Advanced III-Nitride Technology for mm-Wave Applications

Anna Malmros

Microwave Electronics Laboratory
Department of Microtechnology and Nanoscience - MC2
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden, 2019
Advanced III-Nitride Technology for mm-Wave Applications

Anna Malmros

© Anna Malmros, 2019

Doktorsavhandlingar vid Chalmers Tekniska Högskola
Ny serie nr 4594
ISSN 0346-718X

Technical Report MC2-414
ISSN 1652-0769
Microwave Electronics Laboratory

Department of Microtechnology and Nanoscience – MC2
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Tel. +46 (0)31 772 1000

Printed by Chalmers Reproservice
Gothenburg, Sweden, May 2019
Abstract

Within wireless communication, there is a continuously growing need for more bandwidth due to an increasing number of users and data intense services. The development within sensor systems such as radars, is largely driven by the need for increased detection range and robustness. In such systems, power amplification and generation at high frequency are of importance. High-electron mobility transistors based on gallium nitride (GaN HEMTs) offer efficient generation of high output power at high frequency. This is enabled by the unique characteristics of GaN and its heterostructures, with a large breakdown field, related to the wide bandgap, and excellent electron transport properties. Due to this, it is today used in high-performing radar, telecommunications, as well as power electronic systems. Despite substantial progress over the last decade, the GaN HEMT is still the subject of intense research to reach its full potential.

Recent development within epitaxy has significantly improved the quality of III-nitride semiconductors, and enabled indium aluminum nitride (InAlN) and InAlGaN as alternatives to AlGaN in the conventional AlGaN/GaN heterostructure. The higher polarization charge in these materials allows for considerable downscaling of the barrier layer thickness with a sustained high sheet carrier density. This has opened new possibilities for optimization of the high frequency performance.

In this work, HEMTs with downscaled InAl(Ga)N barrier layers have been developed with the goal to optimize the devices for power amplification in the mm-wave range. Electron trapping and short-channel effects have been addressed in the design of the epi and in the optimization of the process modules. Different surface passivation layers and deposition methods have been evaluated to mitigate electron trapping at the surface. The output power density of a HEMT increased from 1.7 to 4.1 W/mm after passivation with a SiN$_x$ layer. The deposition method for Al$_2$O$_3$ passivation layers showed to have a profound impact on the electron trapping. A layer deposited by plasma-assisted atomic layer deposition (ALD) exhibited superior passivation of the surface traps as compared to the layer deposited by thermal ALD, resulting in an output power at 3 GHz of 3.3, and 1.9 W/mm, respectively. The effect of the channel layer thickness (50 – 150 nm) in InAlGaN/AlN/GaN HEMTs with and AlGaN back barrier demonstrated a trade-off between short-channel effects and deep-level electron trapping in the back barrier. The maximum output power was 5.3 W/mm at 30 GHz, obtained for a GaN layer thickness of 100 nm.

To further enhance the high frequency performance, the ohmic contacts were optimized by the development of a Ta-based, Au free, metal scheme. Competitive contact resistance of $< 0.2 \, \Omega \cdot \text{mm}$ was achieved on both AlGaN/GaN and InAlN heterostructures with a Ta/Al/Ta metal stack. The contacts are annealed at a low temperature (550 – 575 ºC) compared to more conventional contact schemes, resulting in a smooth morphology and good edge acuity.

The implementation of microwave monolithic integrated circuits (MMICs) based on III-nitride HEMTs facilitate the use of III-nitride HEMTs in a system where frequency and compactness are key requirements. Thin film resistors (TFRs) are one of the passive components required in MMICs. In this work, a low-resistance titanium nitride (TiN) TFR was developed as a complement to the higher resistance tantalum nitride (TaN) TFR and mesa resistor in the in-house MMIC process. The developed TiN TFR exhibits a sheet resistance of $10 \, \Omega/\square$, compared to 50 and 200-300 $\Omega/\square$ of the TaN TFR and semiconductor resistor, respectively. The critical dissipated power in the TFR showed a correlation to the footprint area, indicating that Joule-heating was the main cause of failure. TiN- and TaN films exhibit different signs of the thermal coefficient of resistance. This feature was used to demonstrate a temperature compensated TFR (TCR = -60 ppm ºC) with application in MMICs operating in a wide temperature range.

List of appended publications

This thesis is based on the following papers:


Other publications


# Table of Contents

Chapter 1  Introduction  

Chapter 2  III-nitride HEMTs  
- 2.1  GaN-based heterostructures  
  - 2.1.1  The AlGaN/GaN heterojunction  
  - 2.1.2  Polarization in III-nitrides and origin of the 2DEG  
  - 2.1.3  Carrier transport  
- 2.2  HEMT operation  
  - 2.2.1  Small-signal operation  
  - 2.2.2  Large-signal operation  
  - 2.2.3  Short-channel effects  
  - 2.2.4  Downscaling of InAl(Ga)N-based HEMTs  
  - 2.2.5  Electron trapping  
- 2.3  Fabrication process  

Chapter 3  Ohmic contacts  
- 3.1.1  Ti-based contacts  
- 3.1.2  Ta-based contacts  
- 3.1.3  Recess etching  

Chapter 4  Critical aspects of InAl(Ga)N HEMTs  
- 4.1  Mobility  
  - 4.1.1  GaN interlayer  
- 4.2  Electron confinement  
- 4.3  Passivation  

Chapter 5  Thin Film Resistors  
- 5.1.1  Thermal dependence  
- 5.1.2  TiN TFRs  
- 5.1.3  Temperature compensated TFRs  
- 5.1.4  TFR reliability  

Chapter 6  Conclusion and future outlook  

Summary of appended papers  

Acknowledgement  

References
Chapter 1 Introduction

Mm-waves encompass the frequency range between 30 - 300 GHz, an interval suitable for a large variety of applications, including satellite communications [1]–[3], imaging [4], and motion sensor radars [5]. In mobile communication systems, the ever-increasing demands for higher data rates are pushing the development towards higher operating frequencies, with potentially wider channel bandwidth. Mobile telecommunication networks are currently going through a transition to the 5th generation (5G), which target a wide range of frequency bands. Several are located near 30 GHz, but even operation at frequencies in the D-band (110 – 170 GHz) is considered [6], [7]. For wireless communication, a large output power density is advantageous due to smaller components, increased efficiency, and increased bandwidth. The gallium nitride (GaN) high-electron mobility transistor (HEMT) is a technology that shows great potential to meet these demands.

The first GaN HEMTs were based on an AlGaN/GaN heterojunction, and this material system is still extensively used today. Initially, the operating frequency of the GaN HEMT was limited to a few GHz. Radar systems with low requirements on high frequency performance constituted the major application area, and the development was mainly driven by defense industries. Radar remains a very important application. The large output power density is a great advantage in radar transmitters. Further, the GaN HEMT is also considered for receiver applications, implemented in a low-noise amplifier (LNA) [8]. The noise figure is comparable to the GaAs pHEMT and in addition the GaN HEMT offers robustness, which leads to high survivability [9]. A monolithically integrated radar transceiver in GaN technology is therefore feasible [10]. The highest level of integration is achieved for implementation in a microwave monolithic integrated circuits (MMIC). MMICs are circuits in which all elements are processed on a single piece of epitaxially grown material. They are typically a few mm² in size, which is considerably smaller than the alternative of mounting the components on a circuit board.

Since the advent of the GaN HEMT in the early 1990’s [11], the technology has undergone a remarkable development in terms of high frequency performance. This has been possible through the improvement of the material quality and of the processing methods, and by the optimization of the heterostructure and the device design, for instance by downscaling. However, the downscaling of the AlGaN layer thickness is limited by the simultaneous loss of carrier electrons. On the other hand, the sheet carrier density is dependent on the strength of the built-in polarization field. Progress within epitaxy has enabled the growth of InAlN, InAlGaN, and AlN, with higher polarization fields, allowing for downscaling while maintaining a high sheet carrier concentration.

In this thesis, HEMTs based on InAlN/GaN and InAlGaN/GaN are explored as an alternative to the AlGaN/GaN technology, with the objective to provide large-signal amplification at higher operating frequencies. The adoption of these III-nitrides has already generated many promising results. A cutoff frequency of 374 GHz was demonstrated in [12]. An output power of 5.6 W/mm at 10 GHz was shown in [13], and 5.8 W/mm at 40 GHz was reported in [14]. Nevertheless, several challenges remain.

1. Electron trapping
   Electrons are captured and emitted in trap states in the buffer and at the surface. This leads to dynamically varying charge close to the 2DEG, which leads to lower output power, decreased efficiency, and memory effects due to its modulation of carriers in the channel [15].
2. **Scaling effects**
   The device dimensions, in particular the gate length in relation to other device geometries, may prevent the HEMT from being fully switched off at high lateral electric fields. The phenomenon is called short-channel effects (SCEs) and leads to reduced gain, output power, and efficiency [16].

3. **Self-heating**
   GaN HEMTs dissipate a lot of power that cause self-heating. Some mitigation can be provided by the use of a SiC and diamond substrate which have a high thermal conductivity [17]–[19]. However, the issue still remains since the conductivity of intermediate layers is often poor [20]. Self-heating reduces the mobility and is also plays a role for the device reliability.

4. **Reliability**
   The growth of HEMT structures on non-native substrates introduces relatively high dislocation densities. This, in combination with high electric fields and current densities, a high level of dissipated power, and the sensitivity to surface states makes the reliability of III-nitride HEMTs demanding [21]–[25]. Under large-signal operation, both the high electric field and the dissipated power can be problematic. A critical point is next to the gate, where the electric field is the highest.

5. **Electron mobility**
   The electron mobility is generally lower in InAl(GaN) based heterostructures compared to AlGaN/GaN, which has been attributed to a larger degree of alloy disorder in the barrier layer [26]–[28].

6. **Ohmic contacts**
   Despite numerous reports on ohmic contacts with a low contact resistance, repeatability is still an issue in GaN HEMT technologies [29], [30]. Furthermore, a smooth surface morphology and edge acuity are important aspects.

Electron trapping and short-channel effects have been given special focus in this work. These issues have been addressed through the optimization of the heterostructure and of the processing modules. The use of a back barrier reduced short-channel effects but introduced more electron trapping. The thickness of the intermediate GaN channel layer was optimized with respect to these two phenomena. To reduce surface trapping, different passivation layers and their impact on the HEMT performance have also been investigated. The issue of lower mobility was addressed through the growth of a thin intermediate GaN layer, which resulted in a higher mobility, but also a loss of carrier electrons. This heterostructure was investigated by temperature dependent characterization of the HEMTs. Further, ohmic contacts based on Ta have been developed. The ambition was to minimize the contact resistance and thereby reduce losses, and to achieve a repeatable process.

To fully take advantage of the HEMTs, the ultimate goal is the implementation in MMIC circuitry, and enable highly integrated circuits operating in the mm-wave range. To reach this goal, a number of building blocks are required. Apart from the transistor technology, passive elements are needed. These include capacitors, inductors, and resistors. New thin film resistor based on TiN and TaN have been developed. Aspects such as reliability and temperature dependence have been taken into account.
Chapter 2 III-nitride HEMTs

2.1 GaN-based heterostructures

The first HEMT was demonstrated by Mimura et al at Fujitsu Laboratories in 1980 [31]. The heterojunction was an AlGaAs/GaAs system, where the top AlGaAs layer was Si-doped. A significantly improved mobility was measured compared to a reference MESFET. A little more than a decade later, a HEMT based on an AlGaN/GaN heterojunction was reported by Khan et al [11]. The mobility was in the 500 cm²/Vs range, and the transconductance of the HEMT 28 mS/mm. Since then, the GaN HEMT technology has experienced a remarkable progress and is now offered by numerous companies, e.g. MACOM Technology Solutions [32], CREE, Inc. [33], Infineon Technologies [34], and NXP Semiconductors [35].

2.1.1 The AlGaN/GaN heterojunction

GaN is a group III-nitride semiconductor. These compounds consist of one or more elements from group III in the periodic table of elements (In, Al, and Ga), and nitrogen. GaN exhibits a unique set of properties, shown in Table 1. It has a bandgap, $E_g$, around three times larger than that of Si, and more than twice that of GaAs. The large bandgap results in a large breakdown field $E_{br}$. SiC has a similar bandgap but lacks one of the properties of the III-nitrides: the possibility of growing heterostructures. Lastly, GaN has a high peak velocity. The great interest in GaN HEMT technology arises from the unique combination of material properties, which enables a large output power density as well as a fast device.

<table>
<thead>
<tr>
<th>Table 1. Material parameters of common semiconductors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandgap $E_g$ (eV)</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>1.1</td>
</tr>
<tr>
<td>Breakdown electric field $E_{br}$ (MV/cm)</td>
</tr>
<tr>
<td>Peak electron velocity $v$ (10⁷ cm/s)</td>
</tr>
<tr>
<td>Electron mobility $\mu$ (cm²/Vs)</td>
</tr>
</tbody>
</table>

The GaN HEMT is based on a heterojunction formed by the growth of GaN layer and a III-nitride barrier layer. The bandgap ($E_g$) of the III-nitrides are determined by their composition and span a wide range from 0.7 (InN) to 6.2 eV (AlN). The bandgap versus lattice constant $a$ for the ternary compounds AlGaN, InAlN, and InGaN is depicted in Fig. 2-1. The barrier layer has a wider bandgap than the GaN layer. This leaves AlGaN and InAlN as potential alternatives. The conventional GaN HEMT uses an AlGaN/GaN heterojunction. Due to the lattice mismatch between AlGaN and GaN, the AlGaN will be subjected to strain, which increases with the Al mole fraction. Strain in the epilayer increases the probability of dislocations and partial relaxation, with negative consequences for the reliability. This limits the maximum Al content in AlGaN to roughly 30% if the layer is not highly downscaled. InAlN on the other hand can be grown lattice-matched to GaN, as indicated by the red dashed line. Lattice
match is obtained at an In content of 17 – 18%. In recent years, the quality of InAlN and the quaternary InAlGaN has improved. Thus, today these two constitute alternatives to the more established AlGaN.

**III-nitride HEMTs**

The principle behind the HEMT as opposed to other field effect transistor (FET) technologies is the spatial separation of carrier electrons from the donor atoms. Thereby, the electron scattering caused by collisions between electrons and ionized donor atoms can be avoided, which increases the mobility. Separation is achieved from the different electron affinities, $\chi_1$ and $\chi_2$, of the barrier- and the GaN layers, shown in Fig. 2-2a. The electron affinity is defined as the difference between vacuum level ($E_{\text{vac}}$) and the bottom of the conduction band ($E_c$). This results in a conduction band discontinuity $\chi_2 - \chi_1$. The carrier electrons originate from the barrier layer but will attain the lower energy level on the GaN side. There they congregate at the interface and form a quantum well. Besides constituting the channel, the GaN layer also serves as a buffer layer, ensuring a good crystal quality near the interface.

The electrons in the quantum well are described to form a “two-dimensional electron gas” (2DEG), as indicated by the dashed line in Fig. 2-2b. In reality, the electron distribution is three-dimensional, which can be seen in Fig. 2-2c, showing the simulated band structure and electron density ($n$) of an Al$_{0.30}$Ga$_{0.70}$N/GaN heterostructure. Most electrons are found within a thickness of a few nm, while $n$ decays exponentially towards the buffer. A small fraction is located in the barrier layer. It is highly desired to confine the carrier electrons to the quantum well. If the electrons are located in the barrier layer, they will have considerably lower mobility, while wide distribution into the buffer layer indicates that the device may suffer short-channel effects at high electric fields. The confinement of the 2DEG at the channel - barrier interface plays a crucial role for the HEMT performance.

**Fig. 2-1.** Bandgap of the III-nitrides versus the lattice constant $a$. The dashed red line indicates the point where InAlN is lattice-matched to GaN.

**Fig. 2-2.** a) Energy band structure of a heterojunction, b) a GaN HEMT heterojunction, and c) the simulated energy band diagram and electron distribution of an AlGaN/GaN heterojunction with an Al mole fraction of 0.30.

Values of the electron affinity found in the literature vary to some extent due to the difficulties of accurate determination of this property. One method for the estimation of the electron affinity is to measure the ionization energy of the valence band ($E_v$) electrons, $E_{\text{vac}} - E_v$, by ultraviolet photoemission spectroscopy (UPS), and calculate the affinity by subtraction of the known bandgap of the material [36]. Fig. 2-3a shows the electron affinities for GaN [37], AlN [38], and InN [38], and for AlGaN and InAlN.
versus Al content, assuming a linear dependence. The resulting offset in the conduction band for these barrier layers grown on GaN is shown in Fig. 2-3b. Lattice-matched InAlN/GaN exhibits a larger conduction band offset compared to Al0.30Ga0.70N/GaN, which promotes better confinement from towards the barrier.

![Fig. 2-3. a) Electron affinity of AlGaN and InAlN with respect to the Al mole fraction, and b) the resulting band offset for an AlGaN/GaN and an InAlN/GaN heterojunction. Two commonly used compositions for AlGaN- and InAlN-based heterostructures are indicated.](image)

### 2.1.2 Polarization in III-nitrides and origin of the 2DEG

In Fig. 2-2c above, it can be seen that the energy bands of an AlGaN/GaN heterojunction are not flat. This is due to a built-in polarization field, an inherent property of the III-nitrides. The field arises because of a difference in electronegativity between the elements in the compound, which results in a displacement of the electron cloud. The crystal can therefore be envisioned as being made up of small dipoles. Due to the lack of symmetry, the dipoles at the boundaries are left dangling, resulting in a boundary charge, Fig. 2-4. This type of polarization is called spontaneous polarization, denoted \( \vec{P}_{sp} \). A second type of polarization is the piezoelectric polarization, \( \vec{P}_{pz} \), which arises in a layer under strain.

The total polarization \( \vec{P}_{tot} \) is

\[
\vec{P}_{tot} = \vec{P}_{sp} + \vec{P}_{pz}
\]

The strength of the built-in field varies with the composition of the compound. For AlGaN and InAlN, \( \vec{P}_{sp} \) increases with the Al mole fraction. AlGaN grown on relaxed GaN is subjected to tensile strain that results in a \( \vec{P}_{pz} \) with the same polarity as \( \vec{P}_{sp} \), thus adding to a stronger \( \vec{P}_{tot} \). Generally, the Al content in AlGaN is restricted to around 30% to avoid strain relaxation. Very thin layers can however be grown at a larger Al content.

For InAlN, \( \vec{P}_{pz} \) can be either positive or negative. For lattice-matched InAlN, \( \vec{P}_{pz} \) is zero, which means that the strain is nonexistent. This is an advantage, since strain may cause dislocations that reduce the mobility and have a negative impact on the device reliability. Fig. 2-5 shows \( \vec{P}_{tot} \) versus the Al mole fraction for InAlN and AlGaN. For lattice matched InAlN, \( \vec{P}_{tot} \) is nearly 40% larger than for AlGaN with an Al content of 30%. The strongest polarization field among the III-nitrides is exhibited by AlN. On the other hand, AlN grown on GaN suffers so much strain that the thickness is limited to around 2 – 4 nm.
In other HEMT material systems such as AlGaAs/GaAs, carrier-donor spatial separation is achieved through intentional n-doping of the barrier layer. The donated electrons will attain the lower energy level at GaAs side and thereby form a 2DEG. GaN HEMTs are different in this respect, since the 2DEG is formed without intentional doping. There is today a high level of consensus holding that the electrons originate from donor-like surface states. An intuitive model for the formation of the 2DEG has been provided by Ibbetson et al [40]. It assumes a single donor state and makes no assumption about Fermi level pinning at the surface. Fig. 2-6a depicts the band structure of a HEMT with a thin barrier layer (solid lines). A surface donor state is located at a level $E_D$ from the conduction band. When the donor state is below the Fermi level, it is highly probable that the state is filled. If the barrier is grown thicker (dashed lines), $E_F - E_D$ will decrease, and $E_D$ will eventually reach the Fermi level. An empty state implies that the electron has been excited to the conduction band, and if so, it will attain the lower energy level on the GaN side and contribute to the formation of a 2DEG. For a 2DEG to form, a certain critical thickness of the barrier layer is required, and $n_s$ continues to increase for a thicker barrier layer.

Ensuing from Ibbetson’s model, a barrier layer with a higher polarization also leads to a higher $n_s$, as illustrated in Fig. 2-6b. A third example is presented in Fig. 2-6c, showing the band structure when an additional GaN layer caps the surface. GaN caps are often used for improving the reliability [41],[42], but are accompanied by a drop in $n_s$. The model is useful for the prediction of the impact that a variation in the layer structure has on $n_s$. 

---

**Fig. 2-4.** Illustration of the polarization charges at the layer boundaries in the III-nitrides.

**Fig. 2-5.** The total polarization of AlGaN and InAlN grown on GaN versus Al mole fraction. The dashed lines indicate resulting polarization of commonly used mole fractions for each of the alloys [39].

**Fig. 2-6.** Principle of 2DEG formation according to Ibbetson’s model. The impact of growing a) a thicker barrier layer, b) a barrier with increased polarization, and c) a GaN cap layer.
2.1.3 Carrier transport

The sheet conductivity $\sigma$ in a HEMT structure can be calculated from

$$\sigma = q \mu n_s$$  \hspace{1cm} (2)

where $q$ is the elementary charge, and $\mu$ is the electron mobility. In the description of a HEMT heterostructure, the sheet resistance $R_{sh} = 1/\sigma$, is most often used instead. At low electric fields the electron velocity shows a linear dependence on the electric field $E$, where $\mu$ is the proportionality factor:

$$v = \mu E$$  \hspace{1cm} (3)

$\mu$ is limited by different scattering mechanisms, which all have an associated limiting mobility $\mu_s$, set by the scattering rate. The total effective mobility of a semiconductor can be approximated by Matthiessen’s rule:

$$\frac{1}{\mu_{tot}} = \frac{1}{\mu_1} + \frac{1}{\mu_2} + \ldots + \frac{1}{\mu_n}$$  \hspace{1cm} (4)

A scattering event is essentially a change in the electron energy and momentum. In GaN-based heterostructures, there are numerous scattering mechanisms. Phonon scattering (i.e. lattice vibrations) has a strong temperature dependence and dominates at room temperature but is insignificant at cryogenic temperature [43], [44]. Therefore, the mobility in GaN-based heterostructures generally improves quite dramatically as the temperature is reduced. Further scattering mechanisms in GaN include e.g. interface scattering and carrier-carrier scattering. In ternary compounds such as InAlN and AlGaN, alloy disorder causes scattering due to the varying potential within the non-perfect crystal. The scattering rate associated with this mechanism is assumed to be high in InAlN layers due to large compositional variations [45]. Charged impurities in the GaN channel is well-known scattering mechanism, sometimes referred to as Coulomb scattering [46]. This mechanism has shown a certain temperature dependence [47], opposite to that of phonon scattering and not as strong. Other scattering mechanisms such as alloy scattering and interface scattering are assumed to be temperature independent [44].

Several non-destructive measurement techniques are available for the determination of $n_s$, $\mu$, and $R_{sh}$ of the as-grown wafer, e.g. eddy current technique, microwave reflectance, and mercury probe C(V) characterization. These are commonly performed on the as-grown wafer, prior to processing. The data presented in this work have been measured by Hall characterization on van der Pauw [48] structures processed simultaneously with the HEMTs. Characterization of the material properties before- and after processing allows the study of any impact on the epitaxial material from the processing.

By the optimization of $n_s$ and the low-field mobility, $R_{sh}$ can be minimized and thereby the parasitic resistances can be reduced. For the intrinsic HEMT, the high-field properties of the material are of vital importance: The cutoff frequency $f_T$ is directly dependent on the saturation velocity $v_{sat}$ by

$$f_T = \frac{v_{sat}}{2\pi L_g}$$  \hspace{1cm} (5)

At high electric fields, the electron velocity is no longer linearly dependent on $E$, and $\mu$ is no longer constant. Fig. 2-7 shows $v$ as a function of the electric field for GaAs, Si, SiC, and GaN as obtained from Monte Carlo simulations. GaN has a very high $v_{sat}$. Similar to GaAs, the peak in $v$ is followed by a region with a negative differential mobility, but for GaN, the peak occurs at a very high electric field, and the drop in $v$ is less pronounced. The behavior of $v$ at high electric fields can be described by (from [37])

$$v(E) = \frac{\mu_0 E}{(1+(\mu_0 E/v_{sat})^{\beta})^{1/\beta}}$$  \hspace{1cm} (6)

where $\mu_0$ is the low-field mobility, and $\beta$ is a fitting parameter. However, Eq. (6) does not take the negative differential mobility into account. This behavior has been verified empirically up to the peak velocity by pulsed characterization of un-gated structures [49].
2.2 HEMT operation

The HEMT is a FET device with three terminals: gate, source, and drain. HEMTs are most often implemented with several gate fingers, like the two-finger HEMT shown in Fig. 2-8a. The gate is a Schottky contact with ideally no current flowing through it when reversed biased, while the source and drain terminals are low-loss ohmic contacts. The drain-source current $I_{ds}$ is set by the laterally applied $V_{ds}$ and the vertically applied $V_{gs}$. The magnitude of $I_{ds}$ in the saturation region depends on the number of carriers in the 2DEG, which is set by $V_{gs}$, as illustrated in Fig. 2-8b. For the investigated HEMTs in this work, a 2DEG is present at zero gate voltage. These HEMTs are therefore said to be normally on, or depletion-mode devices. The opposite, a normally-off device (enhancement mode device), can be fabricated by e.g. recessing the barrier layer in the gate region [51], [52], or by modifying the barrier by F-implantation [53]–[55].

Examples of DC characteristics ($I_{ds}$–$V_{ds}$) of a HEMT are shown in Fig. 2-8c. Some important parameters that can be directly observed are the on-resistance ($R_{on}$), the transconductance ($g_m$), see inset of Fig. 2-8c, and the output conductance ($g_{ds}$). $R_{on}$ is the slope of $I_{ds}$ in the linear region and has a direct impact on the knee voltage and current ($V_{knee}$ and $I_{knee}$). $g_m$ and $g_{ds}$ are defined by the equations

$$\begin{align*}
g_m &= \frac{\delta I_{ds}}{\delta V_{gs}} \\
g_{ds} &= \frac{\delta I_{ds}}{\delta V_{ds}}
\end{align*}$$

$g_m$ serves as a quality measure of the gate modulation, describing the incremental change in $I_{ds}$ for a change in $V_{gs}$. Similarly, $g_{ds}$ describes the incremental change in $I_{ds}$ for a change in $V_{ds}$. In the saturation region, $g_{ds}$ is ideally zero. As the gate length is reduced, the output conductance increases, making $g_{ds}$ a direct measure of short-channel effects, discussed further in 2.2.3.
2.2.1 Small-signal operation

In small-signal operation, the input- and output signals are so small that non-linear characteristics can be regarded as linear. By measuring the scattering parameters (s-parameters), the small-signal current- and unilateral power gains, $h_{21}$ and $U$ can be calculated:

$$h_{21} = \frac{-2S_{21}}{(1 - S_{11})(1 + S_{22}) + S_{12}S_{21}}$$

(9)

$$U = \frac{|S_{21}|^2}{(1 - |S_{11}|)^2 (1 + |S_{22}|)^2}$$

(10)

From these gains, two important figures of merit, $f_T$ and $f_{\text{max}}$, can be extracted. $f_T$ is the transit frequency (or cutoff frequency), which is the frequency where $h_{21}$ equals zero dB. $f_{\text{max}}$ is the maximum frequency of oscillation where $U$ equals zero dB. These figures are used to enable direct comparison of different technologies and gives an indication of the maximum frequency of operation for circuits implemented in these technologies. An example of the measured $h_{21}$ and $U$ for an InAlN/AIN/GaN HEMT is shown in Fig. 2-9. The figure also shows the extraction of $f_T$ and $f_{\text{max}}$ from these measurements.

![Fig. 2-9. The small-signal gains $h_{21}$ and $U$ and the determination of $f_T$ and $f_{\text{max}}$ by extrapolation.](image)

Small-signal characterization permits the extraction of an equivalent circuit model [56]–[58]. The model is conventionally divided into a bias independent extrinsic part and a bias dependent intrinsic part (Fig. 2-10). For accurate determination of the intrinsic elements, it is important that the extraction of the extrinsic elements is precise. These are obtained in two additional cold-FET (off-state) measurements, as described in [58].

![Fig. 2-10. Equivalent small-signal electrical model.](image)

A small-signal model is valid for one bias point only. The bias dependence of the different elements is obtained by multi-bias s-parameter measurements and model extraction. The bias-dependent parameters are often used as a first approximation for the extraction of large-signal, non-linear models.
III-nitride HEMTs

Fig. 2-11 shows the extracted values of \( g_m \) and \( C_{gs} \) (gate capacitance) versus \( V_{gs} \) for different values of \( V_{ds} \). Small-signal models provide valuable insight into the limiting parameters of the HEMT, which to some extent can be related to physical parameters of the device.

The intrinsic \( f_T \) can be described by \( g_m \) and \( C_{gs} \) and \( C_{gd} \) (gate-drain capacitance):

\[
f_{T,\text{int}} = \frac{g_m}{2\pi(C_{gs} + C_{gd})}
\]

Eq. (11) is sometimes approximated by omitting \( C_{gd} \), since \( C_{gs} \) is often considerably larger than \( C_{gd} \). \( f_{\text{max}} \) can be approximated with the following expression [59]:

\[
f_{\text{max}} \approx \frac{f_T}{2\sqrt{(R_i + R_s + R_g)g_{ds} + 2\pi f_TR_gC_{gd}}}
\]

To reach higher operating frequencies, the optimization of \( g_m \), \( C_{gs} \), and \( g_{ds} \) is paramount. It is desired to maximize \( g_m \) while minimizing \( C_{gs} \) and \( g_{ds} \). These parameters are related to the physical dimensions \( t_B \) (barrier thickness) and \( l_g \) (gate length), and gate width (\( W_g \)), shown in

\[
\begin{align*}
g_m & \sim \frac{\varepsilon}{t_B} V \\
C_{gs} & \sim \frac{\varepsilon}{l_B} l_g \\
g_{ds} & \sim \frac{q\mu n_s(t_B)W_g}{l_B}
\end{align*}
\]

\( t_B \) and \( l_g \) are clearly important factors in the high frequency optimization. The general approach to increase \( f_T \) and \( f_{\text{max}} \) is to simultaneously scale down \( l_g \) and \( t_B \). In this way \( g_m \) is increased, while both \( C_{gs} \) and \( g_{ds} \) are reduced. However, as discussed in 2.1.2, the downscaling of \( t_B \) implies a loss of carrier electrons, which limits the maximum drain current and ultimately the output power. This results in a trade-off between the high frequency performance and the output power.

2.2.2 Large-signal operation

Power amplifiers (PAs) and frequency converters operate under large-signal conditions. For PAs, it is desired to maximize the output power \( P_{\text{out}} \), gain, efficiency, and linearity. The current and voltage waves at the output port depend on the connected load impedance in combination with device properties. PAs are categorized into different classes depending on the bias conditions. In this work, large-signal characterization was generally performed with the HEMTs biased in class AB operation, which means that the transistor is conducting more than half of the cycle. Class AB is a good compromise in terms of gain, output power and efficiency.

Fig. 2-13. IV curves and superimposed load line of a HEMT in large-signal operation.
Maximum output power $P_{\text{out, max}}$ is achieved when the RF swing spans the entire load line. Hence, $P_{\text{out, max}}$ is limited by the knee in the IV characteristics, and also by the breakdown voltage. 

$$P_{\text{out, max}} = \frac{(V_{\text{max}} - V_{\text{knee}})(I_{\text{knee}})}{8}$$  \hspace{1cm} (16)

The efficiency is often given as the power-added efficiency, PAE:

$$PAE = \frac{P_{\text{out}} - P_{\text{in}}}{P_{\text{DC}}}$$  \hspace{1cm} (17)

In this work, large-signal characterization has been performed using load- and source pull to find the output matching for maximum output power and efficiency. After the load/source pull sequence, a power sweep is performed, where the input signal is increased until a specific gain compression is reached. Both passive- and active load pull has been performed. The active load-pull system is based on a large-signal network analyzer (LSNA), which enables the direct measurements of the current- and voltage waveforms [60]. Active load-pull has been used up to a fundamental frequency of 10 GHz, allowing for enough harmonics to reconstruct the waveforms. For higher frequencies (up to 40 GHz), passive load-pull was used. Passive load-pull relies on impedance tuners that can present different load impedances to the device under test.

### 2.2.3 Short-channel effects

The downscaling of the HEMT dimensions and high-voltage operation can lead to the phenomenon known as short-channel effects (SCEs). It occurs when the electrons gain enough energy from the applied lateral electric field to pass underneath the depleted area of the reverse biased HEMT. The consequence is that the HEMT channel cannot be completely pinched. HEMTs with short gates are particularly sensitive to SCEs, since they cannot deplete the channel as efficiently, and require a larger magnitude of the applied voltage to fully pinch the device.

In Fig. 2-14, the phenomenon of SCEs is visualized with TCAD simulations. The current in two AlGaN/GaN HEMTs with gate lengths of 200 and 50 nm, respectively, is compared at a bias $(V_{\text{gs}}, V_{\text{ds}}) = (-6, 20) \text{ V}$. For the longer gate this bias condition results in a pinched channel. For the short gate the depletion region is much smaller, which facilitates conduction through the buffer. The thickness of the barrier layer also plays a role. A thin barrier layer alleviates SCEs, but also results in fewer carriers in the 2DEG.

![Fig. 2-14. Simulated current density of two HEMTs on AlGaN/GaN with gate lengths of (a) 200 nm and (b) 50 nm. The HEMTs are biased at $(V_{\text{gs}}, V_{\text{ds}}) = (-6, 20) \text{ V}$.](image)

The impact of SCEs is visible in the IV characteristics, as shown in Fig. 2-15. The maximum negative gate voltage of -4.5 V is sufficient to pinch the HEMT at a drain voltage up to 10 V, whereas at higher drain voltage, the HEMT cannot be fully pinched-off.

The gravity of the SCEs is clearer in the transfer characteristics $(I_{\text{ds}}$-$V_{\text{gs}}$), Fig. 2-16. As $V_{\text{ds}}$ is increased, the voltage required to pinch the HEMT $(V_{\text{po}})$ is shifted towards more negative values. A
useful measure to quantify the SCEs is therefore the drain-induced barrier lowering, DIBL \[61\], defined as

\[
DIBL = \frac{V_{po,\text{high}} - V_{po,\text{low}}}{V_{ds,\text{high}} - V_{ds,\text{low}}}
\]

The determination of DIBL requires the definition of a pinch-off condition, in this work defined as a drain current \(I_{po}\) of 1 mA/mm. DIBL is given in mV/V. Another measure of SCEs is the subthreshold swing (SS) \[62\], \[63\], which considers the behavior of \(V_{gs}\) with respect to \(\log_{10}(I_{ds})\) below the threshold voltage, as shown in the inset of Fig. 2-16. SS is given in mV/decade and should be minimized.

\[
SS = \frac{\delta V_{gs}}{\delta \log_{10}(I_{ds})}
\]

In the extraction of DIBL it is important to choose the \(V_{ds}\) interval within a region which is relevant for the intended application, and similar reasoning should be applied to SS. Since there is no consensus for the extraction of these figures, it can be difficult to compare results in the literature.

In the IV characteristics, various SCEs are visible. Figure 2-15 shows the IV characteristics, while Figure 2-16 illustrates the extraction of DIBL and SS from \(I_{ds}-V_{gs}\) data.

**2.2.4 Downscaling of InAl(Ga)N-based HEMTs**

As described in 2.1.2, the sheet carrier concentration depends on the barrier thickness and the size of the built-in polarization field. Since InAlN has a stronger polarization field compared to conventional AlGaN (with an Al content up to around 30%), the required layer thickness for a given \(n_s\) is smaller. The realization of InAlN barrier layers has therefore offered a way to circumvent the limitations in the downscaling of AlGaN. It has been shown that Ga is easily incorporated during the growth of InAlN, generally attributed to residual Ga in the growth chamber \[64\]. This presumably means that some reported results on InAlN-based heterostructures may in fact be quaternary InAlGaN. Both InAlN and InAlGaN have been evaluated as barrier layers in this work, collectively addressed as InAl(Ga)N.

The dependence of \(n_s\) on the barrier thickness is illustrated for different barrier layers in Fig. 2-17. It is shown that AlGaN must be grown thicker to achieve the same 2DEG density as the InAl(Ga)N barrier. A result that stands out is the pure AlN barrier (although capped with a thin GaN layer). Having the highest polarity of the III-nitrides, a mere 3.5 nm thick AlN layer produces an \(n_s\) of \(1.2 \cdot 10^{13} \text{ cm}^2\).
To give an idea of the size of the small-signal equivalent circuit parameters obtained for the HEMTs on a typical epi in this work, $g_m$ and $C_{gs}$ for an InAlN HEMT from [F] are plotted in Fig. 2-18. A comparison is made with two HEMTs fabricated on a downscaled AlGaN-based epi. The two heterostructures, shown in Fig. 2-18a, have total barrier thicknesses of 7.5 and 13.5 nm, respectively. The InAlN HEMT has a gate length of 170 nm, while the AlGaN HEMTs have gate lengths of 100 and 200 nm. The HEMTs are biased at a drain voltage of 5 V. The InAlN HEMT exhibits a $g_m$ around 140 mS/mm higher than that of the AlGaN HEMTs Fig. 2-18b. $C_{gs}$ of the InAlN HEMT was in the same order as for the AlGaN HEMT with a gate length of 200 nm, Fig. 2-18c. The data are consistent with the general results in this work, in which the maximum $g_m$ has ranged from 750 to 900 mS/mm, and $C_{gs}$ from around 1 – 2 pF/mm, largely dependent on the gate length.

### 2.2.5 Electron trapping

Surface states as well as imperfections in the buffer layer, such as dislocations and impurities, can act as electron traps. Filled electron traps exert the 2DEG to an electric field, which partially depletes the GaN channel layer of electrons (Fig. 2-19). This leads to dispersive effects, generally defined as unwanted variations of device response dependent on frequency or bias conditions. For instance, a parameter extracted from DC characterization may not exhibit the same value under RF operation, or the device exhibits different behavior when transiting from different quiescent bias points.

During large-signal operation, electron trapping manifests mainly in a reduction in the output power, efficiency, and linearity. To counteract electron trapping at the surface, a passivation layer, commonly SiNx or Al$_2$O$_3$, can be used, while the mitigation of buffer trapping is mainly a matter of the heterostructure design and growth technique.

During large-signal operation, electron trapping manifests mainly in a reduction in the output power, efficiency, and linearity. To counteract electron trapping at the surface, a passivation layer, commonly SiNx or Al$_2$O$_3$, can be used, while the mitigation of buffer trapping is mainly a matter of the heterostructure design and growth technique.

There are many phenomena related to trapping effects. Some can be readily observed by pulsed IV characterization. In these measurements, double pulsing of $V_{gs}$ and $V_{ds}$ is performed from a quiescent
III-nitride HEMTs

point (V_{gs,Q}, V_{ds,Q}) to an active point where the IV data is measured. At certain quiescent voltages, the electrons will be trapped, while at (V_{gs,Q}, V_{ds,Q}) = (0, 0) V quiescent voltage represents the ideal case, e.g. with no traps activated by an electric field.

![Illustration of surface- and buffer traps, depleting the 2DEG.](image)

In Fig. 2-19, the results of pulsed IV characterization of two HEMTs from [D], here denoted HEMT 1 and 2, are shown. A quiescent voltage (V_{gs,Q}, V_{ds,Q}) = (-5, 0) V is compared to (0, 0) V. HEMT 1 shows no signs of trapping, while HEMT 2 exhibits a lower current for a negative quiescent gate voltage. The phenomenon, known as current collapse, is caused by partial depletion of the 2DEG. The impact of electron trapping by the application of a gate voltage is also known as the gate lag effect. The notion of a virtual gate is also commonly used [82], referring to the impact of filled surface states, which can be regarded as a gate extension.

In Fig. 2-20a, a HEMT from [E] exhibits an increase in the on-resistance for a non-zero quiescent voltage, in this case (-5, 30) V. The phenomenon is caused by depletion of the 2DEG in the channel region and in the adjacent access regions. The HEMT exhibits a simultaneous displacement of the knee of the IV curve, referred to as knee walkout. Similar to the gate lag effect, dispersion caused by electron trapping incited by a voltage applied to the drain terminal, is referred to as the drain lag phenomenon.

![Dispersive effects revealed in pulsed IV measurements. a) Current collapse related to surface trapping [D], and b) Dynamic on-resistance and knee walkout phenomena caused by buffer trapping [E].](image)

A definite distinction between surface- and buffer traps is hard to make. As a rule of thumb, a negative gate voltage fills the surface traps, causing current collapse, while a high drain voltage fills the buffer traps, leading to an increased on-resistance and knee walkout. However, the quiescent voltage must be chosen so that the HEMT is fully pinched in order to avoid any impact from self-heating. For a depletion mode device this means that a negative gate voltage below V_{pc} is applied also when the impact of a quiescent drain voltage is measured, which may add a contribution from surface trapping to the measured results. Further, the drain voltage also increases the field near the gate and may contribute to surface trapping. However, conclusions can often be drawn based on the observed phenomena (e.g. current collapse and knee walkout), and the design of the experiment.

The impact of electron trapping has been quantitatively estimated by a number of figures in this work. They include slump ratios, defined as

\[
Z_n = \frac{(I_{ds}(BP_n) - I_{ds}(BP_0))}{I_{ds}(BP_0)} \cdot 100
\]

(20)
where $I_{ds}$ is the current measured in the active point. Under the influence of trapping, $I_{ds}$ depends on the quiescent bias point BP$_0 = (V_{gs,Q}, V_{ds,Q})$. BP$_0$ represents $(V_{gs,Q}, V_{ds,Q}) = (0, 0)V$. Further, the impact of trapping can be estimated by the shift in the knee voltage $\Delta V_{knee}$ and the associated drop in the output current $\Delta I_{knee}$. Another very useful measure is the relative increase in the on-resistance, $R_{on,inc}$. The extraction of these measures, as performed within this work, is depicted in Fig. 2-21. The figure shows the pulsed IV characteristics from two quiescent bias points BP$_1$ (for estimation of surface trapping) and BP$_2$ (for estimation of buffer trapping). The slump ratio $Z_1$ is extracted at any voltage above the knee, whereas $Z_2$, $\Delta V_{knee}$, and $\Delta I_{knee}$ are extracted at the knee, indicated by the black points. $R_{on,inc}$ is extracted in the linear region at low drain voltage.

![Fig. 2-21. Pulsed IV characterization comparing the bias points BP$_0$, BP$_1$, and BP$_2$, and an indication of where the parameters $Z_1$, $Z_2$, $\Delta V_{knee}$, $\Delta I_{knee}$, and $R_{on,inc}$ are extracted.](image)

### 2.3 Fabrication process

The fabrication of a HEMT includes the definition of an active area, ohmic contacts for the drain- and source terminals, gate formation, contact pads, passivation, and opening of the passivation layer to enable probing. The sequence and methods may vary between different process houses. The active area can be defined by mesa formation through dry etching, or by ion implantation, commonly utilized in industrial GaN technologies. Ion implantation relies on heavy ion bombardment of the region outside the active areas, causing crystal damage that renders the material isolating. A wide range of ions can be used, e.g. Ar$^+$ [83] or O$^+$ [84]. Ion implantation is generally preferred due to its planar characteristics, which is favorable for the repeatability of the gate definition and lack of any issues regarding the mesa sidewall [85], [86].

Conventionally, the ohmic contact is formed by annealing of a metal stack to form a tunneling contact on a highly doped layer. Because of the high anneal temperature, the ohmic contact formation is usually performed prior to the gate step. Optical lithography generally meets the demands on the resolution and alignment for the definition of mesas and ohmic contacts. For gates in the range of 100’s of nm range, e-beam lithography (EBL) is required. Ni is frequently used as Schottky contact for its good adhesive properties and large work function, $\Phi_{m}$, resulting in a large barrier height. Contact pads may be deposited by sputtering or evaporation.

In conventional GaN HEMT processing, Au is used for coverage of ohmic contacts, gates, and electrodes to avoid oxidation. However, it has been shown that long-term degradation of ohmic contacts may occur if Au diffuses and intermix with Al [87]. An Au-free metallization scheme would improve reliability and reduce the production cost. Furthermore, and Au-free process would be compatible with Si-processing. In particular, ohmic contacts can be challenging to fabricate without Au, as will be discussed in Chapter 3. Suggestions for Au-free Schottky diode metallization schemes include e.g. Al/W [88], pure Ni [89], and Ti/Al/W [90].
Passivation may be performed after the gate formation ("passivation last-process") [91]. In this case, it is important that the deposition temperature does not cause any harm to the device. In a metal-insulator-semiconductor HEMT (MISHEMT), the gate dielectric, which constitutes at least part of the passivation layer, is deposited prior to gate formation [92], [93]. Surface passivation may also be performed as the very first step in the process ("passivation first-process"). In this approach, openings are defined for the gate-, source-, and drain contacts. This is frequently done with SiN\(_x\) passivation deposited in-situ [94],[95], or ex-situ deposition using low-pressure chemical vapor deposition (LPCVD) [96]. The advantage is the immediate protection of the semiconductor surface. The most commonly used dielectric for surface passivation is SiN\(_x\), deposited by plasma-enhanced chemical vapor deposition (PECVD) [97], [98].

In this work, the fabrication process is a basic process, suitable for the evaluation of the epitaxially grown heterostructure, the process modules, and the device performance. The process flow is depicted in Fig. 2-22. Active areas are defined by mesa isolation, performed by optical lithography and a Cl\(_2\)/Ar-based dry-etching process. Ohmic contacts are formed by metallization of the contact area followed by annealing. Details of the work on ohmic contacts are presented in Chapter 3.

In this work, the fabrication process is a basic process, suitable for the evaluation of the epitaxially grown heterostructure, the process modules, and the device performance. The process flow is depicted in Fig. 2-22. Active areas are defined by mesa isolation, performed by optical lithography and a Cl\(_2\)/Ar-based dry-etching process. Ohmic contacts are formed by metallization of the contact area followed by annealing. Details of the work on ohmic contacts are presented in Chapter 3.

Gates are defined by e-beam lithography (EBL), and metallized with a Ni/Pt/Au scheme. The gate length of the fabricated HEMTs range from 50 to 200 nm. Following the gate formation, Ti/Al electrodes are deposited. Notably, Au free contacts are prone to form a surface oxide, resulting in an additional resistance between the contact metal and the electrode. Sputtered electrodes combined with in-situ Ar\(^+\) cleaning has shown to be successful in terms of forming a non-resistive direct connection between the electrode and the ohmic contact. Passivation is performed by the deposition of Al\(_2\)O\(_3\) by plasma-assisted atomic layer deposition (ALD), with an exception in [C], where SiN\(_x\) is deposited by PECVD. In the final step, openings in the passivation layer in the regions of the electrodes are defined by optical lithography. The removal of the passivation layer is in the case of Al\(_2\)O\(_3\) passivation done by wet etching HF:H\(_2\)O 1:10, and in the case of SiN\(_x\) passivation by dry etching.
Fig. 2-23 shows three images related to the fabrication process. A microscope image of the HEMT after the gate formation is shown in Fig. 2-23a. Fig. 2-23b shows a FIB-SEM image of the cross section of a HEMT with a gate length of 170 nm. The finalized HEMTs are shown in Fig. 2-23c.
Chapter 3 Ohmic contacts

Low on-resistance is key for high frequency operation and efficiency of HEMTs. $R_{\text{on}}$ is determined by the sum of the contact resistances ($R_c$) and access resistances, related to the sheet resistance ($R_{\text{sh}}$) of the epi, see Fig. 3-1. Thus, the minimization of $R_{\text{on}}$ necessitates for $R_c$ to be reduced. For the processed HEMTs within this work, $R_c$ often constitutes to a significant part of the on-resistance. As an example, a typical 2x50 $\mu$m InAl(Ga)N HEMT has an $R_c$ of 0.4 $\Omega$mm, a material sheet resistance of 230 $\Omega$/square, and a source-drain distance of 1.5 $\mu$m. This HEMT has a minimum $R_{\text{on}}$ of 12 $\Omega$, where 8 $\Omega$ (67%) corresponds to the contribution from the ohmic contacts.

![Fig. 3-1. Depiction of the elements contributing to the on-resistance](image)

Through its contribution to the on-resistance, the parasitic contact resistance affects the high frequency performance, output power, and efficiency. Some of the highest ever reported values of $f_T$ and $f_{\text{max}}$ were reported for HEMT with an extremely low $R_c \leq 0.01$ $\Omega$mm [99]. By setting the knee of the IV characteristics, the on-resistance limits the RF swing in large-signal operation. Further, the voltage drop over the ohmic contacts results in losses, which reduce the efficiency. The impact can be observed by calculating the extrinsic transconductance, $\Delta m_{\text{ext}}$:

$$
\Delta m_{\text{ext}} = \frac{\Delta m_{\text{int}}}{1 + \frac{R_{\text{sh}}}{R_{\text{sh}}}}
$$

(21)

In Fig. 3-2a, an example of the measured IV data of HEMTs processed on two pieces from the same wafer is shown. The samples differed in the contact resistance, which was 1.0 and 0.4 $\Omega$mm, respectively, resulting in a clear impact on the on-resistance, knee voltage and the maximum current. The extrinsic transconductance is shown for the same devices versus on-resistance in Fig. 3-2b. The transconductance improved from 420 to 480 mS/mm for the piece with the lower contact resistance.
Ohmic contacts

Despite the numerous publications demonstrating a low $R_c$, ohmic contacts to III-nitride heterostructures remain the subject of extensive research. Issues with repeatability and difficulties of achieving a versatile process that works for any heterostructure are revealed in the very wide spread in reported $R_c$ values. Moreover, these results underline the lack of full understanding of the contact mechanism and the impact on $R_c$ by tuning the process parameters. Further motivations for continued exploration include requirements on the contact formation step in terms of integration with the full fabrication process. A smooth surface morphology of the contacts is preferred primarily because the alignment marks used for gate definition are most often defined in the ohmic layer. Furthermore, the surface morphology is related to the edge acuity. A sharp edge facilitates the lateral downscaling and prevents device-to-device variations in the breakdown voltage, caused by a spread in the actual source-drain distance [100]. A minimized thermal budget during fabrication is also desired because of any impact that it might have on the heterostructure. For example, a high anneal temperature has been associated with a higher vertical leakage through the buffer in [101].

The GaN HEMT is essentially a planar device, thus the major part of the current transfers between the ohmic contact metal and the 2DEG through a window marked in red in Fig. 3-3. This part is very small in comparison to the total length $L$ of the contact. The transfer length ($L_T$) is defined as the part through which a fraction $1/e$ of the current flows. As long as $L$ is significantly longer than $L_T$, the resistance of the contact is independent of $L$, but scales inversely with the contact width $W$. Therefore, $R_c$ is generally normalized to $W$ and given in the unit $\Omega\text{mm}$.

An $R_c$ below $0.2\ \Omega\text{mm}$ is generally considered to be excellent, targeted for instance when optimizing for a high $f_{\text{max}}$.

The most straightforward way of measuring $R_c$ is by using the transfer length method (TLM). The method uses a test structure consisting of a set of contacts separated by different spacings $d_1 \ldots d_n$, Fig. 3-4a. Four-point probe measurements are performed, forcing a small current while the voltage drop over each contact pair is measured by a second set of probes. $R_c$ is found by fitting a line to the data and extrapolate to the point where $d = 0$, corresponding to the resistance from two contacts, Fig. 3-4b. The slope of the fitted line is determined by the sheet resistance of the epi. $L_T$ can be determined from TLM measurements since $2L_T$ corresponds the point where the fitted line and the x-axis intersect. The non-normalized resistance $R_{\text{tot}}$ between two contacts, spaced by a distance $d_x$, is written as in Eq. (22). Thus, the slope of the fitted line depends on the sheet resistance of the epi.

$$R_{\text{tot}} = 2R_{\text{contact}} + R_{sh}\frac{d_x}{W} \quad \text{(22)}$$

Sometimes the specific contact resistance $\rho_c$, given in $\mu\Omega\text{cm}^2$, is used to evaluate contacts and enable comparison to vertical devices. $\rho_c$ is often calculated using the equation

$$\rho_c = R_{sh}L_T^2 \quad \text{(23)}$$

However, it should be noted that the sheet resistance under the contact is likely not the same as between the contacts, which makes the equation an approximation. Alternatively, $\rho_c$ can be measured directly on vertical Cross Bridge Kelvin Resistor (CBKR) structures [102], but with limited accuracy.
due to the lateral flow in the contacts. A test structure for a more reliable determination of $L_T$ was proposed in [103], yet such a structure requires a very precise definition of the dimensions.

The conventional contact is based on a Ti/Al metal scheme and requires annealing at high-temperature. This is still the most extensively used ohmic contact for GaN HEMT technologies. An emerging fabrication is re-growth, a method that utilizes epitaxy to grow highly Si-doped GaN in a recess etched contact region, where both MOCVD and MBE can be utilized. A potentially very low contact resistance can be obtained, yet this more advanced method faces similar problems of repeatability as the conventionally processed contacts, and therefore the results are scattered. For MBE-grown contacts, a contact resistance around $0.1 \, \Omega \text{mm}$ [104], [105], and $0.26 \, \Omega \text{mm}$ [106] have been reported, while results on MOCVD grown contacts include $0.23 \, \Omega \text{mm}$ [107], $0.33 \, \Omega \text{mm}$ [108], and $0.45 \, \Omega \text{mm}$ [108]. Notably, recess etching is becoming increasingly common also in conventionally annealed contacts, having shown to enable a lower $R_c$ compared to planar contacts. With this approach, the annealed contacts have also shown to produce excellent contacts with an $R_c$ of e.g. $0.17 \, \Omega \text{mm}$ [109] and even $0.1 \, \Omega \text{mm}$ [110].

A small selection of reported results of $R_c$, that to some degree is representative for ohmic contacts to the III-nitrides is shown in Fig. 3-5. The contacts were formed to AlGaN-, InAlN- and AlN barriers by the different methods discussed in this chapter. Regular metallized - annealed contacts are displayed versus anneal temperature in Fig. 3-5a, whereas Fig. 3-5b shows the results for regrown contacts. The picture is intended to provide an overview of results that are commonly achieved, even though it is not completely fair to compare the fabrication methods for non-identical heterostructures. However, the very low $R_c$ in [A] for Epi I is deliberately omitted since the AlGaN layer in this heterostructure has a significantly lower Al mole fraction compared to the other references.

Some general conclusions from the figure can be drawn. For a metallized, annealed contact, an $R_c \leq 0.2 \, \Omega \text{mm}$ normally requires recess etching, but has been reported as well for planar contacts. Ta-based ohmic contacts can compete with the conventional Ti-based contacts, which require a higher anneal temperature. The very low $R_c$ of $0.08 \, \Omega \text{mm}$ for regrown contacts is very impressive and was achieved for an AlN barrier, for which a planar contact would presumably be highly resistive due to a large $\Phi_h$. Although numerous reports have shown that low-resistive ohmic contacts to GaN heterostructures are possible, the challenge of obtaining a repeatable, versatile process with minimum impact on the epi still remains.

---

**Ohmic contacts**

---

![Fig. 3-3. Illustration of current transfer (red) between the metal contact and the 2DEG, the contact dimensions and transfer length marked out.](image)

![Fig. 3-4. TLM test structure and corresponding $R$ versus distance $d$ plot.](image)
3.1.1 Ti-based contacts

The primary active element in the conventional Ti-based contact approach is Ti. However, Ti alone does not produce good ohmic contacts, but requires an additional Al layer. Au is generally used to cap the layer stack to avoid oxidation, but Al and Au have a tendency to form highly resistive alloys, which leads to poor surface morphology [128]. Therefore an intermediate layer of e.g. Ni [129], or Mo [112] is generally incorporated. This is often not fully adequate, and therefore, for instance W and TiN have been proposed as substitutes to Au [130]. However, it has often been shown difficult to achieve a low $R_c$ without Au, which could indicate that the Au layer has a role besides preventing oxidation. It is speculated in [131] that a low work function Au-containing phase at the metal – semiconductor interface promotes current flow, and further that reported optimum Ti/Al layer thickness ratios [132] are highly influenced by which reactions with Ni and Au that occur.

The optimum anneal temperature of the Ti-based contact lies around 800 - 900 °C. The advantages of a lower thermal budget have incited the development of different strategies to reduce the anneal temperature. Recess etching prior to metallization has shown to facilitate ohmic contact formation, and resulted in an $R_c$ of 0.21 Ωmm for the Ti-based metal scheme annealed at 550 °C [114]. A recent approach is microwave annealing [133]–[135]. The method has been claimed to locally increase heating by means of the eddy current caused by a varying magnetic field [136].

According to the theory that has gained most acceptance regarding the contact mechanism, ohmic behavior in the conventional contact is obtained by the reaction between the Ti and N, forming TiN [137],[138],[139] or other phases containing Ti [140]. This creates N vacancies in the barrier layer, which act as n-dopants [141]. The result is a heavily doped barrier layer at the metal-semiconductor interface, and potentially further down into crystal. The heavy doping bends the conduction band, permitting electrons to tunnel between the contact metal and the 2DEG through field emission (FE). The conclusion is based on the identification of the compound(s) adjacent to the semiconductor surface, which can be performed by e.g. TEM/EDX analysis [142] or by XRD [111]. The role of Al is not as clear and has been debated. It has been proposed that Al works as a catalyst for Ti to react with N [143]. It has also been claimed that Al slows down the aggressive reaction in which TiN is formed, based on the observation of voids beneath the TiN when pure Ti is annealed on GaN and AlGaN [144].

TCAD simulations of a metal/Al$_{0.25}$Ga$_{0.75}$N/GaN structure was performed for different superficial doping levels $N_d$ to demonstrate the impact of N vacancies on the conduction band. The barrier height $\Phi_b$, equal to

$$\Phi_b = \Phi_m - \chi$$ (24)
where $\chi$ is the electron affinity of the AlGaN, was assumed to be 1.5 eV [145]. The AlGaN is 25 nm thick and the doping of the 1 nm top part is stepped from zero to $1.5 \times 10^{21}$ cm$^{-3}$. For comparison, regrown ohmic contacts have been fabricated with a doping level of $1 \times 10^{19}$ to $1 \times 10^{20}$ cm$^{-3}$ [105], [126]. For the highest doping level, the conduction band within the AlGaN is approaching $E_F$, leaving only a very narrow barrier, which promotes tunneling.

![Fig. 3-6. Simulations of a metal - Al$_{0.25}$Ga$_{0.75}$N/GaN contact with an electron barrier of 1.5 eV. The top 1 nm part of the AlGaN is varied from intrinsic to heavily doped at the indicated doping concentrations.]

Although the Ti-based contact is essentially a tunneling contact, it does not exclude other factors from having an impact on $R_c$. More commonly seen in earlier publications is the reporting of metallic protrusions stretching along dislocations in the barrier layer, thereby forming a direct contact to the 2DEG [116]. Some have emphasized the low work function of TiN (3.74 eV [143]) as being a major contributing factor through the reduction of the effective barrier height [111]. If the band structure looks like in Fig. 3-6, tunneling is facilitated by a higher thermal energy of the electrons. For the highest simulated doping level, an additional energy of around 0.2 eV above $E_F$ would be required for a high electron tunneling probability. This current mechanism, combining thermionic emission and field emission, is referred to as thermionic field emission, TFE [146]. In [F], $R_{on}$ showed an increase at very low temperature. Since no increase in the measured $R_{sh}$ was observed, the increase was concluded to arise from $R_c$. This could be an indication of a lower probability of tunneling through the barrier when the thermal energy of the electrons is very low, and that the contact mechanism is TFE, dominated by tunneling.

From the discussion above it is clear that a low $\Phi_b$ promotes current flow. Since $\chi$ decreases with the Al mole fraction [147], the benefits of a high Al content in terms of $n_s$ and electron confinement comes at the cost of a more challenging ohmic contact fabrication. The tuning of $\Phi_m$ as a way of decreasing $\Phi_b$ is further complicated, since the main role of the contact metal is the extraction of N, which Ti has shown to do very efficiently, inevitably forming TiN at the interface.

### 3.1.2 Ta-based contacts

Ta has previously been incorporated in the metallization schemes for ohmic contacts to III-nitrides. It has been used as a diffusion barrier between Al and Au [148], and together with Ti at the bottom of the metal stack [149]. In a patent application by Kanamura et al [150] a Ta/Al-based scheme for ohmic contact formation was proposed. This metallization has also been used previously on n-GaN in [151], and to AlGaN in [152]. The contact mechanism of Ta-based contacts is believed to be similar to that of the Ti-based contact, since TaN has been identified at the interface [153]. A potential disadvantage of using the Ta/Al metallization scheme is that TaN has a larger $\Phi_b$ (> 4 eV) than TiN [143].

A thorough investigation of Ta/Al as an alternative to Ti/Al was made in [A]. Planar contacts were formed on two AlGaN-based heterostructures, with Al mole fractions of 0.14 and 0.25, respectively. Optimization of the metal layer stack in terms of thicknesses and cap layers was performed, as well as of the anneal temperature and duration. From the discussion on $\Phi_b$ above it follows that a low $R_c$ should be more easily obtained on the material with the low Al mole fraction. Indeed, the contacts on this
Ohmic contacts

material exhibited a very low $R_c$ below $< 0.1 \, \Omega \text{mm}$. Due to the lack of reported results for contacts fabricated on AlGaN/GaN with a similarly low Al mole fraction, it is difficult to make a comparison with the literature. The minimum $R_c$ of the epi with the higher barrier Al mole fraction was $0.28 \, \Omega \text{mm}$, which is in line with what is commonly seen for Ti-based contacts on comparable heterostructures [132][154].

Capping the stack with Au and an interlayer of Ni or Ta like in conventional ohmic contacts generally led to a higher $R_c$ and a reduced process window in terms of the thickness of the Al layer. TEM/EDX analysis indicated a high level of interdiffusion between the metals, resulting in Ni and Au near the interface, which likely explains the results. A plain Ta/Al/Ta stack on the other hand led to a considerable improvement. The scheme has an $R_{sh}$ of around $0.5 \, \Omega/\square$, as measured on test structures. Thereby the contact metal itself constitutes a very small part of the measured $R_c$.

A striking feature of the Ta-based contact is the low optimum anneal temperature of 550 – 575 ºC. The absence of Au combined with the low anneal temperature resulted in a very smooth surface morphology. The Ta/Al/Ta metallization has shown to be compatible with early passivation [78], [155]. Fig. 3-7 shows AlGaN/GaN samples with ohmic contacts formed after the deposition- and opening of a SiN$_x$ passivation layer. The samples were metallized with Ta/Al/Ta and Ti/Al/Ni/Au schemes and annealed at 550 and 830 ºC, respectively. As can be seen, the rough morphology of the Ti-based contact has severe consequences for the edge acuity.

![Fig. 3-7. Ohmic contacts separated by 3 µm, processed after the deposition and opening of a SiN passivation layer. a) A Ta/Al/Ta contact annealed at 550 ºC and b) a Ti/Al/Ni/Au contact annealed at 830 ºC.](image)

The evolution of $R_c$ versus anneal duration in [A] shows that the process saturates, with no risk for over-annealing. This opens for a sequential anneal process, interrupted when the contact resistance is no longer decreasing. For a thin bottom Ta layer, $R_c$ saturates faster, which can be seen in Fig. 3-8, comparing two contacts with Ta thicknesses of 50 and 200 Å, respectively. This could indicate that Al near the interface promotes the contact formation. The Ta/Al/Ta scheme has shown as well to produce good ohmic contacts to InAl(Ga)N/GaN. It was further investigated in [B] combined with recess etching and was utilized as well in [C-F].

![Fig. 3-8. Development of $R_c$ versus time for two different thicknesses of the bottom Ta layer [A].](image)
3.1.3 Recess etching

Recess etching provides a way to circumvent the limitations imposed by a high $\Phi_b$, since it enables the contact to be formed with a reduced barrier thickness (enhancing tunneling probability), or at the sidewall of the trench, in close proximity to the 2DEG. It has been shown that recessed contacts can reduce $R_c$ compared to planar contacts [113], and the target $R_c$ is often in the 0.15 – 0.3 $\Omega$mm range [156][157]. On the other hand, repeatability is a greater challenge for recessed contacts due to the introduction of more parameters, such as recess depth, sidewall angle, resist hardness, and etching damage. The recessing is performed by dry-etching, often in a Cl$_2$ or Cl$_2$/BCl$_3$ plasma [156], [158], but the more mechanical bombardment by Ar sputtering has also been suggested [159]. Full consensus on the optimum recess depth has not been reached. Previous studies have reported scattered results. A minimum $R_c$ was found by etching part way through the barrier [113], while the complete removal of the barrier provided the best results in [117].

In this work, recessing was performed by the combined inductively couple plasma etching (ICP) and reactive ion etching (RIE) in a Cl$_2$/Ar-based plasma. Descumming in a low-power oxygen plasma and subsequent oxide stripping in diluted HCl prior to metallization was performed in order to remove any etching products and superficial oxide. In [B], the Ta/Al/Ta metallization was applied to recess etched contact areas of an InAlN/AlN/GaN heterostructure, with a primary focus of optimizing the recess depth. Recess etching for 5 different durations was performed, spanning a recess depth from 0 (planar contacts) to beyond the 2DEG (8 nm below the surface), see Fig. 3-9. The optimum level was found if etching was interrupted just above the 2DEG, with approximately 1 nm of AlN remaining. At this level, there is presumably still a 2DEG in the recessed region due to the high polarization of AlN. Reduced sensitivity to other parameters such as the bottom Ta layer thickness was observed at the optimum etch depth, a minimum contact resistance was 0.14 $\Omega$mm was obtained.

In order to improve the repeatability and versatility of the ohmic contact fabrication process step, it is important to identify the parameters that play a role in the contact formation. It is highly suspected that the semiconductor surface is affected by the etching, which is believed to increase $\Phi_b$. This was discussed in [160], where a higher $R_c$ was observed for a shallow recess compared to the planar contact. Another factor that could play a role in the contact formation is the angle of the sidewall [161]. As shown in [B], the etched trenches do not exhibit vertical sidewalls, Fig. 3-11. The assumed impact of the sidewall angle $\alpha$ is illustrated in Fig. 3-12. Near the trench at the level of the 2DEG, the barrier is thin, which promotes tunneling. On the other hand, a too thin barrier does not support a 2DEG, why a depletion region is formed, which decreases the tunneling probability.
Ohmic contacts

In [162], the recess of re-grown contacts on GaN/AlGaN/GaN was investigated. It was found that a more vertical wall provided a smaller $R_c$. It was also suggested that a large $\alpha$ is favorable due to a comparatively small density of dangling bonds, since these raise the conduction band above the Fermi level. In addition to this, recessing the ohmics may require process adjustments to obtain the desired effective layer thicknesses and ensure proper metal coverage of the sidewall, as pointed out in [d]. Notably, the situation is likely to be different in InAl(Ga)N/AlN/GaN heterostructures due to the large polarization in the AlN layer, i.e. an AlN can be very thin without a complete depletion of the 2DEG underneath. From this perspective, a smaller $\alpha$ is presumably possible without a degradation in $R_c$, which was partly supported by TEM analysis in [B].
Chapter 4 Critical aspects of InAl(Ga)N HEMTs

High frequency optimization of HEMTs requires a shorter gate length. However, the downscaling of the gate length alone eventually leads to short-channel effects, which limit the high frequency performance, output power and efficiency. The possibility of scaling down the barrier layer by replacing the conventional AlGaN with InAl(Ga)N while maintaining a high $n_s$ was discussed in 2.2.4. This permits a reduced gate length without a simultaneous increase in short-channel effects. In addition, it is possible to grow a lattice-matched layer, which allows for growth without any strain-induced dislocations.

The high polarization in InAlN and the associated large $n_s$ were predicted [163] before the first appearance of InAlN/GaN structures around 2005 [164], [165]. Since then, progress within epitaxy has improved the quality of these heterostructures. Still, the growth of InAlN is still considered more challenging than AlGaN, and it is well-known that InAlN may suffer from lower crystalline quality, compositional variations [166], [167], and increased gate leakage [168]. Part of the explanation lies in the fact that the optimum growth temperatures of InN and AlN differ by several hundred degrees [169]. It is therefore difficult to find the growth conditions that suit both. There is also a large difference in the bond strength between InN and AlN, which makes the InAlN prone to phase separation [167].

Additional Ga in the layer is not necessarily a disadvantage. It has been pointed out by many that the difficulty of alloying InN with AlN arising from the difference in bond strength is mitigated in InAlGaN growth [170][171]. Furthermore, numerous excellent results on HEMTs fabricated with this quaternary barrier have been reported. Medjdoub et al reached an $f_{\text{max}}$ of nearly 200 GHz using an In$_{0.16}$Al$_{0.70}$Ga$_{0.12}$N barrier [172]. A breakdown voltage of 850 V was measured for a MISHEMT based on InAlGaN epi in [173], and a power density of 6W/mm at 30 GHz was reported in [124]. Hence, many epitaxy groups intentionally grow InAlGaN layers. The InAlGaN barrier was adopted in [E], with a composition of 12% In, 80% Al, and 8% Ga, respectively. At this composition the layer is no longer lattice matched to GaN. The calculated lattice, as calculated by Vegard’s law [174], is 3.174 Å, which is smaller than that of GaN, indicating tensile strain. The resulting piezoelectric polarization adds to the spontaneous polarization field.

4.1 Mobility

InAl(Ga)N-based material systems have generally exhibited a lower mobility compared to AlGaN/GaN, which is a disadvantage to the otherwise very promising material heterostructure. To fully take advantage of InAl(Ga)N as a barrier layer, it is therefore of interest to understand this limitation and explore methods of increasing $\mu$. A compilation of reported values of $\mu$ at RT for different heterostructures is provided in Fig. 4-1. In the following discussion, the presence or absence of an additional GaN cap is ignored. Within each category of barrier layers, results are scattered. Nevertheless, some conclusions can be drawn. Pure AlN barriers have demonstrated moderate $\mu$. In both the AlGaN/GaN and InAl(Ga)N/GaN systems, the insertion of a thin AlN exclusion layer has resulted in a higher $\mu$ [175], [176]. This is most apparent for the InAl(Ga)N barrier. The InAlN/GaN heterojunction has produced a mobility in the order of 300 cm$^2$/Vs [177], while an intermediate AlN
Critical aspects of InAl(Ga)N HEMTs

layer usually raises the mobility to 1100-1600 cm²/Vs [178],[179],[126]. Due to the substantial difference, InAl(Ga)N/GaN is today practically never grown without an AlN exclusion layer. The enhanced mobility is often attributed to reduced interface scattering and better confinement of the 2DEG to the GaN layer, thereby reducing alloy scattering [28], [180].

![Fig. 4-1. Mobility from references](image)

Through its direct impact on the sheet resistance, a high mobility is crucial for the reduction of the on-resistance. The on-resistance affects both the small-signal- and large-signal performance. With a minimized on-resistance, a higher \( f_T \) and \( f_{\text{max}} \) can be expected. For the large-signal performance, a smaller on-resistance results in higher output power, gain, and efficiency. To illustrate the extent of this effect, a set of simulations were performed, see Fig. 4-2. The generated s-parameters of a small-signal model corresponding to a 2x50 µm InAlGaN HEMT were used to extract \( f_T \) and \( f_{\text{max}} \), while varying the on-resistance (\( R_s + R_d \)). Results are plotted in Fig. 4-2a. A major impact is observed, primarily for \( f_{\text{max}} \). In Fig. 4-2b-d, the results of a large-simulation are shown. The model corresponds to a 2x25µm InAlN HEMT, which is biased at \( V_{ds} = 5 \) V and operated at a frequency of 10 GHz. \( R_s + R_d \) were varied from 0.13 to 3.75 Ωmm, while the other parameters remained unchanged. By increasing the extrinsic resistances within the selected range, the output power dropped by 2.7 dBm to 0.4 W/mm, Fig. 4-2b. The transducer gain, shown in Fig. 4-2c, dropped by 3 dB. The impact was distinct in the power-added efficiency, which dropped from 42 to 26 %, Fig. 4-2d.

![Fig. 4-2. Simulated impact of the on-resistance on a) \( f_T \) and \( f_{\text{max}} \), b) the output power, c) the transducer gain, and d) efficiency](image)

---

Through its direct impact on the sheet resistance, a high mobility is crucial for the reduction of the on-resistance. The on-resistance affects both the small-signal- and large-signal performance. With a minimized on-resistance, a higher \( f_T \) and \( f_{\text{max}} \) can be expected. For the large-signal performance, a smaller on-resistance results in higher output power, gain, and efficiency. To illustrate the extent of this effect, a set of simulations were performed, see Fig. 4-2. The generated s-parameters of a small-signal model corresponding to a 2x50 µm InAlGaN HEMT were used to extract \( f_T \) and \( f_{\text{max}} \), while varying the on-resistance (\( R_s + R_d \)). Results are plotted in Fig. 4-2a. A major impact is observed, primarily for \( f_{\text{max}} \). In Fig. 4-2b-d, the results of a large-simulation are shown. The model corresponds to a 2x25µm InAlN HEMT, which is biased at \( V_{ds} = 5 \) V and operated at a frequency of 10 GHz. \( R_s + R_d \) were varied from 0.13 to 3.75 Ωmm, while the other parameters remained unchanged. By increasing the extrinsic resistances within the selected range, the output power dropped by 2.7 dBm to 0.4 W/mm, Fig. 4-2b. The transducer gain, shown in Fig. 4-2c, dropped by 3 dB. The impact was distinct in the power-added efficiency, which dropped from 42 to 26 %, Fig. 4-2d.
4.1.1 GaN interlayer

Despite the enhancement in µ ensuing from an AlN exclusion layer, a µ exceeding ~1600 cm²/Vs is rarely encountered in InAl(Ga)N-based heterostructures. In [F], the incorporation of a GaN interlayer between the InAlN and AlN layers was demonstrated as a viable way of improving µ. At RT, a µ of 1930 cm²/V was measured, compared to 1600 cm²/Vs of a reference sample without IL. The finding provide insight into the current transport and limiting scattering mechanisms in these materials. When comparing the GaN IL material with the reference, it was found that the reference suffered additional scattering corresponding to a temperature independent limiting mobility of around 12700 cm²/Vs.

The adoption of a GaN IL has previously produced a high mobility (2603 cm²/V at T = 270 K) in InAlN-based heterostructures [187]. The enhanced µ was attributed to a reduced number of electrons in the InAlN, leading to reduced alloy scattering, and to improved surface morphology, which was related to a smoother GaN/InAlN interface. In paper [F], two other possible scattering mechanisms are suggested, which are more related to the compositional quality of the barrier layer

i) Coulomb scattering due to variations in the boundary charge, caused by a non-uniform composition of the InAlN layer. The increased spatial distance between introduced by the GaN IL mitigates this type of scattering, which has previously been claimed to be temperature independent [188].

ii) Scattering due to subband energy fluctuations, here referred to as SEF. In a quantum well, the potential energy of the electrons is quantized into discrete levels. The subband energies ε₁, ..., εₙ (with respect to the Fermi level) depend on the shape of the quantum well, which changes with the size (and distribution of) of n. Since n correlates to the composition of the barrier layer, it follows that the composition has an impact on ε₁, ..., εₙ. This scattering mechanism was investigated in [189]. It was found that a more SEF scattering can be expected in an InAlN/GaN heterojunction compared to AlGaN/GaN for a given percental variation in the Al mole fraction due to the larger variation in polarization in InAlN. Fig. 4-3a the simulated ε₁ of the reference sample in [F] for two different Al mole fractions of the InAlN layer. In Fig. 4-3b, the simulated variation in ε₁ as a function of the Al content is shown for the GaN IL and reference samples. Also included are an AlGaN/GaN and an InAlN/GaN heterostructure without interlayers. The simulation indicates that the variation in ε₁ for the GaN IL sample is insignificant above the measured Al concentration of 86 %. This finding supports mitigated SEF as the cause reduced scattering. The figure also shows that the subband variations are considerably larger in InAlN/GaN compared to AlGaN/GaN, which could be contributing reason for the low µ seen in InAlN-based heterostructures without any interlayers.

Fig. 4-3. a) Illustration of the impact on the first subband ε₁ by a variation in the barrier polarization [F], and b) the variation in ε₁ versus Al mole fraction of the 6 nm InAlN layer a heterostructure without interlayers, and two heterostructures with AlN exclusion layers of different thicknesses.
Critical aspects of InAl(Ga)N HEMTs

Notably, the higher μ that is obtained with a GaN IL comes at the expense of fewer electrons in the 2DEG, in accordance with Ibbetson’s model for the 2DEG formation. Therefore, the difference in $R_{sh}$ is not as large as if the wafers would differ only in μ.

4.2 Electron confinement

Although the use of a downscaled barrier reduces SCEs, there is still a need for improved confinement due to the large $n_s$. A method to counteract SCEs is to provide better confinement by the buffer. There are two general methods that are frequently employed to accomplish this. One is to introduce a deep electron trap in the buffer, which main purpose is to compensate the non-intentional background doping. The two most common dopants are C and Fe, but also Be has been suggested for MBE-grown buffers [190]. The downside of compensational doping is that it can lead to electron trapping [191]. Therefore, it is desired to minimize the doping near the channel. C-doping has the advantage of offering a better control of the doping profile. A stepped profile has shown to result in good confinement and moderate dispersion in [78]. The incorporation is generally achieved by changing growth condition, which could cause a non-optimized crystal quality. Also, the C-doping level is very sensitive to the growth conditions, and there is a risk that it may act as an n-dopant, depending on which site (Ga- or N) occupies [192], [193].

The second commonly used method to improve the buffer confinement is to introduce a second heterojunction by the growth of a back barrier (BB). This approach was used in [E] in which an 800 nm thick AlGaN BB was grown underneath the GaN channel. The BB had a graded Al content with the highest mole fraction of 6 % at the BB/channel interface. Notably, the incorporation of a BB does not exclude the requirement of compensational doping, since background doping exists also in this layer. A proposed alternative to the AlGaN BB is to use a thin InGaN layer, sometimes referred to as an InGaN “notch” [194].

In this work, a specific AlGaN BB design was used, while the thickness of the channel layer and the gate length were varied to study the impact of back barriers to mitigate SCEs [E]. A clear effect was observed in the DC characterization. For a thick channel layer, DIBL increases quite dramatically as the gate length is scaled down to 50 nm, Fig. 4-4. In a comparison with the HEMTs processed on the materials in [F] without a back barrier, it can be seen that DIBL decreases from 125 mV/V to below 50 mV/V for a 100 nm gate length.

![Fig. 4-4. Comparison of DIBL between InAlGaN/AlN/GaN HEMTs with a BB [E], and InAlN/AlN/GaN HEMTs without BB [F]. DIBL was extracted for $\delta V_{th} = 116$ V.](image)

The design of a HEMT is often a trade-off between SCEs and electron trapping. These phenomena both affect the RF swing, but at the opposite extremes. As discussed earlier, electron trapping affects the knee voltage and knee current, while short-channel effects hinder the complete pinch-off, as shown in Fig. 4-5. It is shown in [E], that while effectively suppressing SCEs, a thin channel layer also introduced more electron trapping. This was reflected in the knee-walkout, in the increase in the on-resistance, and
Critical aspects of InAl(Ga)N HEMTs

in the slump ratio. The increased trapping was associated with a closer proximity of the 2DEG to the C-doped AlGaN back barrier. Similar observations of a trade-off between electron trapping and SCEs in the design of the channel layer thickness were reported in [191]. However, the scaling of the gate length was not considered. In [E], the gate length showed to be highly relevant for the maximum output power. A thicker channel layer requires a longer gate to maximize the output swing. A thin channel layer on the other hand reduces the output power in all HEMTs, regardless of the gate length. In Fig. 4-6, the measured maximum output power at 30 GHz for a gate length of 50 and 100 nm, respectively, is plotted. The two regions where the output power is limited by dispersion and SCEs, respectively, are marked out.

Fig. 4-5. The limiting impact of dispersion and short-channel effects on the large-signal RF output swing.

Fig. 4-6. Variation in the maximum $P_{out}$ of HEMTs in [E] with respect to channel thickness and gate length.

4.3 Passivation

In order to inactivate the surface traps and thereby avoid current collapse, a high-quality passivation layer is crucial. Numerous passivation schemes have been explored by different research groups, for instance SiO$_2$ [195], [196], SiON [197], BCB [198], MgO [199]. The use of a GaN cap, both intrinsic and p-doped, has also been reported to mitigate current collapse, [200], and may be combined with a surface passivation dielectric such as SiN$_x$ [201]. The improved performance has been explained by the screening of surface potential fluctuations by the cap layer, by a reduction of the peak electric field near the gate, and to some extent the increased distance between the surface states and the 2DEG [202], [203].

Despite many investigated options, SiN$_x$ remains the most extensively used surface passivation layer in GaN HEMT technology. It has proven to efficiently passivate surface traps [204],[124], and is a commonly available in most laboratories. The denotation “x” is used since the composition in these layers varies. SiN has an $\varepsilon_r$ of around 7 – 8 (corresponding to a refractive index of 2.1-2.2 at a wave length of 633 nm). The most commonly used deposition method is PECVD, but several other methods exist, e.g. reactive sputtering [205] and other CVD-based methods such as LPCVD [206], ICP-CVD [124], and catalytic CVD [207].

Besides the role of passivating the surface traps, there are other crucial aspects of a passivation layer. A low dielectric constant $\varepsilon_r$ is preferred to minimize the additional parasitic capacitance. In the special case when a gate dielectric is used (MISHEMT structure), it may however be desired to have a high $\varepsilon_r$ to promote a high $g_m$, as given by Eq. (13). A gate dielectric can be combined with a second layer for proper surface passivation [123]. It has been shown that the gate leakage, can increase after passivation. The topic was investigated in [208], where a dedicated test structure was used to identify the leakage path. It was demonstrated that the gate current was in fact flowing vertically through the barrier and not within the SiN$_x$ film itself.
Critical aspects of InAl(Ga)N HEMTs

The properties of a passivation layer, which have a major impact on the HEMT performance, can be altered by the tuning of process parameters such as temperature, pressure, plasma power, and gas flows [209]. Analysis of the film can be performed for example by ellipsometry to determine thickness and the refractive index. For SiN passivation layers, it has been shown that stress has a major impact on dispersion as well as gate leakage. In [210], it was observed that a less compressive stress resulted in a more efficient surface passivation. On the other hand Cho et al found that compressive strain mitigated the gate leakage [211]. Recent studies have shown that stress in the passivation layer has an even more profound impact on the HEMT performance, contributing to the piezoelectric polarization, with a direct impact on the 2DEG [212]. In [213], it was shown that the threshold voltage could be controlled by varying of mechanical stress in SiN films, locally affecting the 2DEG underneath the gate.

For InAl(Ga)N surfaces in particular, Al₂O₃ deposited by ALD [214][215] has emerged as an alternative to SiN. It has also been used on AlGaN surfaces, mainly as a gate insulator [214]. The εₛ of Al₂O₃ is slightly higher (~8-11) than for SiN. ALD is typically performed in the 200 – 350 °C range, which makes it suitable for a process in which the gate is defined before passivation. and produces very uniform layers and thickness control. Several subgroups of ALD exist such as thermal- [216] and plasma-assisted ALD. The methods differ in the precursor providing the O component, which is water vapor in the case of thermal ALD, and oxygen plasma in the case of plasma-assisted ALD. In addition to the evaluation of different passivation schemes, the sensitive surface of III-nitride heterostructures has motivated the investigation of different pre-treatments, such as wet chemical treatment by e.g. HCl and TMAH [217][218], or low-power N₂ plasma exposure [219]. Many positive results indicate that the surface can indeed be modified by pre-treatment [220], yet the issue is complex, and no general method exists.

The primary focus in this work has been the impact of different passivation layers on the InAl(Ga)N-based HEMT performance. The method of early passivation has been avoided to protect the InAl(Ga)N from potential plasma-induced damage that gate recess etching may cause. Thus, the passivation layer has been deposited after gate formation. This in turn imposes restrictions to the maximum temperature during deposition to prevent degradation of the Schottky contact, excluding methods like LPCVD.

Different passivation layers were evaluated in terms of their impact on the HEMT performance. The necessity of using a passivation layer was demonstrated in [C], in which an InAlN was characterized prior to- and after passivation with SiN. Deposition was performed by PECVD at 340 °C in a process developed at III-V Lab. A substantial improvement in the drain current and were observed in the IV characteristics, shown in Fig. 4-7. This is associated with a reduced Rₛ, according to Eq. (21), due to less trapped surface electrons that deplete the 2DEG. In the study the gate length was around 50 nm. SCEs became more pronounced after passivation in the absence of the virtual gate effect. The passivated HEMT exhibited a lower fₘₐₓ, which was mainly associated with a higher output conductance and parasitic capacitance. The reverse gate leakage increased. Yet, the necessity of surface passivation for large signal operation is confirmed by an increase in the output power of 240 % to 4.1 W/mm at 31 GHz.

![Fig. 4-8. HEMT performance before and after passivation with SiN by PECVD ([C]). a) IV characteristics and b) gₘ.](image-url)
The impact of the deposition method on the HEMT performance was investigated in [D], in which Al$_2$O$_3$ films were deposited on InAlN by thermal- and plasma-assisted ALD. The HEMTs were processed on a sample that was split in two pieces before the passivation step, and thereafter passivated separately with the two different methods. Both HEMT types exhibited negligible gate leakage, in contrast to the SiN$_x$-passivated HEMT in [C]. The HEMTs were subjected to pulsed-IV characterization, showing a considerable current slump for the HEMT passivated by thermal ALD, while the HEMT passivated by plasma-assisted ALD remained nearly unaffected at the applied quiescent bias voltage, Fig. 4-9a. The HEMT implemented with Al$_2$O$_3$ by plasma-assisted ALD exhibited 74 % higher output power (3.3 W/mm) at 3 GHz compared to the HEMT passivated by thermal ALD. This was reflected in the output waveforms, Fig. 4-9b, showing that plasma-assisted ALD minimized current collapse and maintained a high current level at microwave frequencies.

Fig. 4-9. Results corresponding to HEMTs passivated with Al$_2$O$_3$ by thermal- and plasma-assisted ALD, paper [D]. a) Pulsed IV characteristics for estimation of surface trapping an b) the output waveforms of the HEMTs in large-signal operation.

The reason for the large differences in terms of passivation was not investigated further. A slightly higher refractive index $\eta$ of the film deposited by plasma-assisted ALD (1.64 versus 1.63 for thermal ALD) indicates a denser film [221]. Another difference between the films is the incorporation of hydrogen, which is expected to be higher during thermal ALD. However, it has been discussed whether the presence of hydrogen may in fact promote the passivation of surface traps [220], [222]. A plausible explanation could also be that the plasma exposure itself is beneficial, which was supported in [223], in which plasma pre-treatment of the surface prior to passivation reduced dispersion. It is interesting to note that the same samples were also characterized in [a], in which the 1/f noise was investigated. The HEMTs passivated with thermal ALD exhibited better performance, which indicates that the 1/f noise is affected by different mechanisms.
Chapter 5 Thin Film Resistors

The first MMICs were fabricated in the 1970’s in GaAs MESFET technology [224]. Today, MMICs exist in a number of different technologies, e.g. GaAs pHEMT and MHEMT, SiC MESFET, InP HEMT, SiGe HBT, and InP HBT. The first AlGaN/GaN MMICs were developed in the early 2000’s [225]–[228]. GaN MMIC technologies primarily on SiC are offered by many commercial foundries due to the high thermal conductivity of SiC and relatively low lattice mismatch between GaN and SiC [229]–[232]. GaN on Si has however gained more ground over the last years the progress in growth technique that enables a high crystal quality despite the large lattice mismatch [233].

Circuits based on GaN HEMT technology are generally considered for power switching applications and for PAs. Increasing need for higher data rates motivates the development of transistors that operate at higher frequencies, which could benefit from larger bandwidth. The emerging fifth generation mobile communication network (5G) targets different frequency spectra, whereof some are in the mm-wave range, e.g. 37 – 40.5 GHz , 66 – 76 GHz [234], 130-175 GHz [235].

The number of results on GaN HEMT-based MMICs operating below the Ka band (< 40 GHz) are abundant. Advances in GaN HEMT technology have enabled power amplification also at W-band. HRL Laboratories have presented PAs delivering an output power density of 1.25 W/mm at 92 – 96 GHz with an associated PAE of 18 % [236]. Another PA, designed in a more downscaled AlGaN/GaN technology, exhibited an output power density of 3 W/mm in the 90 – 97 GHz range [237].

Circuits based on InAl(Ga)N technology are not as common. Marti et al presented W-band amplifiers based on InAlN/GaN-on-Si technology [184]. The peak power density was relatively low (0.7 W/mm), which they attributed to dispersion. The potential of InAl(GaN)N-based epi for PA applications was however shown in [238], demonstrating a coplanar PA with a power density of 3.6 W/mm with an associated PAE of 12 % at 86 GHz, in an InAlGaN/GaN MMIC technology. A microstrip process was utilized in [239], demonstrating several PAs operating at E-band. Among them, a two-stage PA exhibiting an output power of 1.63 W with an associated PAE of 15 %.

Active components in MMICs include transistors and diodes, while the passive components comprise resistors, inductors, and capacitors. In addition, MMIC circuitry contains transmission lines, and in the case of a microstrip circuit, also via-holes for proper grounding (via-holes are also often used in coplanar designs to suppress substrate modes). Microstrip designs are often preferred over coplanar (CPW) designs, since the latter easily become bulkier and sets a rather low limit to the level of integration. The diodes are implemented as lateral structures using the gate Schottky contact as anode and the source and drain contacts as cathodes [240].

As resistive elements, either semiconductor resistors or resistive films can be used. The semiconductor resistor exploits the sheet resistance of the heterostructure. It follows that the applied voltage must not exceed the point where current saturation is reached in order to avoid a non-linear IV behavior of the component. The semiconductor resistor exhibits a large temperature dependence, which is generally unwanted, but can be taken advantage of in some applications. For instance, the impact of different substrates on the thermal dissipation was monitored by semiconductor resistors in [19]. Thin film resistors (TFRs) are more common. An advantage is that $R_o$ can directly be controlled by the film thickness. Further, these components suffer less impact by the environmental conditions, whereas the semiconductor resistor can be affected by for instance light. Inductors are usually implemented as spiral inductors, while capacitors are generally metal-insulator-metal (MIM) capacitors.
Thin Film Resistors

At Chalmers, a GaN HEMT microstrip MMIC process was developed [75]. Former published results include an X-band mixer [241], an X-band receiver front end [8], and LNAs [242]. Two types of TFRs are available. A TaN TFR with a sheet resistance of 45 Ω/square, and a TiN TFR with a sheet resistance of 10 Ω/square, described in [G]. Spiral inductors and transmission lines are defined in the same layer, metallized with Ti/Au, and thereafter electroplated with a thicker Au layer for minimization of losses. MIM capacitors are formed as two metal plates with an intermediate 300 nm thick SiN layer deposited by PECVD. The thickness is set with the purpose of obtaining a high breakdown voltage (200 V) and the capacitance per unit area (CPUA) is 200 pF/mm². Via holes are etched from the backside using a thick Ni mask and metallized with sputtered Ti/Au and plated gold.

TFRs are used in MMICs for e.g. circuit stabilization, spanning a broad range of resistance values. A large variety of films have been proposed for application in TFRs, e.g. CrSi [243], CuNi [244], and ZrN [245], but by far the most commonly used TFRs are based on TaN [246] and NiCr [247]. TaN TFRs are often preferred over NiCr TFRs due to their chemical stability, whereas NiCr is known to be prone to oxidation and issues with compositional control, which will affect the resistance value [248], [249].

The primary parameter of a film intended for TFR applications is the sheet resistance R\text{sh}. The resistance value R of the TFR is set by the TFR dimensions, (Fig. 5-1) through

\[ R = R_{\text{sh}} \frac{l}{w} \]  

(25)

where l is the length and w is the width. With a suitable R\text{sh}, very bulky dimensions of the TFRs can be avoided. A suitable value of R\text{sh} depends on the desired resistance value R. For a high-resistance TFR, a large sheet resistance is required to maintain a reasonable l/w aspect ratio. Conversely, a low sheet resistance is required for a low-resistance TFR. It can therefore be advantageous to use more than one TFR type in an MMIC process. R\text{sh} can be adjusted by the film thickness t:

\[ R_{\text{sh}} = \frac{\rho}{t} \]  

(26)

where \( \rho \) is the resistivity of the material. However, the maximum film thickness is limited by mechanical stress that may cause cracks or adhesion problems.

Depending on deposition parameters, the density of the films may vary, which will have a direct impact on \( \rho \). A low pressure leads to fewer collisions of the ejected molecules with the plasma ions. Thereby the sputtered material has higher kinetic energy when reaching the substrate, which has shown to result in a denser film [250]. Other ways of increasing the density include increasing the power of the plasma [251] or to use substrate biasing [252]. On the other hand, a high density of the film increases the mechanical stress. The thickness uniformity is largely influenced by the chamber geometries, i.e. the target – substrate distance, the size of the target relative to the substrate, and the erosion zone of the target [253].

Regarding the fabrication of TFRs, there are a number of aspects to take into account. The most obvious ones are those that determine the properties of the film, but equally important is a process that can be implemented in fabrication process. The TFRs are preferably defined via a lift-off process. Although earlier works have proposed to fabricate TFRs by the deposition of a film followed by patterning and subsequent by dry etching [254], this is hardly a viable option in MMIC processing due to the sensitivity of the surface and structures to the dry etching plasma.
TaN films are generally prepared by reactive sputter deposition. For NiCr, evaporation exists as an alternative option. Evaporation offers an easy lift-off, while sputtering imposes higher requirements on the lithography profile. On the other hand, the composition of evaporated films is more difficult to control. For NiCr films, a composition of Ni:Cr 80:20% is often preferred due to its high resistivity and low thermal dependence [255]. However, because of preferential evaporation of Cr, the Cr content is often higher [256], with a risk for larger differences after repeated evaporations with the same source.

Sputtered films are generally polycrystalline and vary in the stoichiometry, structure, grain sizes, and thickness uniformity. A wide range of methods for the characterization of thin films are available. The composition of the film can be investigated with electron spectroscopy for chemical analysis (ESCA). Alternatively, the film can be analyzed by auger electron spectroscopy (AES) or Rutherford backscattering (RBS). X-ray diffraction can be used for identification of phases [257]. The surface morphology can be investigated by AFM.

5.1.1 Thermal dependence

For circuits operating in a wide ambient temperature span, thermal stability is of importance. TFRs used for circuit stabilization often have a resistance that cannot be much altered for the risk of a degradation in the circuit performance, or the appearance of circuit oscillations. Thin films exhibit a dependence on temperature, which can be described as

\[ \rho(T) = \rho_0 + \rho_1 T \]  

(27)

For a TFR the thermal dependence can be quantified by the thermal coefficient of resistance (TCR), given in ppm/ºC. The TCR is calculated as

\[ TCR = \frac{1}{R_{low}} \frac{R_{high} - R_{low}}{T_{high} - T_{low}} \]  

(28)

where \( R_{low} \) and \( R_{high} \) are the resistances at a lower temperature (\( T_{low} \)) and higher temperature (\( T_{high} \)), respectively. Hence, a positive TCR means that the resistance increases with temperature and a negative TCR means a decrease of resistance vs temp. TiN_xO_y films have exhibited an increase in TCR as the temperature is increased [258], whereas TaN and SiCr films have shown a negative dependence [259],[260],[243]. In [G], TiN TFRs exhibited a positive TCR. For NiCr films, the TCR has been observed to vary with the composition, exhibiting both positive and negative values [261]. In the given references, the magnitude of the TCR varied from 5000 to near-zero, but commonly the TCR is in the range of a few tens to a couple of hundreds of ppm/ºC.

Several studies have investigated the thermal dependence of different thin films and the correlation to different process parameters. Mireles et al observed a large impact on the TCR in TiSiON films by the composition, in particular the N content [262]. Also for TaN films it has been shown that an increase in the N partial pressure, which lead to an increased N content, led to a higher magnitude of the (negative) TCR [263]. Birkett et al correlated a low TCR of CuAlMo TFRs to a low pressure during sputtering and obtained a near-zero TCR with additional annealing [264]. Although no general method for the minimization of the TCR exist, thermal annealing in a temperature range of 250 – 600 ºC has frequently been used [265], [266], [267]. It may however also increase the sheet resistance [268]. The doping of CrSi films with Mo and Ni combined with annealing at 550 ºC resulted in a very low TCR of around 10 ppm/ºC [269]. If thermal annealing of the TFRs is performed in an MMIC process, it is important that the high temperature exposure does not have an impact on the remaining circuitry.

5.1.2 TiN TFRs

In this work, a TiN TFR process was developed in [G], starting with the investigation of the N_2/Ar flow ratio and the composition. The Ar flow was kept constant while the N_2 flow was incremented. The bias voltage of the sputter system increased initially and stabilized at an N_2/Ar ratio of around 6-7%. Fig. 5-2a, indicating nitridation of the target. Films were sputtered on Si/SiO_2 substrate at selected N_2/Ar
Thin Film Resistors

flow ratios. ESCA characterization of the sputtered films (Fig. 5-2b) showed that the composition remained reasonably constant from a ratio of 5%. At this point the film consisted of 50% Ti, 40% N, and an additional and unwanted O component of 10%. The O component increased at higher $N_2$ flows, indicating that it is related to the purity of the gas supply. The higher concentration leads to a higher $\rho$, Fig. 5-2c.

TiN TFRs with thickness of 3560 Å exhibited in a $R_{sh}$ of 10 Ω/square. These were found to be a suitable complement to the TaN TFR in the Chalmers MMIC process, and was therefore implemented as a low-resistivity option.

5.1.3 Temperature compensated TFRs

The TiN films exhibit a positive TCR of 540 ppm/ºC, while the in-house TaN TFRs exhibit a TCR of -470 ppm/ºC. The different signs of the TCRs of TiN and TaN was utilized in [H] to fabricate TFRs with a low TCR. It was shown that the TCR could be significantly reduced by processing TiN and TaN TFRs connected in series. The design formula of the series configuration using two films with different TCRs and a fixed thickness is:

$$\frac{t_{TiN} w_{TiN}}{t_{TaN} w_{TaN}} = \frac{\rho_{TiN}}{\rho_{1, TaN}}$$  \hspace{1cm} (29)

If a certain TFR resistance $R_s$ is targeted, the thicknesses $t_{TiN}$ and $t_{TaN}$ are fixed and $w$ is set equal for the films, and $l_{TiN}$ and $l_{TaN}$ are related through Eq. (28), then the required length of the TaN TFR becomes

$$l_{TaN} = R_s w_{TaN} / (\rho_{0, TaN} - \frac{\rho_{0, TiN} \rho_{1, TaN}}{\rho_{1, TiN}})$$  \hspace{1cm} (30)

A more convenient solution was demonstrated by the fabrication of a single, double layer TFR, which is equivalent to a shunt connection. A mitigated TCR was anticipated by Taylor expanding the expression for the conductance of the shunted TFRs around a temperature $T_x$ and choosing the TFR dimensions so that the thermal dependence is cancelled. This resulted in the design formula:

$$\frac{t_{TiN}}{t_{TaN}} = \frac{\rho_{1, TaN}(\rho_{0, TiN} + \rho_{1, TiN} T_x)}{\rho_{1, TiN}(\rho_{0, TaN} + \rho_{1, TaN} T_x)T_x}^2$$  \hspace{1cm} (31)

The targeted resistance $R_p$ can be set by the dimensions $l$ and $w$ through

$$R_p = \frac{l}{w} \left( \frac{t_{TiN}}{\rho_{0, TiN} + \rho_{1, TiN} T_x} + \frac{t_{TaN}}{\rho_{0, TaN} + \rho_{1, TaN} T_x} \right)$$  \hspace{1cm} (32)

The same multilayer approach has been used with NiCr capped by NiCrSi, which exhibit opposite signs of the TCR, resulting in a minimum TCR of 35 ppm/ºC [270]. The two configurations are shown in Fig. 5-3a.
5.1.4 TFR reliability

The TFR reliability was investigated by exposing the TFRs to electrical stress. By sweeping the current through the TFR, a critical power $P_c$ could be identified, at which the resistance started to increase rapidly, Fig. 5-4a, and eventually leading to a catastrophic failure. The investigated TFRs featured a variety of length – width combinations. A correlation of $P_c$ to the footprint area of the TFR was observed, shown in Fig. 5-4b. As an example, the measured $P_c$ coincides for two TFRs with a large difference in the resistance value, but with nearly equal footprint area. It was therefore concluded that the thermal dissipation through the substrate must be of major importance for the device reliability, and self-heating was believed to be the dominating failure mechanism. Thermal imaging performed with an IR camera during the current sweeps showed that the temperature within the TFR could rise to over 200 °C. It could also be seen that the TaN TFR with a negative TCR exhibited a higher $P_c$ compared to the TiN TFRs for the same footprint area.

![Fig. 5-4. a) Definition of the critical power of a TFR ($P_c$) and b) dependence of $P_c$ on the footprint area.](image-url)
Chapter 6 Conclusion and future outlook

Several advanced concepts in III-nitride HEMTs and MMIC have been investigated to promote power amplification and generation in the millimeterwave frequency band.

InAl(Ga)N was explored as an alternative to the AlGaN barrier for GaN HEMTs due to its larger polarization, which facilitates downsizing with maintained electron density. The utilized heterostructures were grown by MOCVD, and had a total (InAl(Ga)N+AlN) barrier layer thickness in the range from 7 to 9 nm. An \( f_{\text{max}} \) of around 200 GHz was reached for HEMTs with a nominal gate length of 50 nm. Throughout the work, the impact of short-channel effects has been observed in the HEMTs, despite a very downscaled barrier. This imposes limitations to further downsizing, which is necessary to reach higher frequencies.

To mitigate short-channel effects, HEMT structures with a back barrier have been investigated, and the thickness of the intermediate GaN channel layer optimized. The use of a back barrier reduced the short-channel effects, with a correlation to the channel layer thickness. For HEMTs with a nominal gate length of 50 nm, the drain induced barrier lowering increases from 40 mV/V to 93 mV/mm as the channel layer thickness is increased from 50 to 150 nm. However, the proximity of the 2DEG to the back barrier for thinner channels increased electron trapping. Hence, the GaN channel layer thickness is a trade-off between short-channel effects and dispersion. Large-signal measurements at 30 GHz showed a maximum output power, 5.3 W/mm, and an associated PAE of 32 % for a HEMT with a gate length of 50 nm on the heterostructure with a thickness of 100 nm. For the 150 nm channel layer, similar output power and efficiency (5.0 W/mm and 29%) was measured, but for a HEMT with a longer gate length (100 nm), as a result of short-channel effects. The study highlights one trade-off in the design of a HEMT and its performance in terms of output power and short channel effect.

The enhancement in electron mobility in a heterostructure with a thin GaN layer sandwiched between the InAlN and AlN layers of an InAlN/AlN/GaN HEMT structure was evaluated. At room temperature the mobility was 1600 and 1930 cm²/Vs in the InAlN/AlN/GaN and InAlN/GaN/AlN/GaN HEMT structures, respectively. However, the increased mobility came at the cost of decreased carrier density, which counteracted an improvement in HEMT performance. The increase in mobility was further enhanced at cryogenic temperature (5K), where the mobilities were 6900 cm²/Vs and 16000 cm²/Vs. Temperature dependent s-parameter measurements showed an improvement in \( f_{\text{max}} \) from 128 GHz to 153 GHz at 5 K. The difference was mainly attributed to the reduction in the extrinsic access resistance. The higher mobility obtained when incorporating a GaN interlayer could be useful in low-noise-applications. The study provides useful information for the understanding of scattering mechanisms in InAlN-based heterostructures. The investigations suggest that the improvement in \( \mu \) is either due to the suppression of fluctuations in the quantum well subband energies, or to reduced Coulomb scattering, both related to compositional variations in the InAlN.

Low-resistive ohmic contacts based on Ta have been investigated on AlGaN-, and on InAlN-based heterostructures. The contacts have been optimized in terms of metal scheme, recess depth, and anneal temperature. The main advantage over Ti-based ohmic contacts is the lower anneal temperature of around 550 °C, which produces a smooth morphology and facilitates a more versatile process flow compared to Ti-based contact that are annealed at higher temperatures.

The effect of different passivation layers has been studied. The results underline the considerable impact of the passivation layer on several critical parameters of the HEMT performance. By measuring
Conclusion and future outlook

The collective results in this work have provided a rough indication of the limitations and status of InAlN HEMT technology today. Due to the challenging epitaxy, InAlN-based heterostructures may still not have reached its full potential. To improve the high frequency performance and large-signal power generation, a lot could be gained with a more homogenous InAlN layer. This would lead to a higher mobility and presumably a higher breakdown voltage, improved reliability, and a larger maximum output power density. Continued close collaboration with epi-suppliers are therefore of essence to facilitate a combined effort at both material and device level.

With the existing InAl(Ga)N layers of today, the tuning of parameters such as the barrier layer thickness and the gate length would likely only lead to minor improvements. To accomplish power amplification at higher frequencies a different approach is required. The growth of a more highly polarized barrier layer would permit further downscaling while maintaining a large $n_s$. It would therefore be interesting to increase the Al content of the InAlN. The issue of short-channel effects would remain and motivate the use of a back barrier. However, to take full advantage of a back barrier, it is desired to decrease electron trapping, for instance by optimizing the C doping level.

A possible way forward in the could perhaps be offered by the growth of N-polar heterostructures, in which the channel layer is grown on top of the barrier layer, reducing the distance between the gate and the 2DEG. HEMTs based fabricated on N-polar heterostructures have shown very promising results [271]–[273]. Further proposed innovative solutions include the fabrication of so-called Fin-HEMT structures, which provide improved gate control. A high $g_m$ (645 mS/mm) combined with a low DIBL (28 mV/V) has been reported for FiN-FETs based on InAlN [274].

Regarding surface passivation, the reasons behind the very different impact of different films on the HEMT performance are still open for further investigation. To start with, the test structures in [275] could be used to investigate whether the increased gate leakage in [C] after passivation was due to conduction within the SiNx layer, or within the heterostructure. Mechanical stress in passivation layers has shown to have a large impact on the heterostructure, locally affecting the 2DEG [213]. It would therefore be interesting to compare the stress in the two Al$_2$O$_3$ films and the SiNx film.

For ohmic contacts, the issue of repeatability is still not fully resolved. It is necessary to increase the understanding of the physical impact of parameters like the sidewall angle effect on the contact resistance, and moreover how to optimize the recess etching to achieve a certain geometry of the trench.
Summary of appended papers

In the following text a short description of each paper appended to the thesis is provided, with a statement of my own contribution.

**Paper A**


Ta-based ohmic contacts for GaN HEMTs are developed. Two metallization schemes were considered: Ta/Al/Ni(Ta)/Au and Ta/Al/Ta. The latter was superior in terms of lower contact resistance ($R_c$) and wider process window. The lowest measured $R_c$ was 0.06 $\Omega\cdot$mm. The main advantage of the Ta-based ohmic contacts over conventional Ti-based contacts was the low anneal temperature of 550–575 °C. The contacts exhibit excellent surface morphology and edge acuity, which facilitates lateral scaling of the GaN HEMT. TEM images were used to investigate the microstructure and the contact mechanism. Storage tests at 300 °C for more than 400 h in air ambient showed no deterioration of $R_c$.

*My contribution:* I designed the experiments, fabricated the test-structures, performed the measurements, and wrote the paper with feedback from the co-authors.

**Paper B**


The study investigates the formation of recessed, Au-free ohmic contacts to an InAlN/AlN/GaN heterostructure. A Ta/Al/Ta metal stack is used to produce contacts with contact resistance ($R_c$) as low as 0.14 $\Omega$mm, which was found for an etch depth almost reaching the 2DEG. A reduced sensitivity to other process parameters such as anneal temperature was found at the optimum etch depth. For deeper recesses, $R_c$ remains low but required annealing at higher temperatures for contact formation. An optimum bottom Ta layer thickness of 5–10 nm is found. Two reliability experiments preliminary confirm the stability of the recessed contacts.

*My contribution:* I co-designed the experiments, participated in the measurements and the analysis, and co-authored the paper with JB and with feedback from the co-authors.
Summary of appended papers

Paper C

The paper presents an InAlN/AIN/GaN HEMT with Au-free Ta-based ohmic contacts and a high-quality PECVD SiN passivation. The ohmic contacts were annealed at 550 °C, resulting in a contact resistance of 0.64 Ω·mm. HEMTs with a gate length of 50 nm were evaluated by performing DC-, pulsed IV-, RF-, and load-pull measurements. It was observed that current slump was mitigated by the passivation layer. The DC channel current density increased by 71 % to 1170 mA/mm at the knee of the IV curve, and the transconductance increased from 382 to 477 mS/mm after passivation. At the same time the gate leakage increased, and the extrinsic $f_{\text{max}}$ decreased from 207 to 140 GHz. Output power levels of 4.1 and 3.5 W/mm were measured at 31 and 40 GHz, respectively.

My contribution: I fabricated the HEMTs, performed the measurements, and wrote the paper with feedback from the co-authors.

Paper D

This paper investigates thermal and plasma-assisted atomic layer deposition (ALD) of Al2O3 films as passivation for InAlN/AIN/GaN HEMTs. A conventional plasma enhanced chemical vapor deposition deposited SiNx passivation was used for comparison. The difference in sheet charge density, threshold voltage, $f_T$ and $f_{\text{max}}$ was moderate for the three samples. The gate leakage current differed by several orders of magnitude, in favor of Al2O3 passivation, regardless of the deposition method. Severe current slump was measured for the HEMT passivated by thermal ALD, whereas near-dispersion free operation was observed for the HEMT passivated by plasma-assisted ALD. This had a direct impact on the microwave output power, where passivation with plasma-assisted ALD exhibit 77% higher output power at 3 GHz compared to passivation with thermal ALD.

My contribution: I fabricated the HEMTs, performed the measurements, and wrote the paper with feedback from the co-authors.

Paper E

The impact of varying the GaN channel layer thickness ($t_{\text{ch}}$) in InAlGaN/AIN/GaN HEMTs with C-doped AlGaN back barriers is investigated. $t_{\text{ch}}$ was 50, 100, and 150 nm, and the gate length of the fabricated HEMTs ranged from 50 to 200 nm. It is found that
Summary of appended papers

short-channel effects (SCEs) are significantly mitigated with a small $t_{ch}$. For HEMTs with a gate length of 50 nm, the drain-induced barrier lowering changes from 40 to 93 mV/V as $t_{ch}$ is increased from 50 to 150 nm. On the other hand, it is shown that dispersive effects are more severe for a smaller $t_{ch}$, as demonstrated by a six-fold increase in the dynamic on-resistance for $t_{ch} = 50$ nm compared to $t_{ch} = 150$ nm. The trade-off between dispersion and SCEs is reflected in large-signal measurements at 30 GHz. The 50-nm channel, mainly limited by dispersion, exhibits an output power of 3.5 W/mm. The thicker channels reach a maximum of around 5 W/mm, but for different gate lengths due to the difference in severity of the SCEs. This paper elucidates the interplay between SCEs and dispersion related to $t_{ch}$, its consequences for the large-signal performance and the limitation in downscaling of the gate length.

My contribution: I co-designed the experiments, fabricated the HEMTs, performed the measurements, analyzed the results, and wrote the paper with feedback from the co-authors.

Paper F


A mobility ($\mu$) enhancement in InAlN/AlN/GaN heterostructures is demonstrated by the incorporation of a thin GaN interlayer (IL) between the InAlN and AlN. The introduction of an IL increases $\mu$ at room temperature (RT) from 1600 (for a reference InAlN/AlN/GaN structure) to 1930 cm$^2$/Vs. The difference is larger at cryogenic temperature (5 K), where the GaN IL sample exhibits a $\mu$ of 16000 cm$^2$/Vs, compared to 6900 cm$^2$/Vs for the reference sample. The results indicate the reduction of one or more scattering mechanisms. We propose that the improvement in $\mu$ is either due to reduced Coulomb scattering, or to the suppression of fluctuations in the quantum well subband energies, both related to the compositional variations in the material composition of the InAlN-barrier. The HEMTs on the GaN IL sample demonstrate larger improvement in dc- and high frequency performance at 5K. Compared to RT, $f_{\text{max}}$ of the GaN IL sample increases by 51 GHz, while the reference exhibits an increase of 35 GHz. The difference in improvement between the samples was associated mainly with the drop in the source resistance. Our study also underlines the importance of further growth optimization of the InAlN barriers.

My contribution: I fabricated the HEMTs, performed the measurements, analyzed the results, and wrote the paper with feedback from the co-authors.

Paper G


The paper presents the fabrication of titanium nitride (TiN) thin film resistors (TFRs) by reactive sputter deposition. The TFRs were characterized in terms of composition, thickness, and resistance. Furthermore, a first assessment of the resistor reliability was made by measurements of the resistivity ($\rho$) versus temperature, electrical stress, long-term stability, and thermal infrared measurements. TiN layers with thicknesses up to 3560 Å, corresponding to a sheet resistance of 10/□, were successfully deposited without any signs of stress in the
The critical dissipated power showed a correlation with the resistor footprint-area, indicating that self-heating was the main cause of failure. This was partly substantiated by the thermal infrared measurements.

**My contribution:** I fabricated the TFRs, performed the measurements, and wrote the paper with feedback from the co-authors.

**Paper H**


The opposite signs of the temperature coefficient of resistance (TCR) of two thin film materials, titanium nitride (TiN) and tantalum nitride (TaN), were used to form temperature compensated thin film resistors (TFRs). The principle of designing temperature compensated TFRs by connecting TFRs of each compound in series or in parallel was demonstrated. TiN, TaN, and combined TiN and TaN TFRs for monolithic microwave integrated circuits (MMICs) were fabricated by reactive sputtering. DC characterization was performed over the temperature range of 30 – 200 °C. The TiN TFRs exhibited an increase in resistivity with temperature with TCRs of 540 and 750 ppm/°C. The TaN TFR on the other hand exhibited a negative TCR of −470 ppm/°C. The shunted TFRs were fabricated by serial deposition of TiN and TaN to form a bilayer component. The TCRs of the series- and shunt configurations were experimentally reduced to −60 and 100 ppm/°C, respectively. The concept of temperature compensation was used to build a Wheatstone bridge with an application in on-chip temperature sensing.

**My contribution:** I fabricated the TFRs, performed the measurements, and wrote the paper with feedback from the co-authors.
Acknowledgement

Finally finding myself at the end of this demanding journey, I would like to express my deepest gratitude to a number of people who have helped me reach the end of the tunnel. First of all, I would like to thank my examiner, Prof. Herbert Zirath, for giving me the opportunity to pursue this work. I cannot enough thank my main supervisor Prof. Niklas Rorsman for his consistent support throughout the entire period, for his engagement in my work, for always taking a minute to discuss a topic despite a tight schedule, and on top of that enrich his colleagues with a great sense of humor. I also deeply thank my co-supervisor Dr. Mattias Thorsell for his assistance with different measurements, for valuable input on characterization and other topics, and for scrutinizing my written words. A special thanks to Dr. Hans Hjelmgren for all the physics-discussions, and for teaching me TCAD simulations.

Further, I am grateful to Dr. Jr-Tai Chen and Olof Kordina at SweGaN AB, to Dr. Piero Gamarra, Prof. Sylvain Delage, and Prof. Marie-Antoinette Di Forte Poisson at III-V Lab, and to Dr. Hervé Blanck at United Monolithic Semiconductors (UMS) GmbH for our collaboration over the years.

I will also take the chance to thank the clean room staff, in particular Mats Hagberg, Johan Andersson, Niklas Lindvall, and Henrik Frederiksen, for helping me solve the many process-related issues that tend to arise in the lab. In addition, Dr Alexei Kalaboukhov and Gustav Sundell are thanked for their assistance with Hall characterization and FIB-SEM.

I am surrounded by great colleagues, many of whom have helped me in one way or another over the years, Dan, Christian, and many more. Thank you all for your assistance, and for contributing to a good work environment. Special thanks to Ding for being a superb office mate. And lastly, thank you my husband Mikael for always supporting me and for making me happy.

This work was supported by:


“Advanced III-Nitrides-based electronics for future microwave communication and sensing systems” (RE10-0067) financed by the Swedish Foundation for Strategic Research (SSF).

“Manufacturable GaN” (B-0924-IAP1-ERG) and “European Gallium Nitride Industry Supply Chain” (B-1447-IAP1-GP) financed by the European Defence Agency (EDA) funded by Germany, Great Britain, France, Italy, and Sweden, through the EDA Project n° B-1447-IAP1-GP.
Acknowledgement

“Swedish Manga” financed by the Swedish Defence Materiel Administration (FMV).


“Bridging the THz-gap” financed by Knut and Alice Wallenberg Foundation (KAW)

“Ultra-Compact AESA Technology for Autonomous Aircrafts” (2017-04870) with financial support from the VINNOVA and Saab AB
References


[33] https://www.cree.com/

[34] https://www.infineon.com/

[35] https://www.nxp.com/


References


[57] G. Dambrine, A. Cappy, F. Heliodore, and E. Playez, “A New Method for Determining the FET
References


References


References


[123] B. P. Downey, D. J. Meyer, D. S. Katzer, J. A. Roussos, Ming Pan, and Xiang Gao,


References


[152] F. M. Mohammed, L. Wang, D. Selvanathan, H. Hu, and I. Adesida, “Ohmic contact formation mechanism of Ta/Al/Mo/Au and Ti/Al/Mo/Au metallizations on AlGaN/GaN HEMTs,” J. Vac.
References


References


References


[229] https://www.qorvo.com/innovation/technology/gan

[230] https://www.ums-gaas.com/foundry/

[231] www.winfoundry.com

[232] https://www.wolfspeed.com/rf/foundry-services


References


References


References


