, the reader should note that the optimal f_{\max} of the physical device is likely less than what can be deduced from figure 4.15b. As such, the peak value in the vicinity of 255 GHz do not serve as an indicator to the upper frequency limit of the device.

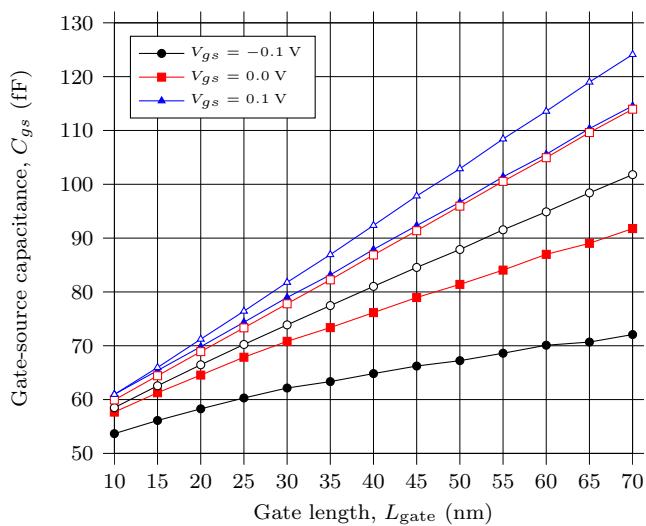
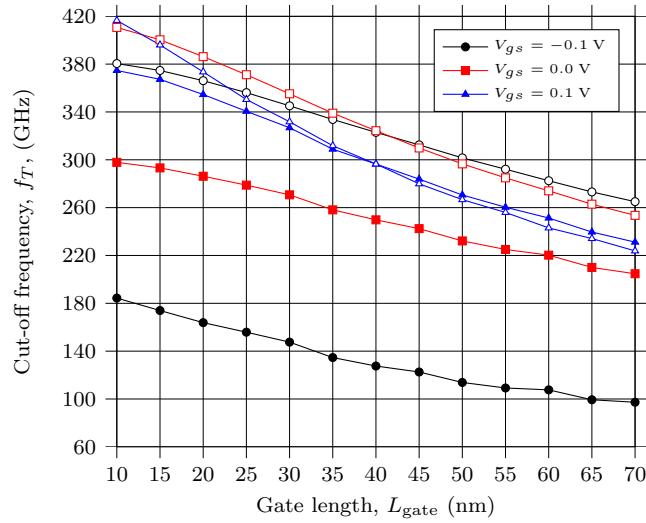
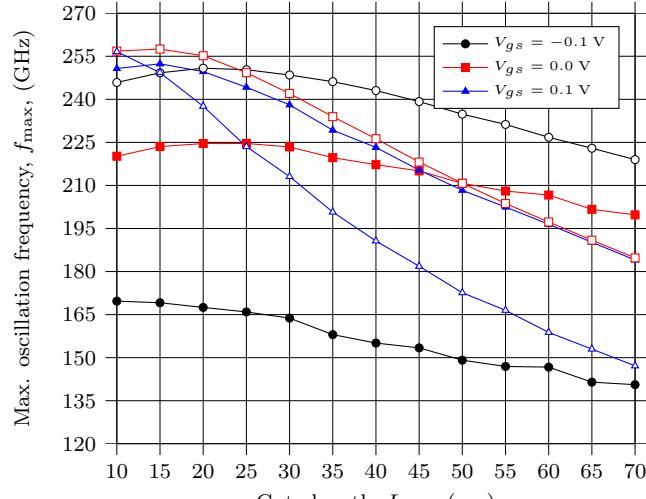


Figure 4.14: Gate-source capacitance as a function of gate length for insulator thicknesses 11 nm (hollow markers) and 8 nm (solid markers). The drain voltage bias is 0.5 V.



(a)



(b)

Figure 4.15: Unity gain frequencies f_T (a) and f_{\max} (a) as a function of gate length for insulator thicknesses 11 nm (hollow markers) and 8 nm (solid markers). The drain voltage bias is 0.5 V.

Chapter 5

Conclusion and future work

A virtual device model of a 130 nm pseudomorphic InP HEMT has been modeled for the purpose of being used in a device scaling study. The physical description is based on data acquired through DC and RF measurements performed on a $2 \times 10 \mu\text{m}$ physical implementation of the HEMT at 300 K using a cryogenic probe station. The virtual model has been implemented and tested with the aid of the Sentaurus brand of TCAD software.

As accurate description of the electrical characteristics requires knowledge of the device physics, as well as awareness of inconsistencies with theory arising from variability in device fabrication and measurements uncertainties, a qualitative description was aimed at. A macroscopic description was therefore implemented which omitted a number of minor contributions expected to arise from the generation/recombination of electron-hole pairs and various scattering mechanisms.

The current-voltage characteristics of the model which was derived yields acceptable agreement in the ohmic operating region of the reference device. The model was found to come short in faithfully describing the effect of certain SCEs, causing the model to exhibit far higher output impedance than what can be expected from a physical device. An adverse effect of the model's inability to reproduce the output impedance is an over estimation of the intrinsic small-signal power gain in the saturation region of the device.

In view of time constraints, the modeling of a gate leakage mechanism which concerns the real-space transfer of free holes between the active region of the 2DEG and the gate electrode has been omitted. Instead, a qualitative argument regarding its impact on the scaling limits is put forward.

The device scaling study was limited to investigating the impact of scaling the gate length as well as reducing the separation between gate and channel through additional etching of the recess cavity.

Nearly all aspects of the Chalmers HEMT, from the design of the epitaxial structure, the device fabrication process, and the choice of contacts, have been carefully selected with cryogenic (4 K) low-noise operation in mind. In order to establish the biasing conditions required for low-noise operation and, in extension, the limitations in regard to device scaling, the device model will need to be adapted for this condition. As the quantized nature of the charge carriers have a predominant effect on the device characteristics and cannot be neglected at 4 K, a future model will likely need to employ a quantum-corrected approach

using hydrodynamic charge transport. If done successfully, the model will benefit from more precise knowledge of carrier and temperature distributions within the device structure, which would for instance allow for the modeling of carrier temperatures and the transferred electron effect [7].

Additional work might involve extending the scaling study to include modifications to the heteroepitaxial structure. It is reasonable to believe that the gate-source voltage required for conduction can be increased by scaling the sheet donor density of the delta-doping, which would allow the application of thicker insulator less prone to leakage [13]. The confinement of the electron current to the channel quantum well can potentially be improved by increasing the In content of the channel. Further inquiry into what effect this scaling will have on the level of crystal defects would then be needed in order to estimate the scaling of other features, such as the barrier thickness. Finally, several authors have reported that the current-driving capabilities can be improved by scaling of the channel thickness [5], as well as through engineering of the spacer layer thickness and material composition [30].

Acknowledgements

I, the author of this work, would like to extend my gratitude towards my examiner, professor and deputy head of laboratory Jan Grahn, for his guidance and support in making this project a reality. In addition, I would also like to acknowledge the support of my supervisors. Associate professor Hans Hjelmgren, for providing me with invaluable advice and direction in the TCAD environment and for aiding me in my understanding of the device and the underlying device physics. Researcher Per-Åke Nilsson, for introducing me to the exciting subject of microfabrication and for allowing me to witness the process first-hand in the MC2 clean room.

Moreover, I would also like to thank the HEMT research group at the microwave electronics laboratory for their input and suggestion, and for sharing their reflection on the topics presented in this thesis. A special thanks goes out to Joel Schleeh for his characterization of the cryogenic HEMT and for generously providing me with the information upon which the conceptual device model is based. Likewise, to Helena Rodilla for taking her time to introduce me to her research on the topic of MC simulations on InP HEMTs and, apart from giving me hands-on access to data, providing me with better understanding of the practical manifestation behind the operation of the device.

Bibliography

- [1] Marian W. Pospieszalski. “Extremely low-noise amplification with cryogenic FETs and HFETs: 1970–2004”. In: *Microwave magazine* 6.3 (2005), pp. 62–75.
- [2] C. K. Peng et al. “Microwave performance of InAlAs/InGaAs/InP MOD-FETs”. In: *IEEE electron device letters* 8.1 (1987), pp. 24–26.
- [3] Loi D. Nguyen et al. “50-nm self-aligned-gate pseudomorphic AlInAs/-GaInAs high electron mobility transistors”. In: *IEEE transactions on electron devices* 39.9 (Sept. 1992).
- [4] Joel Schleeh. “Cryogenic ultra-low noise InP high electron mobility transistors”. PhD thesis. Gothenburg, Sweden: Chalmers university of technology, 2013.
- [5] Dae-Hyun Kim and Jesús A. del Alamo. “Scalability of sub-100 nm InAs HEMTs on InP substrate for future logic applications”. In: *IEEE transactions on electron devices* 57.7 (July 2010), pp. 1504–1511.
- [6] Arnulf Leuther et al. “35 nm mHEMT technology for THz and ultra low noise applications”. In: *The 25th international conference on indium phosphide and related materials*. Kobe, Japan, 2013.
- [7] *Sentaurus Device user manual*. Synopsys Inc. June 2012.
- [8] E. F. Schubert. “Delta doping of III-V compound semiconductors: Fundamentals and device applications”. In: *Journal of vacuum science and technology* 8.3 (1990).
- [9] S. M. Sze and K. K. Ng. *Physics of semiconductor devices*. 3rd ed. Hoboken, New Jersey, USA: John Wiley & Sons incorporated, 2007.
- [10] Joel Schleeh et al. “Characterization and modeling of cryogenic ultralow-noise InP HEMTs”. In: *IEEE transactions on electron devices* 60.1 (Jan. 2013), pp. 206–212.
- [11] N. A. Semikolenova et al. “Energy band narrowing in InAs”. In: *Soviet physics: Semiconductors* 12.10 (1990), pp. 1139–1142.
- [12] S. C. Jain, J. M. McGregor, and D. J. Roulston. “Band-gap narrowing in novel III-V semiconductors”. In: *Journal of applied physics* 68.7 (Oct. 1990), pp. 3747–3749.
- [13] Mikael Malmkvist. “Optimization of narrow bandgap HEMTs for low-noise and low-power applications”. PhD thesis. Gothenburg, Sweden: Chalmers university of technology, 2008.

- [14] Javier Mateos et al. “Improved Monte Carlo algorithm for the simulation of δ -doped AlInAs/GaInAs HEMT’s”. In: *IEEE transactions on electron devices* 47.1 (Jan. 2000), pp. 250–253.
- [15] Helena Rodilla et al. “Cryogenic performance of low-noise InP HEMTs: A Monte Carlo Study”. In: *IEEE transactions on electron devices* 60.5 (May 2013), pp. 1625–1631.
- [16] J. L. Thobel et al. “Electron transport properties of strained InGaAs”. In: *Applied physics letters* 56.4 (1990), pp. 346–348.
- [17] V. Balyunas et al. “Time-resolved, hot-electron conductivity measurement using an electro-optic sampling technique”. In: *Applied Physics A* 51.4 (Oct. 1990), pp. 357–360.
- [18] M. Son et al. “Simulation of the DC and the millimeter-wave characteristics of $0.1\text{ }\mu\text{m}$ offset Γ -shaped gate $\text{In}_x\text{Ga}_{1-x}\text{As}/\text{In}_{0.52}\text{Al}_{0.48}\text{As}/\text{GaAs}$ MHEMTs with various $\text{In}_x\text{Ga}_{1-x}\text{As}$ channels”. In: *Journal of the Korean physical society* 44.2 (2004), pp. 408–417.
- [19] Yutaro Nagai et al. “Comparative study on frequency limits of nanoscale HEMTs with various channel materials”. In: *The 25th international conference on indium phosphide and related materials*. Kobe, Japan, May 2013.
- [20] S. D. Cho et al. “Comprehensive characterization and modeling of the abnormal gate leakage current in pseudomorphic HEMTs”. In: *Journal of the Korean physical society* 40.4 (Apr. 2002), pp. 577–583.
- [21] Dae-Hyun Kim and Jesús A. del Alamo. “Lateral and vertical scaling of $\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$ HEMTs for post-Si-CMOS logic applications”. In: *IEEE transactions on electron devices* 55.10 (Oct. 2008), pp. 2546–2553.
- [22] R. Pierobon et al. “Study of breakdown dynamics in InAlAs/InGaAs/InP HEMTs with gate length scaling down to 80 nm”. In: *2004 international conference on indium phosphide and related materials*. Kagoshima, Japan, 2004.
- [23] John Michael Golio. *Microwave MESFETs and HEMTs*. Artech House, 1991.
- [24] Madhu S. Gupta. “Power gain in feedback amplifier a classic revisited”. In: *IEEE transactions on microwave theory* 40.5 (May 1992), pp. 864–879.
- [25] Yoshimi Yamashita et al. “Pseudomorphic $\text{In}_{0.52}\text{Al}_{0.48}\text{As}/\text{In}_{0.7}\text{Ga}_{0.3}\text{As}$ HEMTs with an ultrahigh f_T of 562 GHz”. In: *IEEE electron device letters* 23.10 (Oct. 2002), pp. 573–575.
- [26] Jung-Ho Ahn et al. “Novel method for accurate extraction of f_{\max} for nano-scale MOSFETs”. In: *Journal of the Korean physical society* 45.5 (Nov. 2004), pp. 1201–1204.
- [27] Kevin J. Chen et al. “High-performance InP-based enhancement-mode HEMT’s using non-alloyed ohmic contacts and Pt-based buried-gate technologies”. In: *IEEE transactions on electron devices* 43.2 (Feb. 1996), pp. 252–257.
- [28] Akira Endoh et al. “InP-based high electron mobility transistors with a very short gate-channel distance”. In: *Japanese journal of applied physics* 42.4B (Apr. 2003), pp. 2214–2218.

- [29] Niklas Rorsmann et al. “Accurate small-signal modeling of HFETs for millimeter-wave applications”. In: *IEEE transactions on microwave theory and techniques* 44.3 (Mar. 1996), pp. 432–437.
- [30] Tae Whan Kim and Do-Hyung Kim. “InGaAs HEMT with InAs-rich InAlAs barrier spacer for reduced source resistance”. In: *Electronic letters* 47.6 (Mar. 2011), pp. 406–407.
- [31] M. Schirmer et al. *Chemical semi amplified positive e-beam resist (CSAR 62) for high resolution*. Allresist GmbH. 2014.
- [32] *UV60 DUV photoresist*. Rohm and Haas Company. Jan. 2014.
- [33] Gary S. May and Costas J. Spanos. *Fundamentals of semiconductor manufacturing and process control*. Hoboken, New Jersey, USA: John Wiley & Sons incorporated, 2006.

Appendix A

Gate Experiments

One of the desired outcomes of the investigation into the scalability of the Chalmers HEMT were to be able to utilize the results in future decisions regarding the in-house manufacturing process. This chapter aims to describe the HEMT fabrication process through the documentation of a series of experimental trials carried out as a means to evaluate it's usefulness in the fabrication of shorter gates.

A.1 Fabrication

A number of T-gates reminiscent of the structure shown in figure 3.1 on page 11 were manufactured with differently sized gate feet using the in-house process. A test pattern was used, consisting of 7 individual gates differing only in the size of the lithographic region forming their feet. The structures were produced 20 μm apart on a 16 mm \times 16 mm chip from an epi-wafer of the kind described in section 3.1. In order to facilitate the process of cleaving the chip prior to inspection of their cross-section, the structures were made 2 mm wide. The lithographic pattern was repeated in a 4×4 grid in order to some extent account for the limited yield of the manufacturing process.

As the purpose of the experiment was to investigate only the possibility to manufacture the gates and not the transistor as a whole, the process described below was modified to omit the processing steps involving the passivation of the device and the formation of the drain and source contacts.

A.1.1 Preprocessing of the epi-wafer

Prior to applying the electron beam (e-beam) resists used in the formation the mask, a few actions were taken to clear the surface from contaminants. The initial cleansing was performed by placing the chip in an ultrasonic cleaner (utilizing acetone as its solvent). In order to remove possible acetone residues, the step was followed by a chemical cleansing of the surface using isopropyl alcohol (IPA), followed by a regular cleaning using deionized water. The presence of oxides produced on the cap surface due to air exposure the last step was reduced by a final treatment using a 1:10 mixture of hydrochloric acid (HCl) and deionized water.

A.1.2 Formation of mask

The epi-wafer was heated to 160 °C for 5 min to improve the adhesion between the exposed barrier layer and the layer of resist used in the formation of the gate feet. A chemically amplified e-beam resist [31] was spin coated onto the surface at a rate of 4000 r/min with the intent to produce a 200 nm thick film. The film was then heated at 160 °C for an additional 5 min as an effort to prevent chemical binding with the upper layer of resist.

The uppermost film, consisting of a more sensitive resist [32], was spun onto the chip at a rate of 4000 r/min, producing a secondary film approximately 650 nm thick.

The patterning of the masks was carried out through direct exposure of the resist using e-beam lithography. As opposed to optical lithography, the resolution of e-beam lithography is limited by electron scattering rather than by optical diffraction [33]. Electrons scatter as they penetrate the resist and the underlying material, causing them to either leave the surface through backscattering or to transfer their energy to resist in close proximity of the area of exposure. The latter occurrence is known as the proximity effect and effectively places a limit on the minimum spacing that can be achieved between features.

Immediately after exposure, a post exposure bake (PEB) was performed on the chemically amplified resist for 2 min at 130 °C in order to complete the photo-reaction initiated by the lithographic process.

A.1.3 Development of e-beam resist

The PEB was followed by the development of the resists. The exposed parts of the uppermost resist were dissolved by submerging the sample for 60 s into a special purpose developer [31]. The residues were removed with water.

In order to release possible chemical bonds expected to arise between the molecules of the different resists during baking, the sample was exposed to a highly reactive oxygenic plasma. The remaining resist was developed using ortho-xylene for 2 min. Remaining oxide residues around the base of the foot were removed using a relatively weak solution of 1 part HCl per 25 parts water, due to the critical nature that the interface poses to the device performance.

A.1.4 Recess cavity etching

The formation of the recess cavity was controlled by utilizing the mask prior to forming the gate contact. A highly selective etchant solution based on succinic acid ($C_4H_6O_4$) and hydrogen peroxide (H_2O_2) was prepared with a 9 to 2 parts mixing ratio. The idea is that, due to its high selectivity towards InGaAs, the solution would reach the cap layer through the open areas of the mask and gradually penetrate underneath the resist. As the InAlAs of the barrier was also etched, albeit at a slower rate than the cap, the exposure time was critical to the resulting barrier thickness of the finished device. Aluminum oxides arising from the acidic reaction with the InAlAs were carefully removed from the cavity using a weak (1:25) hydrochloric water-based solution.

A.1.5 Metal deposition

The metal stack that constitutes the gate was deposited onto the mask through evaporation using a thin film deposition system. In order to improve the adhesion between the barrier material and the gate structure, an initial 20 nm layer of Titanium (Ti) was deposited, followed by a 10 nm thick layer of Pt. The latter was added to prevent the diffusion of gold (Au) atoms into the barrier. The remaining part of the gate foot and hat was formed by depositing a final 320 nm thick layer of Au.

A.1.6 Lift-off

The metals which were not part of the gate structures were removed by dissolving the remaining resist underneath. The structuring method is known as lift-off and involves the removal of all material that mechanically supports the T-shaped structures.

The upper film of resist was dissolved using acetone for 15 min. The bottom layer of resist was dissolved using the more delicate solvent anisole, effectively removing the one thing that held the gate hats in place. The fabrication of the device was finalized by dissolving the residues of anisole using acetone, which itself was dissolved in IPA, leaving the chip completed.

A.2 Verification and inspection

The results of the process described above was investigated. The chip was cleaved and the cross section of the gates were inspected in a SEM. The yield of the process was particularly good, apart from a few structures which became distorted during lift-off. SEM micrographs of the fabricated structures are shown in figure A.1.

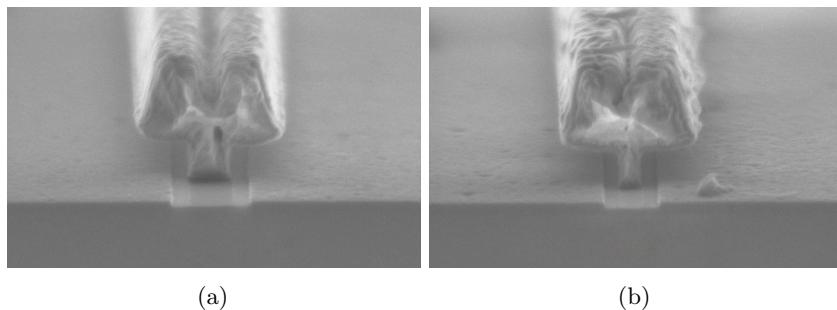


Figure A.1: SEM micrographs of the test structures fabricated with gate feet being 100 nm (a) and 40 nm (b) long.