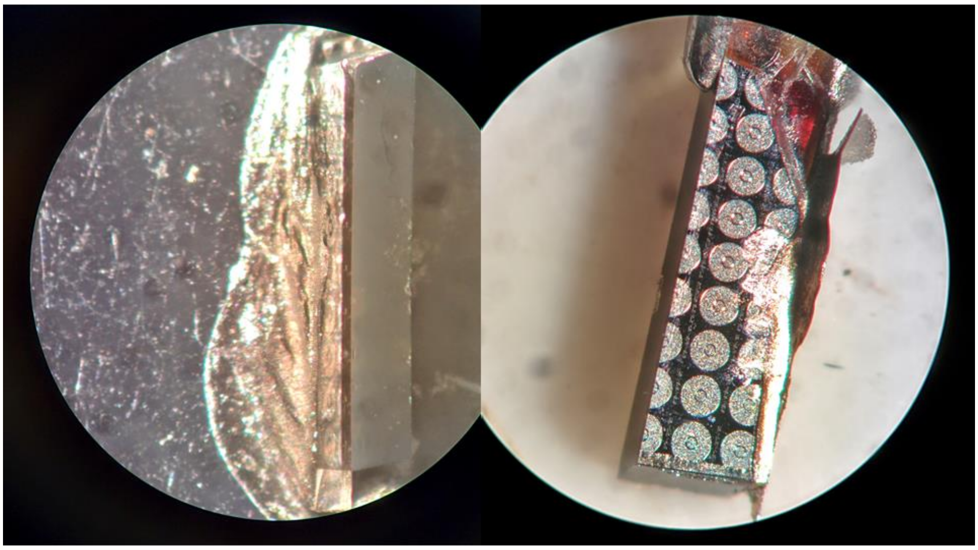
MASTER THESIS

IMAGING OF GAN-ALGAN HIGH ELECTRON MOBILITY TRANSISTORS

AUTHOR:	*GRADUATION COMMITTEE:*
Kay Polders	Prof. dr. ir. Harold J.W. Zandvliet
	Prof. dr. Dirk J. Gravesteijn
	Dr. ir. Herbert Wormeester
	Dr. Marc M.J. Dhallé



UNIVERSITEIT TWENTE.
Physics of Interfaces and
Nanomaterials Group
& NXP Semiconductors

July 31st 2014

Abstract

Presented here is the research done in course of a Master's assignment at the Physics of Interfaces and Nanomaterials (PIN) group at the University of Twente, commissioned by NXP Semiconductors. Research is done into the carrier transport mechanism of gallium nitride – aluminium gallium nitride high electron mobility transistors, using titanium alloyed contacts, designed and produced by NXP. Three hypotheses concerning the transport mechanism are put forward, all concerning the formation of titanium nitride. A cross-section of these transistors is fabricated and measured by way of scanning tunnelling microscopy, atomic force microscopy, and current-sensing atomic force microscopy. Additional measurements, done during fabrication and in between other measurements, by way of scanning electron microscopy and helium ion microscopy, showing unknown and as of yet unexplained electric behaviour in the structure, are discussed as well.

Determination of the transport mechanism proved unsuccessful as the cross-section was fabricated in atmosphere, causing insurmountable oxide build up on the measured surface, which in turn makes definitively determining the electronic structure of the transistor difficult. However, in investigating this result, interesting electronic properties of the stack in question, as well as the substrate used, are found.

Recommendations are done on how to repeat the presented measurements with samples created in situ, preventing oxide build-up, as well as recommendations concerning the fabrication of said samples.

Uittreksel

In dit schrijven wordt verslag gedaan over het onderzoek, gedaan in de loop van een afstudeeropdracht bij de Fysica van Grensvlakken en Nanomaterialen (PIN) groep aan de Universiteit Twente, in opdracht van NXP Semiconductors. Er is onderzoek gedaan naar het ladingsdragertransportmechanisme van galliumnitride-alluminiumgalliumnitride hoge-elektronmobiliteittransistoren, met contacten van titaniumlegeringen, ontworpen en geproduceerd door NXP. Drie hypotheses aangaande dit transportmechanisme worden voorgelegd, welke allemaal gerelateerd zijn aan de formatie van titaniumnitride. Een dwarsdoorsnede van deze transistoren wordt gefabriceerd en gemeten door middel van rastertunnelstroommicroscopie, atoomkrachtmicroscopie, en stroomgevoelige atoomkrachtmicroscopie. Aanvullende metingen, gedaan tijdens fabricage of tussen andere metingen door, door middel van scannende-elektronbundelmicroscopie en heliumionenmicroscopie, welke onbekend en vooralsnog onverklaard elektronisch gedrag in de structuur laat zien, worden ook besproken.

Het bepalen van het transportmechanisme is onsuccesvol gebleken, gezien de dwarsdoorsnede gefabriceerd werd bij atmosferische druk, hetgeen een onoverkomelijke oxidevorming veroorzaakt, wat het zeer moeilijk maakt onherroepelijk vast te stellen hoe de elektronische structuur van de transistor eruit ziet. Aanbevelingen worden gedaan over hoe deze metingen over te doen met in situ gefabriceerde monsters, waardoor oxide opbouw voorkomen wordt, evenals aanbevelingen aangaande het fabricageproces van voorgenoemde monsters.

Table of Contents

Introduction

In today’s industrial age progress moves ever faster: It took humanity some 300 years to get from the first external combustion engine to its internal counterpart; 90 years to get from the Gramophone record to the laser disc; 40 years from man’s first powered flight to the first jet aircraft; less than 12 years from the first artificial satellite in space to the first man on the moon. This trend of speed is well known and recognized by industries worldwide, partly self-fulfilling its forecasts of ever faster progress, but for none more is this true than for the semiconductor industry. From the early days of the first transistor to the first integrated circuit took only eleven years, from then on the amount of circuits on a chip would double, whilst cutting energy usage, heat production and cost in half, every 18 to 24 months.

In order to keep progress going, the industry has focused research and development largely on a single material: Silicon. As one of the most abundant elements on earth, it is very cheap and, given the many years of research, the fabrication, extrusion and purifying are very well understood. However, not all modern applications of semiconductors are possible, or economically viable[1] when using just silicon.

Silicon is a group IV elemental semiconductor, and has a small bandgap of 1 eV. This means it can easily become conducting when a large enough potential is applied. The electrical breakdown field of silicon, then, is only 0.3 MV/cm, meaning it is not suitable for high power or high voltage applications. Furthermore the silicon-bandgap is indirect, meaning electrons need momentum to overcome the bandgap, as opposed to direct-bandgap semiconductors in which the energy is used to produce a photon. The latter type is needed for optoelectronic applications, such as electroluminescence, where electrical energy is converted to light using a light emitting diode (LED), which obviously is a fast growing industry.

---

[1] From a physical point of view: some applications are not scalable to a small enough size when using silicon, due to the small bandgap of the material, to have enough of them on a microchip in order for it to become economically viable.

For applications where high power, voltage, or a direct bandgap are required, silicon no longer is usable. Thus group III-IV compound semiconductors are used. These semiconductors are comprised by two or more elements, which makes them much more tuneable; many different combinations are possible. These materials can, for instance, be combined to create direct-bandgap semiconductors with bandgap energies exceeding that of silicon. One of these materials is galliumnitride (GaN).

GaN is a direct-bandgap semiconductor with a high bandgap of 3,39 eV, making the electrical breakdown field for this material 3,5 MV/cm. This makes GaN much more suitable for applications where high electric fields are applied, and, because of the direct bandgap, the material can be used to create LED's with a wavelength of 405 nm [1].

For the scope of this report, another application of GaN is investigated: using this semiconductor in a GaN-AlGaN heterojunction yields a sophisticated High Electron Mobility Transistor (HEMT), which has a much lower channel resistivity than more regularly used MOSFET's. At this moment, however, far less is known about the precise workings of such a device than there is for its silicon counterparts. Specifically, very little is known about the electronic transport between the metal source and drain to the two dimensional electron gas on the GaN-AlGaN interface. Many different hypotheses have been put forward [2], but none have yet been proven true or false.

Thus a closer look is needed to the interface in the contact region of these transistors, and it is this that the research for this masters assignment set out to achieve. A cross-section of a working HEMT device of GaN on silicon was made, and attempts have been made to investigate the electronic transport mechanism.

In this report, an outline of the basic workings of the HEMT will be given, as well as three distinct hypotheses for the transport mechanism. Various measurement techniques necessary for the investigation into how electronic transport takes place are discussed in theory, as well as a brief explanation of the workings of several real-life devices used in the research. Furthermore, a brief outline of obtaining a good cross-section is made. Lastly, results will be discussed in order to draw conclusions concerning the transport hypotheses and recommendations are given to improve results in the future.

Chapter 1.

Background

In this chapter a theoretical basis will be given for the research discussed in the rest of the report. A basic understanding of semiconductor physics and materials physics is required. For the reader that is not skilled in these subjects, a brief explanation of the need-to-know theories and terms is given in Appendix A.

1.1 GaN – AlGaN HEMT

Although widely used, the MOSFET has a few disadvantages. The most important one, for this thesis, is the reduced mobility of the electrons in the inversion layer. As a semiconductor is doped with impurity atoms, collisions between electrons and these atoms are likely, reducing mobility of charges. A solution to this can be found by getting the charges from another physical event, occurring at an interface between two different semiconductors. A High Electron Mobility Transistor, or HEMT, is a form of transistor that makes use of a heterostructure of two semiconducting materials. In this paragraph, the working of such a system will be explained [3].

1.1.1 HEMT's

Heterojunctions in semiconductor physics are interfaces between two semiconducting materials with different bandgaps. Because E_F must be at thermal equilibrum across the interface[2] despite having different values in each material, the conduction and valence bands near the interface are bent in order to equalize the Fermi level. This would result in a band structure across the interface, akin to figure 1.

---

[2] Compare this with two reservoirs of water at different heights, with a gate in between. If one were to open the gate the water from the highest reservoir would run to the lowest until the level would be equal across both reservoirs.

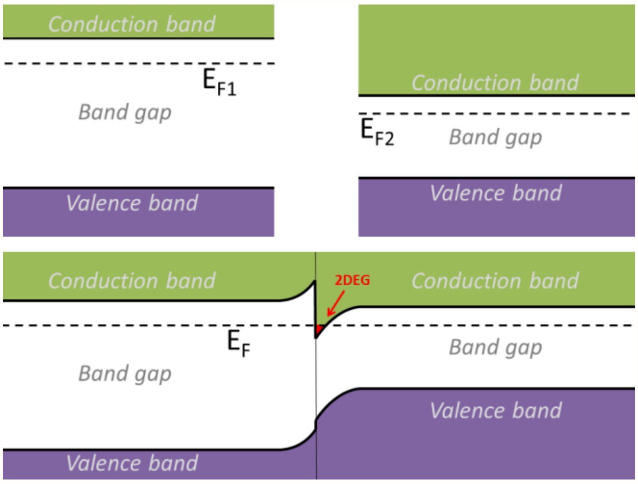


Figure 1. Band bending caused by stacking two different semiconductor materials together.

The result is similar to the electronic structure of a MOSFET, and one could indeed make a transistor by stacking two semiconductors on top of each other as in figure 2. Again by applying a voltage to the gate an electric field across the junction will raise E_F to the point where the conduction band near the interface drops below the Fermi level.

The Heterojunction Field Effect Transistor (HFET), contrary to the MOSFET, gets its negative charge carriers from undoped semiconductors. In the GaN-AlGaN system, used in this research, the electrons come from the depletion of the AlGaN into the (undoped) GaN conduction band, where their mobility is not reduced by doping atoms like in a MOS system [4].

Furthermore the bending in this heterostructure creates a much more narrow well of charge carriers than in the MOS system, causing the electrons to be much closer to the interface. The “canyon” in the electronic landscape is so sharply defined, in fact, that one can speak of a quantum well, with usually only one or a couple of specific energy states available. The result is that electrons can quickly fill the quantum well, where they behave like an ideal gas, moving along the interface without many collisions. Since the gas is confined to the interface this is called a Two Dimensional Electron Gas (2DEG), which gives the channel a very high charge density and charge mobility (meaning a low resistance). Hence its other name: High Electron Mobility Transistor (HEMT).

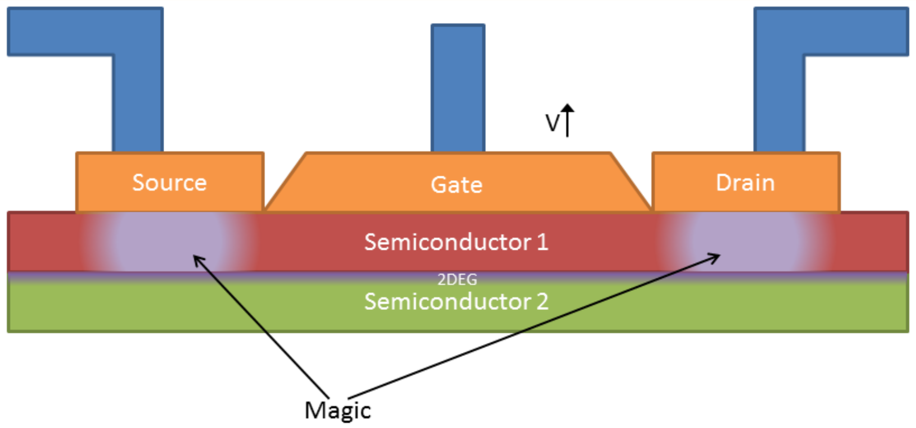


Figure 2. A schematic representation of an HFET or HEMT. The transport from the source and drain to the 2DEG is shown as having no definitive scientific explanation (yet).

1.2 Transport hypotheses

The use of an HFET instead of a MOSFET eliminates the necessity to dope the semiconductor channel, since plenty of charge carriers are available in the 2DEG. However, very little is known about the transport mechanism that allows a current to flow from the source of the transistor, into the 2DEG and out to the drain again. Contrary to MOSFET technology there isn't a 'straightforward' interface between a positively doped bulk and a negatively doped source and drain. Instead current has to flow to and from the terminals through the (insulating) AlGaN layer, into the 2DEG and out on the other side. The mechanism by which electrons cross the insulating layer is not trivial (as is indicated in figure 2).

To reduce the difference in barrier height between the metal and the semiconductors, preference is given to the transition metals with few electrons in their outer shell, but the crystal structure needs to be taken into account as well. Empirical research[3] has shown that the use of titanium and aluminium provides good results for deposition on AlGaN [2]. For the course of this research, titanium, alloyed with aluminium, receives specific attention. The alloys are sputtered, in layers of alternating titanium and aluminium, and annealed. This provides a good Ohmic transport from the outside world to the 2DEG. What the mechanism behind the success of this method is, however, is unclear and the main research question for this report. Three hypotheses

---

[3] Trial and error.

are put forward, all of which have to do with the formation of titanium nitride (TiN) in various ways.

1.2.1 Formation of TiN

As mentioned above, annealing is a key step in providing a low-resistance contact. Simply depositing titanium is not enough. It is therefore put forward that a key reaction takes place during annealing that causes the electron transport to increase across the AlGaN layer. One way this could be induced is by the formation of TiN by the extraction of nitrogen from the AlGaN layer.

It has been shown that the formation of TiN on the interface of deposited Ti on AlGaN, by extracting nitrogen from the latter and bonding it to the former, is essential to create a good bonded contact with low contact resistance [2]. Since TiN has a lower formation enthalpy than GaN (-336 kJ/mol vs -111 kJ/mol respectively) it will form naturally at the interface of these materials, provided the annealing step catalyses the reaction sufficiently. Wang et al. [2] provide a detailed view of the formation of TiN at a Ti/AlGaN interface, which creates islands of TiN in various sizes and shapes depending on temperature and duration of annealing. The formation of these islands can lower the resistivity of the structure from the titanium to the 2DEG in two ways that as of yet have not given any clue to which one is dominant in this structure.

1.2.1.1 Lowering of the Schottky barrier

TiN is a semiconductor with a higher electrical conductivity then AlGaN. The nitride also has a lower work function in comparison with AlGaN (3.74 eV vs 4.33 eV [5]). From the Schottky-Mott theory one can derive that an intermittent layer of TiN between the Ti and the AlGaN can reduce the barrier height compared to the direct interface. This increases the probability of carrier transport through the layer and thus the macroscopic conductivity of the structure.

1.2.1.2 Nitrogen vacancies

Another hypothesis is that the presence of TiN is not of significant importance, but rather the absence of the needed nitrogen in the AlGaN layer. This depleted layer in the AlGaN, where nitrogen vacancies are available, acts as a donor layer in a heavily n-doped semiconductor. The doping causes significant band bending in the AlGaN which in turn reduces the depletion layer width and hence increases the chance of electrons simply tunnelling through the layer to the 2DEG. In course, these type of contacts are called tunnelling contacts. It is stated in literature [2] that both the lowering of the Schottky barrier, as well as the formation of a donor layer contribute to the increased carrier transport, however it is unknown which process dominates the increase [6].

1.2.1.3 Metal penetration

A third hypothesis suggested has to do with a different formation of titanium nitride, specifically caused by the lattice mismatch between the AlGaN and GaN layers. GaN has a wurtzite structure with lattice constants of 3.189 Å and 5.178 Å. AlGaN is formed Metal Organic Chemical Vapor Deposition (MOCVD); chemically combining various organic and metallic compounds at high temperatures. This creates two stable materials: AlN and GaN. by alloying the AlN and GaN together in the required quantities, one can create the correct composition of AlGaN, but although AlN has a wurtzite structure as well, its lattice constants are 3.111 Å and 4.979 Å [7]. This difference causes a significant lattice mismatch when depositing these materials on top of each other as is done in this heterostructure. Lattice mismatch can in some situations lead to dislocations, as can be seen in figure 3.

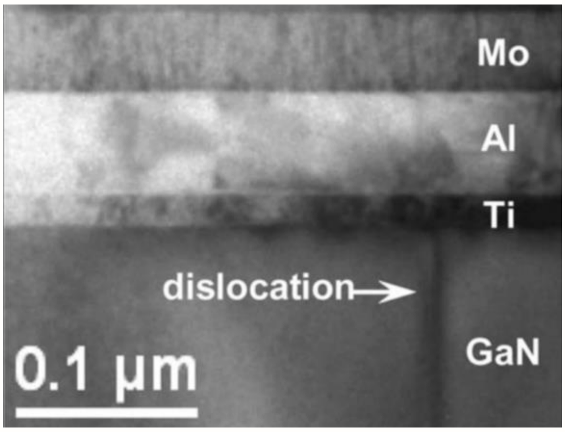


Figure 3. TEM image of a dislocation in crystalline GaN. [2]

When a titanium contact is deposited on the heterostructure and annealed, the combination of these dislocations and the higher temperature possibly creates considerable cracks in the AlGaN layer. Metal can then fill these cracks, creating a direct contact to the 2DEG below.

Although the AlGaN should be homogeneous in distributing N through its structure, chances are a significant pocket of either AlN or GaN is available near dislocations. In fact these inhomogeneous pockets could be the reason of the formation of cracks or dislocations in the first place. If this is the case, titanium could alloy with these materials, forming TiN, which has a much lower resistivity than AlGaN, making electronic transport through anything other than these spikes unlikely. Because the underlying layer of GaN contains nitrogen as well, the spikes can penetrate deeper and deeper into the substrate.

A severe problem with this hypothesis, is that the damage to the AlGaN layer would have a severe effect on the (behaviour of) the electron gas. Since the 2DEG only exists by the combination of electronic properties of the GaN-AlGaN heterostructure, one could imagine that damaging this structure could influence the stability, size, or even existence of the 2DEG. Wang et al. [6] have investigated this phenomenon in an AlGaN/GaN heterostructure with a 13 nm AlGaN layer on sapphire with an 1.5 µm GaN buffer layer. When annealed both TiN islands and Ti-alloy spikes were found, as can be seen in figure 4. The spikes ran through threading dislocations that reached as deep as 130 nm. It was found that the 2DEG has been destroyed in the area surrounding these spikes. Naturally, then, this transport mechanism is not likely to be favourable for applications that require the manufacturer to be able to control the contact resistance accurately; a requirement that is unilaterally demanded in the semi-conductor industry.

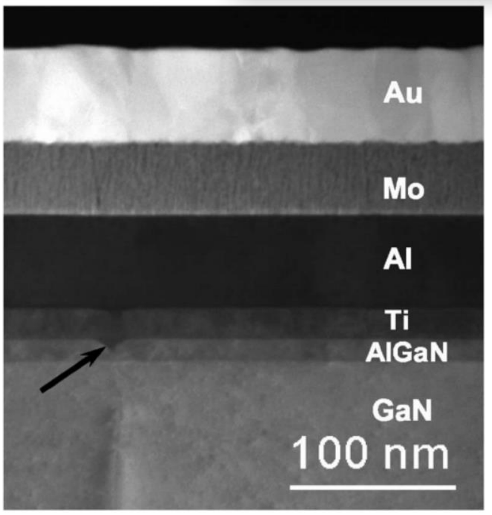


Figure 4. TEM image of a cross-section in the sample made by Wang et al. The arrow shows a significant metallic spike into the AlGaN layer [6].

Chapter 2.

Measurement techniques

In order to investigate the validity of the different transport hypotheses put forward in chapter 1, various measurement techniques have to be applied.

2.1 Scanning Probe Microscopy

Scanning probe microscopy (SPM) is a field where imaging of a surface at an atomic scale is realised by using a physical probe (usually a microscopic tip) to measure various forces and properties close to or on the surface of a material. A topographical image can be achieved by scanning the probe over the surface. Two techniques were used: Scanning Tunnelling Microscopy and Atomic Force Microscopy.

2.1.1 Scanning Tunnelling Microscopy (STM)

STM was invented in 1981 by Binnig and Rohrer et al. [8], and is a technique that enables the imaging of a surface of a conducting sample, while also revealing the local electronic properties down to an atomic scale [9]. The essential components of an STM are an atomically sharp tip, scanned across a surface by piezoelectric tubes, a system to determine and alter the distance between the tip and the sample and an electronic setup to measure the current between the sample and the tip, by way of tunnelling.

2.1.1.1 Tunnelling

Tunnelling is an electronic transport mechanism that arises from the basics of quantum mechanics (Appendix B). From a classical point of view, an object hitting an impenetrable barrier will not be able to pass through it, but when particles of a very small mass, for instance, electrons, are considered, their wavelike probability functions can have a nonzero value on the other side of the barrier, provided said barrier is low or thin enough.

To electrons, a region of low conductivity, like an insulating material, but also a vacuum or simply air, is such a barrier. And in general electrons are thus unlikely to pass through the air from one conductor to another. However, when the distance between the two conductors is small enough, a nonzero chance exists for the electrons to jump from one to the other. This phenomenon is called tunnelling.

From a mathematical point of view tunnelling can be considered by determining the wave function of a beam of electrons. The one dimensional time independent Schrödinger equation:

$$-\frac{\hbar}{2m}\frac{\partial^2\psi_n(z)}{\partial z^2}+U(z)\psi_n(z)=E\psi_n(z)$$

where E is the energy of a given electron and U is a function describing the potential landscape on the z axis. If the energy of the electron is higher than the surrounding potential, as is the case inside a conductor, the solution for the wave function will be of the form of a travelling wave:

$$\psi_n(z)=\psi_n(0)e^{\pm ikz}$$

While in the case of, say, a high potential barrier, $E < U(z)$, the solution will be that of a decaying wave[4]:

$$\psi_n(z)=\psi_n(0)e^{\pm kz}$$

The constant in the exponent k is different for either solution:

$$\begin{array}{cc} travelling\ wave & decaying\ wave \\ k=\frac{\sqrt{2m(E-U(z))}}{\hbar} & k=\frac{\sqrt{2m(U(z)-E)}}{\hbar} \end{array}$$

With the wave function of the electrons one can determine the probability of finding an electron as a function of location z. Given an electron with finite positive energy, this will be free to move like a traveling wave through a conductor, but at the edges, where U exceeds the energy of the electron, the probability will decay exponentially. There is however a small chance of finding the electron just outside the conductor.

---

[4] For those having difficulty spotting the difference: The i is missing in the exponent, meaning we are now using real instead of imaginary exponential functions.

In the case of a conducting tip and sample brought close together, one can find that the wave functions of electron states in both materials will start to overlap, creating a wave function with a nonzero probability of finding an electron on either side of the gap between the tip and the sample. This probability is given by:

$$P \propto |\psi_n(0)|^2 e^{-2kd}$$

where d is the width of the barrier. For this setup, given a positive bias on the tip and a grounded sample, one can determine U:

$$U_{tip} = -qV_{tip}$$

$$U_{gap} = U + qV(z)$$

$$U_{sample} = 0$$

where V_{tip} is the tip bias, $V(z)$ the potential at any point between the tip and the sample, and U the barrier height. If the tip voltage is small compared to the bias voltage, one can simplify k to:

$$k \approx \frac{\sqrt{2mU}}{\hbar}$$

This means electrons close to the Fermi level, can be excited by the bias voltage and tunnel across the barrier when their energy is higher than $E_F - eV$ [5].

2.1.1.2 Local density of states (LDOS)

Because the amount of states that electrons can tunnel to or from is finite, the tunnel current is a sum over all these states:

$$I \propto \sum_{E_F - eV}^{E_F} |\psi_n(0)|^2 e^{-2kd}$$

One can sum the probabilities associated with the energies between $E_F - eV$ and E_F to find the total number of states available in this energy range, which is equivalent to the local density of states. The density of states is given by:

---

[5] eV being electron charge times bias voltage, not to be confused with the unit of electronvolts.

$$\rho(z,E) = \frac{1}{\epsilon}\sum_{E-\epsilon}^{E} |\psi_n(z)|^2$$

So for a given bias and known LDOS of tip and sample:

$$I \propto V\rho_{sample}(d,E)\rho_{tip}(d,E)$$

For the subject of this thesis, in particular the proof or disproof of the hypothesis relating to the Schottky barrier, it was of importance to determine the LDOS from the tunnelling current and known bias, as to find out more about the electronic properties of the sample. For general purposes the one dimensional approximation transmission coefficient, T, introduced by Simmons [10], is used:

$$I(z,V) = \int_0^{eV} \rho_{sample}(d,E)\rho_{tip}(d,E-eV)T(z,V,E)dE$$

Where T:

$$T(z,V,E) = e^{\left(-\frac{2\sqrt{2m}}{\hbar}z\sqrt{U-\frac{eV}{2}-E}\right)}$$

In general the z dependence in this coefficient can be neglected[6], giving a well-defined relation between set bias voltage and measured current as a function of distance to the sample and the LDOS. With these mathematical tools, images can be made and interpreted.

2.1.1.3 Principle of operation

As mentioned earlier an STM uses a tip brought close to a sample to measure a tunneling current. The tip in question usually is made of a tough and well conducting material and is made atomically sharp. This is done to be sure that only one part of the tip (the outermost end) is tunnelling electrons to or from the surface. If the tip would be blunt (i.e. being over 1 nm wide at the end) tunnelling current could come

---

[6] As our research group shared some interest in the theoretical evaluation of this transmission coefficient, some research went into the possibility to measure the LDOS through dz/dV curves. This has not proven useful during the course of this thesis, though, and therefore is not further explained here.

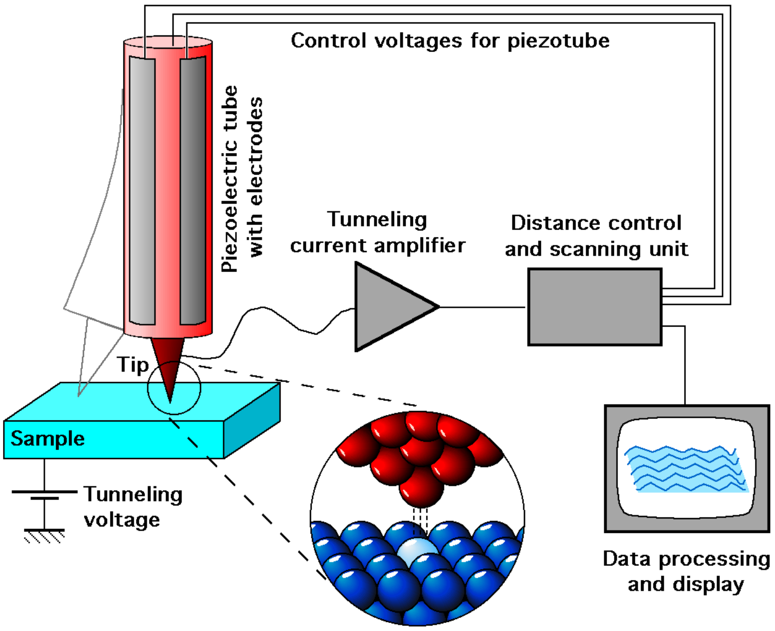


Figure 5. Schematic representation of a Scanning Tunnelling Microscope. [11]

from multiple places on the sample and images would become distorted at atomic resolution. When bringing the tip of an STM close to the sample and applying a bias to the tip or sample a tunnel current can be induced and measured by an electric circuit attached to them. Using piezoelectric tubes the tip can then be moved around over the sample to determine the tunnelling current at various places. When the STM operates in this way, it is said to be in constant height mode.

This is not always enough, however, to image a sample. The distance at which a current is significant enough to measure (~Å), is much smaller than the average height profile of most samples[7], making the chance of crashing the tip into the sample fairly large, which could result in destroying the fragile end of said tip. To prevent this, a feedback system is used to alter the height of the tip. This system measures the tunnelling current and keeps it constant, varying the distance of the tip to the sample to do so; hence this is called constant current mode.

This has, however, severe implications on what an STM image will look like, because a change in current can be caused by more than just a height difference:

---

[7] Imagine flying a plane two meters from the ground, while flying in the Alps; Tough not to hit a mountain (or a tree for that matter).

different atoms, or different electronic properties, a change in bias voltage, etc. To compensate for just the height difference, the operator of an STM needs to adapt the speed at which the feedback system reacts (how much change in current before the height is altered), as well as the speed it reacts with (how much change in height for how much change in current) to the conditions and the sample at hand[8].

2.1.1.4 Spectroscopy

There is another mode of operation for an STM, namely STS: Scanning Tunnelling Spectroscopy, where, apart from topographical images, the electronic properties can be imaged in detail. Here the tip is held over one place on the sample at a specific height, or a specific bias, while respectively the bias and height are varied. Through this, a rate of change of the tunnelling current as a function of either the height or voltage can be determined. This can be used to determine the local density of states of the sample (more on this in paragraph 2.1.1.2). It can also be determined if a material is either metallic or semiconducting, by looking at this rate of change near zero bias voltage. If this rate of change is zero (the curve is flat) no states are available at this voltage. But since this voltage corresponds with the Fermi level, this shows whether states are available around E_F, thus corresponding to a semiconductor when no states are found, and a metal when states are available; or a flat and sloped curve at zero bias respectively.

2.1.1.5 Application in this thesis

Using STS, it should be possible to determine the local density of states over a cross-section of the HEMT under investigation. This should give a detailed picture of how the density of states around the Fermi level develops over the AlGaN-GaN interface. In particular it should give a clear image of the 2DEG and it could prove conclusive in proving or disproving the hypothesis of the lowered Schottky barrier. In theory, the atomic resolution of the STM could also show the possible available states caused by the nitrogen vacancies. This, however, could prove difficult as it requires ideal conditions that are difficult to acquire with this sample (see chapter 4).

---

[8] Take it from the author that this is as a skill that takes quite some time to acquire.

2.1.2 Atomic Force Microscopy

Contrary to an STM, an AFM, invented by Binnig et al. [12], senses not the electric current through a tip, but the mechanical deflection of a cantilever caused by forces applied to it by the sample. Since these forces can in theory be caused by anything, the general name of Atomic Force microscope is used for many different techniques, but in particular applies to mechanical contact forces and van der Waals forces [13].

2.1.2.1 Basic principles

An AFM consists of two basic components: a cantilever system, and a deflection measurement system (figure 6). The cantilever is usually made from a hard (scratch resistant) material and of the order of 300 µm long and 50 µm wide, with a ‘relatively large” tip at the end, used to scan the surface. This process is similar to STM and usually is controlled the same way with piezoelectric tubes.

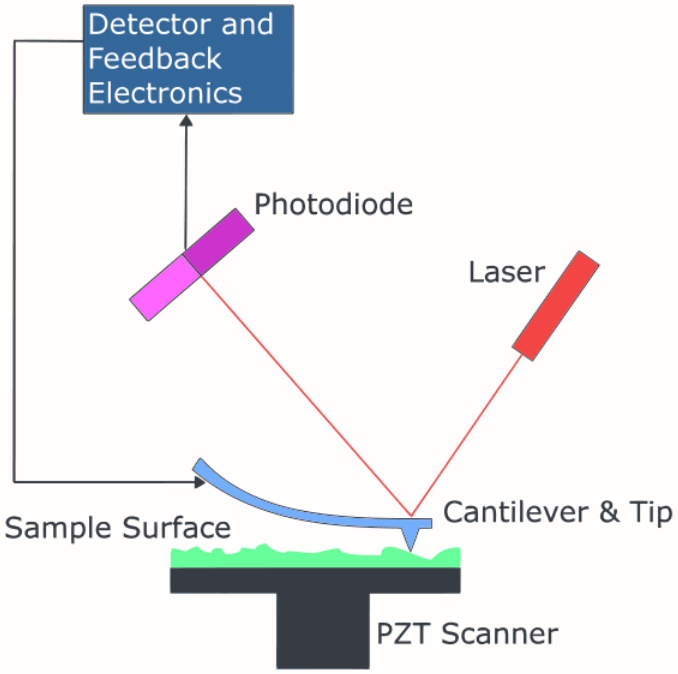


Figure 6. Schematic representation of an Atomic Force Microscope. [14]

To measure how much the cantilever is deflected from the mean average, a laser spot at the end of the cantilever is reflected at a detector, which amplifies the deflection of the cantilever in a translational movement of the laser spot on the detector. The biggest degree of freedom in this device, as mentioned earlier, is how the deflection of the cantilever is induced in the first place.

The two most frequently used modes of an AFM are contact and tapping mode, both of which rely primarily on mechanical contact forces and van der Waals forces. In contact mode, the AFM tip is brought in contact with the sample until the cantilever is deflected a certain distance congruent to a specific force. The tip is then run across the sample and the deflection of the cantilever measured, which is of course a function of the height of the sample.

Similarly to the STM, however, the height profile can exceed the limits of what the AFM can properly measure, as the cantilever will not per se bend down in a deep hole, or the tip could bury into the sample. For an AFM then, a feedback system should be used as well. Instead of keeping the height of the tip constant, it keeps the deflection spot of the laser at the same location, and hence keeping the force applied to the cantilever constant, by varying the height of the tip.

But even with a feedback loop, the contact mode of an AFM can be too destructive for a good measurement: because the tip is really hard and a force is applied to it, some softer materials are left with significant grooves when measured in this way[9]. For this reason, another mode, tapping mode, is commonly used. Here the cantilever is vibrated by piezoelectric actuators on the base of the probe, at a frequency just above its resonant frequency (at an amplitude of ~ 1 nm), and then brought close to the surface. Van der Waals forces or any other forces acting in close proximity to the surface, act to decrease the resonance frequency of the cantilever, meaning the frequency at which the tip is oscillated is removed further from the resonant frequency, thereby reducing the amplitude of the vibration. This in turn reduces the maximum deflection of the tip, which causes the feedback loop to lower the tip to the surface, reducing the resonance frequency, etc. This process is continued until an equilibrium is reached, at which the height of the tip above the sample follows the profile of the sample directly when the tip is moved across it. Again, a topographical image can be made.

It is important to note, though, that contact mode is still widely used and has many advantages over tapping mode. For instance, the resonant frequency of a tip is not always particularly well defined and its variation due to imposing forces from the sample not always known or constant. This makes most tapping mode images less accurate than those made in contact mode. Contact mode furthermore does not have

---

[9] These are sometimes called “ploughmarks”, given the samples resemblance to a recently ploughed field.

to rely on a difficult electronic controlling system, and the deflection of the laser, provided a low enough force is chosen as to not bury the tip, is solely dependent on the height profile of the sample. Lastly, contact mode can have significant advantages to show different regimes on a sample consisting of different materials. Because different materials have different drag coefficients, hardnesses and interactions, it is not unusual to see the boundaries between materials by crossing them with an AFM in contact mode. This will become of significance later on in this thesis.

2.1.2.2 Alternative forces

Apart from forces directly connected to the topography of the sample, other forces can be used to deflect the AFM cantilever as well. A variety of forces has been used in these measurements, amongst which are: thermal forces, magnetic forces, photo thermal forces, electrostatic forces, Casimir forces, chemical bonding forces, and capillary forces. For the subject of this thesis, specifically electrostatic forces have been investigated further. This EFM technique (Electrostatic Force Microscopy) [15], although working from a very different concept than the STM, is able to determine very similar properties of a sample.

EFM is, like tapping mode, a technique where the AFM probe is oscillated above the sample. Here though, the distance to the sample can be significantly higher, as the electrostatic force is much stronger than the van der Waals force. When a bias voltage is applied to the tip an electric field will arise between the tip and the sample, comparable to the field between two parallel plates in a capacitor. The tip will be attracted to the surface and its maximum amplitude will be altered accordingly. Naturally the electric field will vary depending on the work function of the sample, and thus one can image this as a function of location on the sample. This can give an insight into the doping and band-bending in semiconductors, as well as charge trapping in more insulating layers.

The same application of a bias in contact mode is called current sensing AFM. Here a topographical image is made with the normal mechanical contact force, but the current flowing through the tip and sample gives a separate measurement of the local conductivity of the sample. This provides a detailed visualisation of metallic layers and even individual dopant atoms, provided the tip is sharp enough.

Unfortunately, in contact mode, most AFM tips aren’t that sharp, because this would affect the rigidity of the tip and could break it easily. Similarly, because of the high forces electric fields can create, most conducting AFM tips are much wider than STM tips (~ 10 nm).

2.1.2.3 Application in this thesis

AFM in itself is a technique to investigate surface topography, which, apart from checking sample quality, is not of very much use for this research. However, current sensing AFM or EFM could be used to determine the presence of nitrogen vacancies in the AlGaN layer, and, more promising, the penetration of metallic compounds through the AlGaN to the 2DEG. These regions should have a much higher conductivity then the surrounding semiconductors. Finally, the many different stacked layers available in a cross-section could be visualised by an AFM contact mode image, provided they have different interactions with the probe, or different hardnesses. The usefulness of this depends on the preparation of the sample (see chapters 4 and 5).

2.2 Scanning Beam Techniques

Instead of a probe, as in SPM, in the techniques discussed in this paragraph a (charged) particle beam is used to scan the surface. The particles (electrons, protons, ions, etc.) are collimated into a narrow beam that is shot at high velocity on or into the target surface. Particles are either transmitted or reflected by the target and detectors surrounding the device can measure the presence or energy of these particles to determine various properties of the material. Various forms of scanning beams have been used during the research done for this thesis, three of which are explained in more detail here.

2.2.1 Scanning Electron Microscope

An SEM [16] is a device that uses a beam of collimated electrons fired at a conducting sample. A schematic representation of the device is shown in figure 7. Although non-destructive, the electrons can have various interactions with the material. The most commonly used of these is the ejection of secondary electrons. These are electrons that are ejected from the orbitals of the atoms of the sample by inelastic collisions with the incoming electron beam.

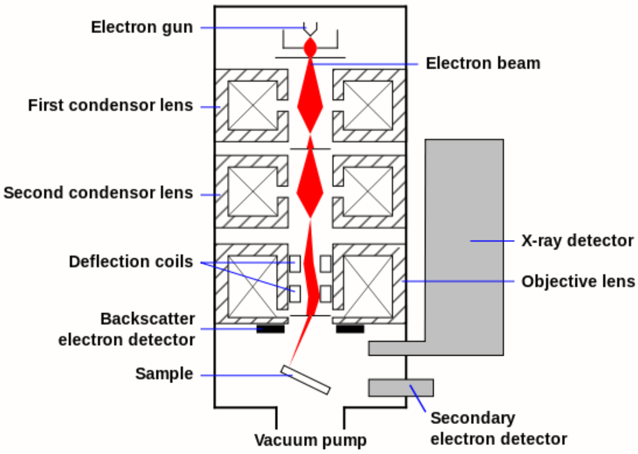


Figure 7. Schematic representation of a Scanning Electron Microscope. Note the various detectors at the bottom near the sample. [17]

Another, more high-energy source of signals is that of backscattered electrons from the beam itself. In order to completely reflect these electrons, instead of just scattering them, a material needs to be heavy, so high-mass elements tend to scatter more of these electrons and will appear brighter in an image.

Both modes of operation can be used to look at the surface but also to look deeper into the sample by increasing the energy of the electron beam (this makes it more likely the electrons will penetrate deeper into the material before reflecting or scattering). There are also many other forms of detection possible, such as the detection of Auger electrons and x-rays. These are not used in this research though.

2.2.1.1 Application in this thesis

SEM was used to image the surface of the cross-section while preparing it for AFM and STM measurements. During this, some unexplained phenomena were observed, which will be further discussed in chapter 5.

2.2.2 Helium Ion Microscope

A Helium Ion Microscope (HIM) [18] is a very similar device to a SEM, as it too uses a collimated beam of charged particles. The particles in question, though, are not electrons but, one could have guessed it, helium ions. Because these are much heavier than the aforementioned electrons, much more energy can be put into the beam. Also, the de Broglie wavelength of the helium ions is much shorter due to the increase in mass[10], resulting in a better confinement of the beam. This, combined with the change in shape of the contributing area of secondary electrons, increases the resolution of a HIM significantly compared to a SEM. The shape in question, instead of parabolic, as is the case with electrons, is pear-shaped, due to the ions having a larger penetration depth before scattering. This leads to a smaller area of the surface contributing to the image.

The interaction of helium atoms with many substrates is quite high, leading to a high secondary electron count, which makes it possible to image at very low currents, making images sharper and more focused on the surface[11]. Helium ions can damage the sample if the current is higher, though, but this is, in general not necessary.

To create a beam of helium ions in a vacuum is not trivial. Unlike electrons there isn't a source material one can generate ions from. The only way is to use helium as a gas and collimate it into a beam by ionising it at a biased tip. This can be difficult in

---

[10] For an explination of this phenomenon, see Appendix B.

[11] All scanning beam techniques using secondary electrons do not image a surface, but a certain volume below that surface. The smaller that volume, the sharper the image will be, and the better the surface can be distinguished and investigated.

a vacuum system, as injecting a gas usually isn't favourable for the vacuum. Also, the aforementioned tip is difficult to create and control.

2.2.2.1 Application in this thesis

A further investigation into the phenomena found in SEM was done in the HIM, because of the higher resolution it provides and the smaller bulk volume contributing to the image. Furthermore, it was used to try to image the stack structure after preparation.

2.2.3 Focused Ion Beam

A Focused Ion Beam (FIB) [19] is a device very similar to a HIM, except for the fact that it uses much heavier ions to shoot at the target; usually gallium is used. The result is a much more destructive form of imaging, as the heavy ions that bombard the sample sputter away material. Given the beam of a FIB is very well confined, the FIB is used in many different fields as a micro-machining tool, drilling holes and levelling planes on a scale of ~100 nm.

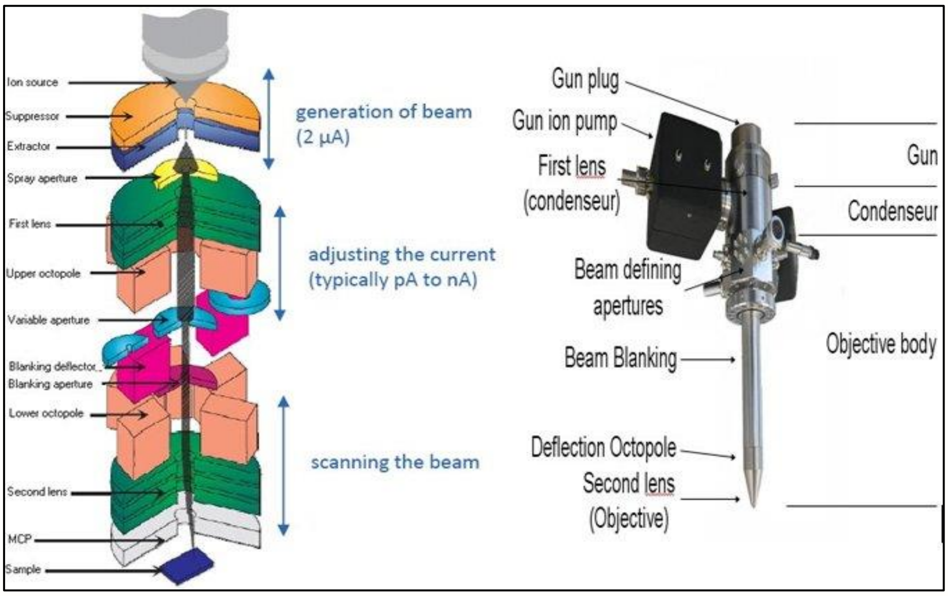


Figure 8. Schematic representation of the beam column of a FIB [20].

As an imaging tool, the FIB offers similar resolution as a HIM or SEM, but because of the heavy ions, imaging can take place at much lower beam currents. This makes secondary electron images from FIB particularly sensitive to grain boundary contrast and material contrast, which is useful for many different applications.

Imaging and sputtering are not the reason FIB is used in this research. Here it is useful because of its third application: assisted chemical vapour deposition.

Chemical vapour deposition is a method in which a gaseous compound is let into a chamber with a substrate, allowing to let the substrate adsorb (parts of) gas molecules onto its surface. When the beam of the FIB is used in a chamber filled with gas, the beam decomposes said gas into its basic components, the nonvolatiles of which (for instance the metals), can be deposited on the sample. This happens, of course, only in the places where the FIB beam is active, meaning deposition is done very locally.

In figure 9 an example of the deposition of a platinum contact (see chapter 4) is shown. As can be seen the deposition is done very locally and is confined to the boundaries of the beam.

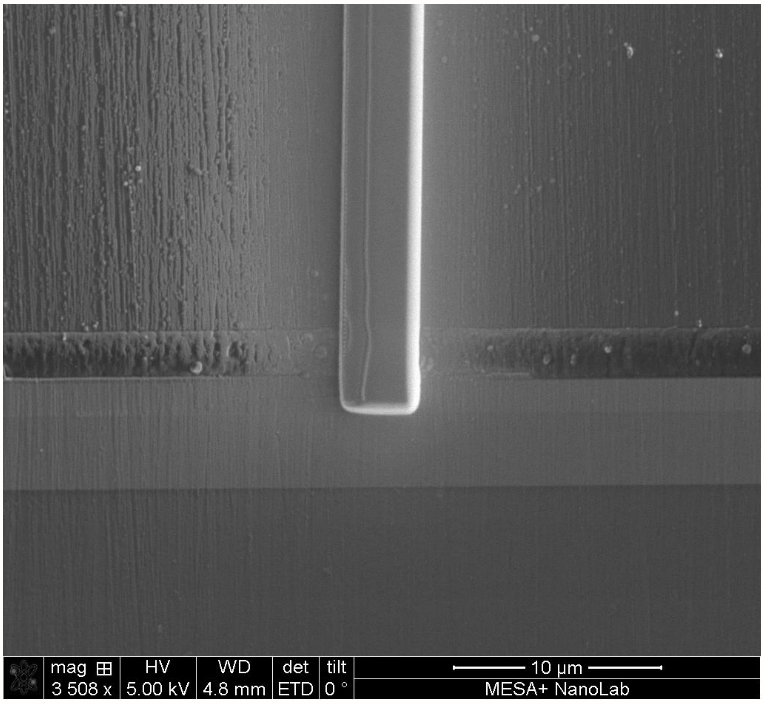


Figure 9. SEM image of a platinum contact grown on a substrate using FIB assisted CVP.

2.2.3.1 Application in this thesis

The FIB was used to deposit additional contacts onto the various transistor patterns, since the original Ti contacts were no longer available (see chapter 4). As mentioned before, the built-in SEM provided additional details requiring further investigation.

Chapter 3.

Measurement setups

Measurements were planned, as was previously stated, using AFM and STM. However, due to unforeseen circumstances, many different devices were actually used. The different real-life devices that were used are described in this chapter.

3.1 Student STM

Measurements were started with a very basic STM: the Nanosurf NaioSTM [21]. This relatively low-tech/low resolution device has been designed with educational purposes in mind and is therefore affectionately called the Student STM. The device is a feat of engineering, being less than 8 cm in diameter, but it does limit the options the operator has in modifying and tweaking measurements and data.

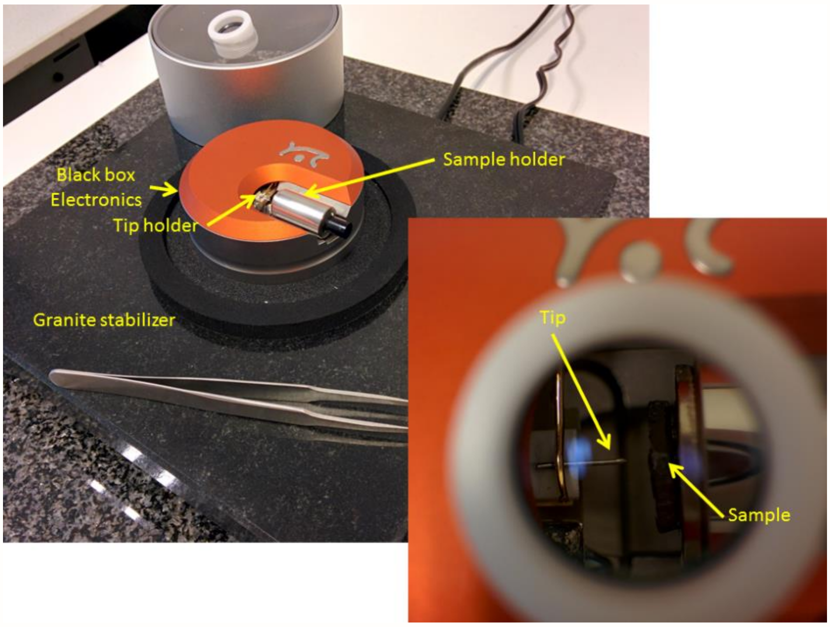


Figure 10. Image of the Student STM. In the detailed image the tip and sample holder are shown.

The Student STM, as can be seen in figure 10, consists of a fixed tip and a sample on a massive cylinder that is moved by piezo tubes. Aside from the scanning range of 500 nm, there is no lateral repositioning of the tip possible. This makes it very difficult to do any measurements on the sample, as measurements have to be taken at a specific point on the sample. Contrary to most measurements where the entire sample portrays the property under investigation, this particular research needs measurements to be done within 400 μm of the edge of the cross-section and within a region of roughly 300 μm wide. The positioning of the Student STM does not allow such precise measurements.

Instead, measurements were done to show the cross-section was smooth enough to do STS measurements on the silicon. The results of this are shown in figure 11, where one can see the average of 10 different I-V curves done on the silicon. The material shows a semiconducting behaviour, having no change in current (zero dI/dV) at zero bias. As mentioned in the previous chapter, the LDOS can be derived from such measurements and from that the bandgap can be estimated. Doing this for these measurements shows the material has a bandgap equal to that of silicon.

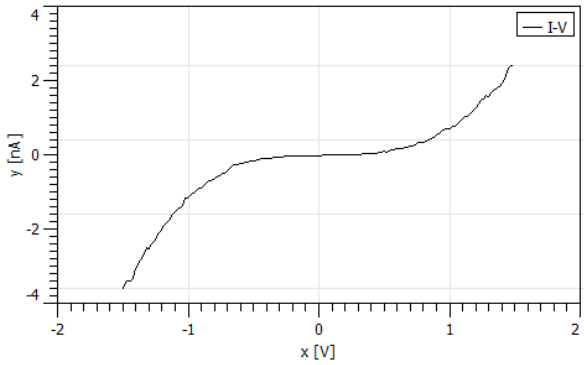


Figure 11. I-V curve derived from an STS measurement on the Student AFM, showing the material measured behaves like a semi-conductor and has a bandgap roughly the size of that of silicon.

3.2 RHK AFM

One setup in our research group contains a variety of equipment that enables it to do several measurements in one vacuum system. The RHK AFM is capable of UHV AFM measurements at a pressure of 10^{-9} Pa, has both contact mode and tapping mode probes and has contacts to do current sensing AFM as well as contacts to run an additional current through the sample (for measuring whilst the 2DEG is conducting current).

An overview of the system, is given in figures 12 and 13. The system has very elaborate control hardware with which the behaviour, position and readout of the probe and sensor can be accurately manipulated. The AFM itself sits in a vacuum system that is accessible via a load lock and is pumped using two turbo pumps. The hole system is connected by a transport rod that can manoeuvre samples and tips through the system, from the load lock to the AFM. A precision manipulator is available near the AFM as well, for moving things around in the system.

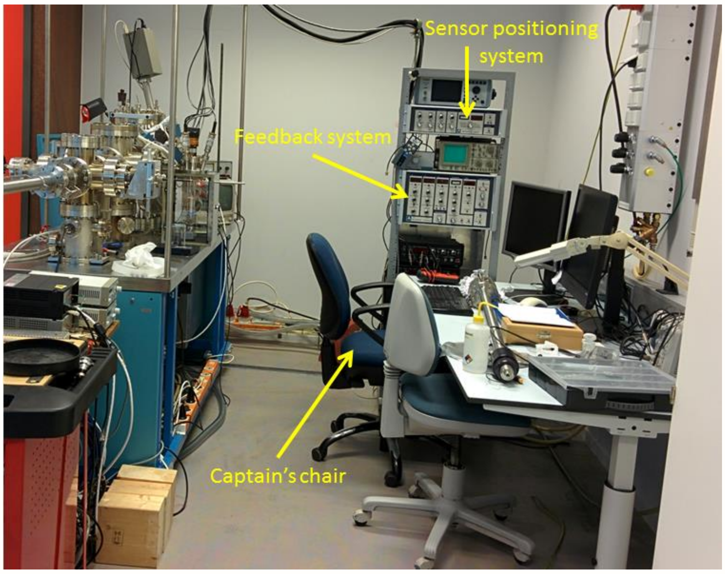


Figure 12. Image of the control systems of the RHK AFM.

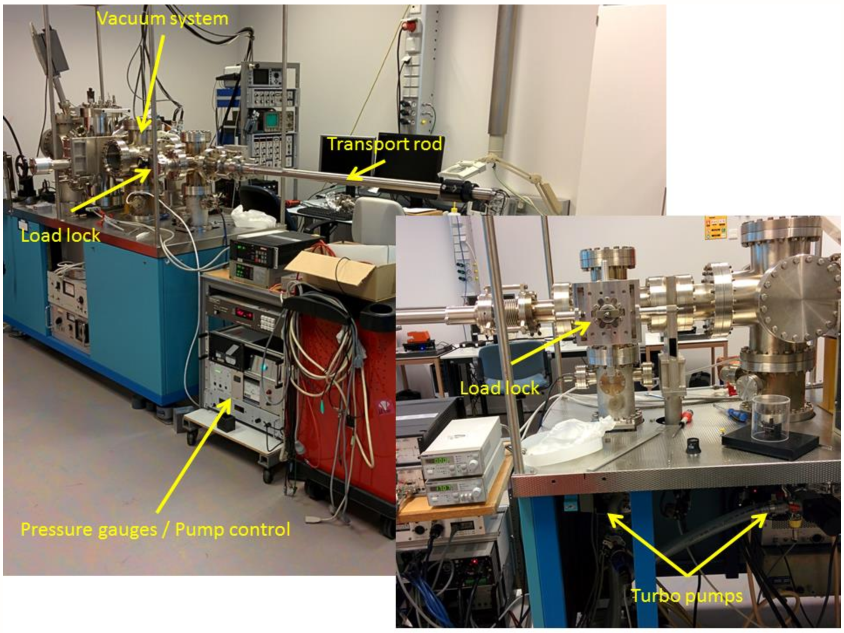


Figure 13. Image of the vacuum system of the RHK AFM.

In figure 14, the scan head of the RHK AFM can be seen. It shows a large amount of copper wires, some of which are used to control the piezo tubes and read out the detector. Several are available however, for additional applications, like running a current through the sample, or doing current sensing AFM. The tip and detector can both be manipulated, to position the cantilever correctly under the laser (coming through the normal of the scan head suspension) and position the detector to register the laser spot correctly. The entire head can be positioned as well using the piezo tubes indicated in the image.

Unfortunately, the RHK AFM setup broke down, several times, in different parts of the system. Because of this, the system, though promising, could not be used for measurements in this research. It is however recommended the system be used for any future research in this system.

For this it does have to have an additional feature, in the form of a side view camera or lens. In other measurements this has proven necessary to position the tip correctly on the sample, but the RHK system currently is unable to do this.

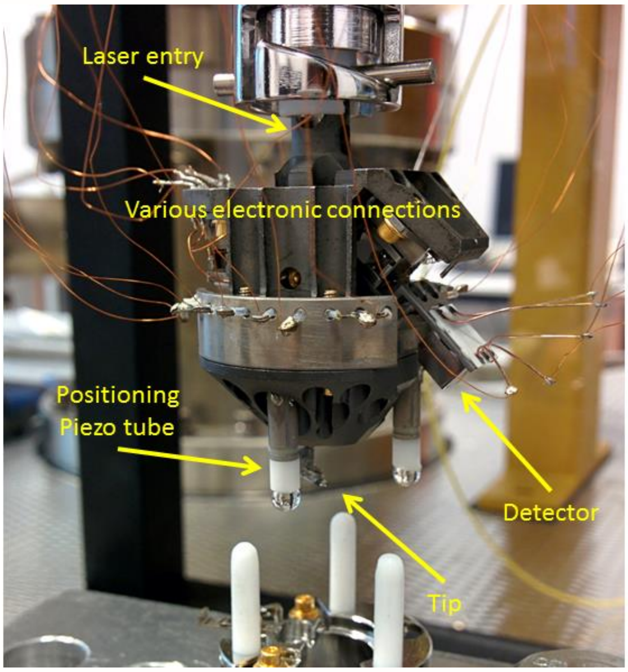


Figure 14. Detailed image of the RHK AFM scan head. The various copper wires are for different electronic connections, modifiable by the user.

3.3 Student AFM

After the RHK AFM was no longer an option and another current sensing AFM had to be found, focus was put on a new device: the NaioAFM [22] from Nanosurf, to satisfy the systematic nomenclature of our devices also called the Student AFM. This, again, is a rudimentary AFM meant for educational purposes and it therefore does not have a particular high resolution. In figure 15 an image of the Student AFM can be seen; the tip, contrary to the high end systems at our disposal, cannot be positioned in any way. Instead it is positioned by a set of grooves on the probes holder, that align it exactly with the laser detector. This makes the device much easier to use, but also less versatile.

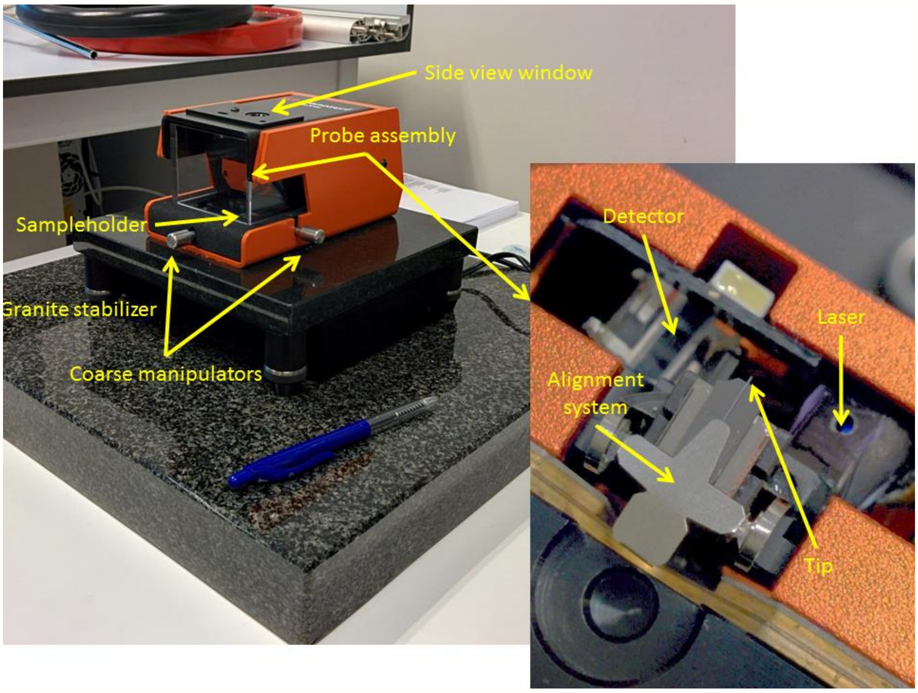


Figure 15. Image of the Student AFM. In the detailed image the tip and the fixed detector system are shown.

A great advantage of this AFM over the RHK system, is that this AFM does have a side view and top view camera. Meaning the probe can be positioned correctly, contrary to the other STM's and AFM's. The view through the side view lens is shown in figure 16. The Student AFM also has a rudimentary current-sensing system, which would make this the ideal AFM for this research. However, the accuracy of the provided I-V converter is too low to do these measurements, so another current sensing AFM will need to be found. The Student AFM can only be used for showing the topography of the sample, and perhaps to determine whether metal is penetrating the AlGaN layer.

Figure 16.Image of the view through the side view lens. The tip is seen, having the laser spot aligned on the cantilever. The sample is the blue surface seen in the centre of the image.

3.4 Zeiss HIM

To get a better idea of the topography of the sample around the Ti contacts, and to pursue some interesting findings discovered during the deposition of additional contacts using FIB (see chapter 4), HIM imaging was done using the Zeiss Orion Plus [18], seen in figure 17. The device is capable of imaging at ~5Å resolution and provides more insight in the structure of the sample.

Figure 17. Zeiss Press image of the Orion Plus HIM. The obligatory operator is not shown in this image.

3.5 Bruker Icon I-sensing AFM

Current sensing measurements could finally be made on the Bruker Icon AFM [23] from the MTP group. This AFM works in atmosphere, but has a very sophisticated positioning system, as well as the required current sensing option. An image of the setup is shown in figure 18.

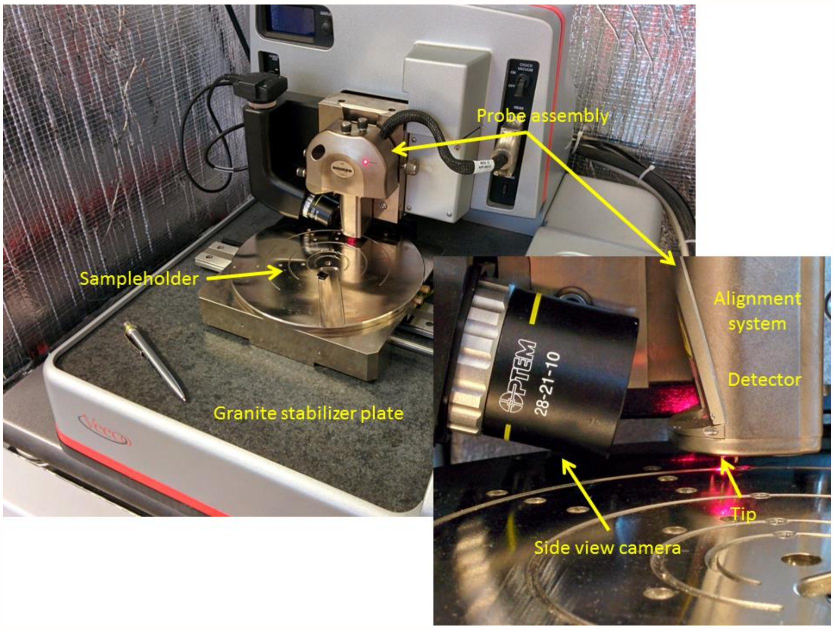


Figure 18. Image of the Icon current sensing AFM. The entire probe assembly is controlled automatically and is not accessable to the user.

The system can be accurately controlled from the software of the connected computer. A side view camera is also available, increasing the ease with which the system can be positioned correctly. The tip-detector positioning is done automatically by the device itself.

It is also interesting to note that in this system, not the tip, but the sample is moved during the measurements, For this the sample is mounted on a giant disc, increasing the stability of the measurement and lowering thermal drift.

3.6 UHV STM

After the findings done using other measurement techniques, additional measurements were done on crystalline GaN without a pattern of transistors grown on them. These measurements were done in the UHV (ultra-high vacuum) STM of the PIN group. An image of the setup is shown in figure 19. The vacuum system of this STM can go to pressures of 10^{-11} mbar; a hundred times lower than the RHK AFM. It also has a second vacuum chamber where the sample can be annealed, a feature that will prove useful in chapter 5.

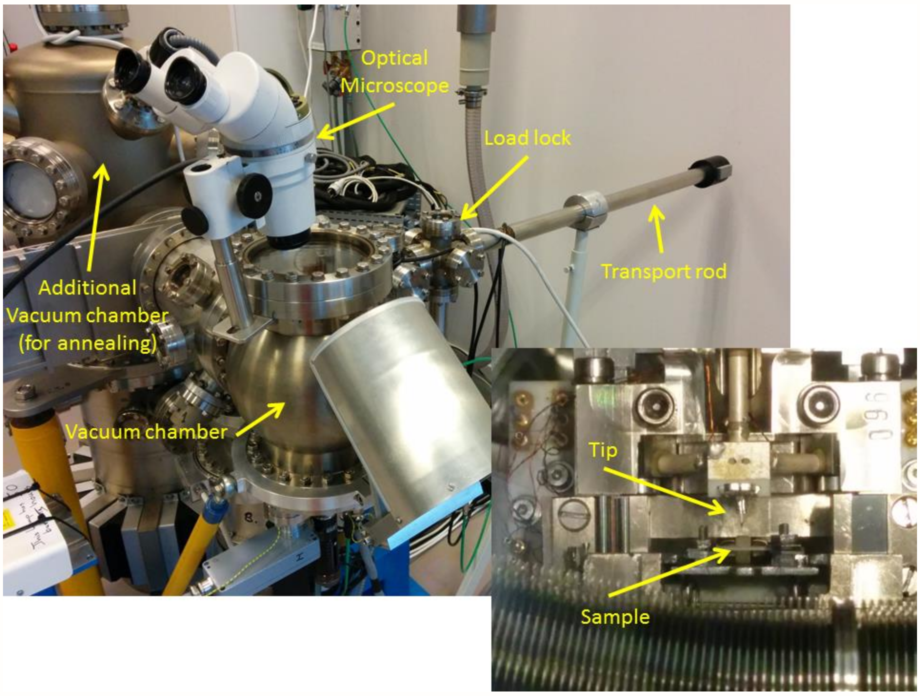


Figure 19. Image of the UHV STM of the PIN group. The additional vacuum chamber where CVP and annealing can be done is indicated on the left.

Chapter 4.

Sample preparation

The samples used for this research are GaN-on-Silicon wafers with a structured pattern of AlGaN and Ti alloyed contacts deposited on them, as can be seen in figure 20. The structure provides multiple contacts bonded on top of an GaN-AlGaN heterostructure. Used for this purpose are multiple spaced rings and circles of a Ti-Al alloy, as is illustrated in figure 21, which form the source and drain of point symmetric circular transistors on this test structure.

As the interest for this thesis lies in the behaviour around the contacts, a region where there is both a metallic contact as well as a bare substrate is preferred for making the measurements. Here we will have to do several different microscopy and spectroscopy measurements in the bulk of the wafer, and for this purpose a cross-section needs to be made.

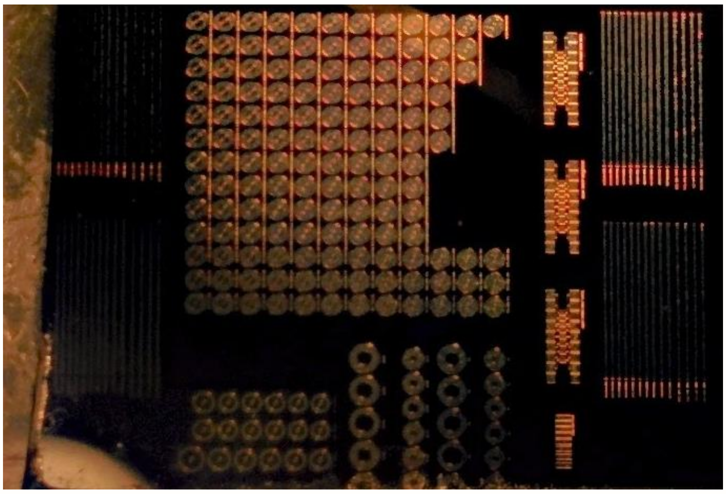


Figure 20. Image of the patterned structure on the GaN-on-Silicon wafer. Width of the image +/- 5mm.

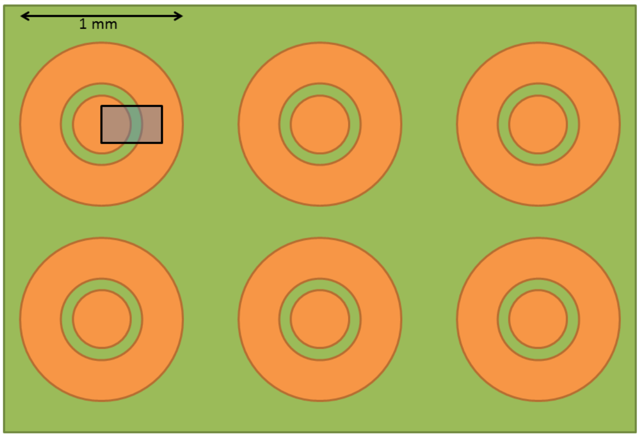


Figure 21. Topview of the sample structure. The region of interest (the edges of the contacts) is highlighted.

Due to confidentiality, a detailed image of a cross-section is not available. However, there are specifications of the fabrication steps done to create the structure. From this, a theoretical cross-sectional map can be created, that is given in figure 22. The stack starts with a 3.6 μm buffer layer of multistacked GaN layers, to reduce strain between the GaN and the Si, followed by a 1.6 μm of crystalline GaN. On top of this is a 20 nm AlGaN layer and a 3 nm layer of GaN to create the heterostructure. The final step is the Ti(alloy) contacts and an insulating layer of SiO_2 or SiN.

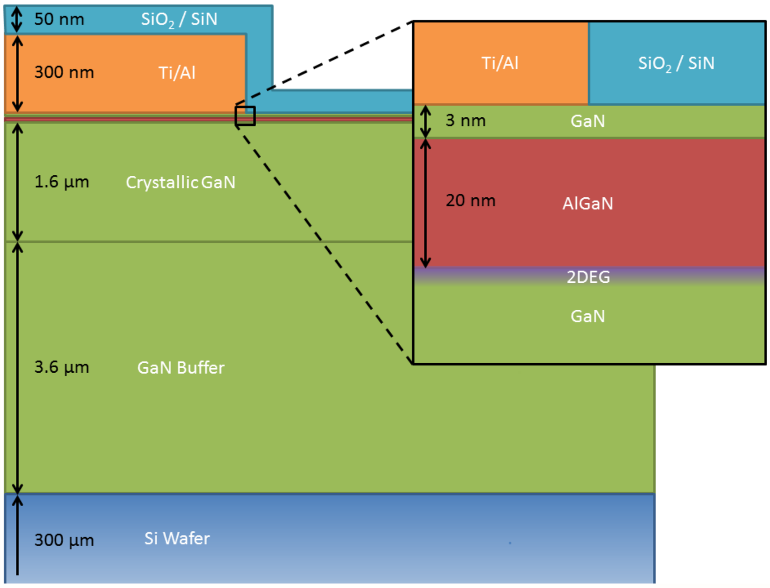


Figure 22. Schematic representation of a cross-section of the heterostructure (components not to scale). The location where the 2DEG should form is highlighted.

4.1 Creating a cross-section

To make a cross-section, basically two methods are worthy of further exploration: sawing and cleaving. From these two cleaving is the preferred choice for this research, since it usually provides a cleaner, possibly smoother surface then sawing, and has a smaller chance of destroying the heterostructure.

4.1.1 Cleaving

Cleaving is a method where one uses the prevailing crystallographic lines in the substrate to break the sample along a very straight line. This has regularly been done with silicon wafers in order to produce straight and very clean breakages, although usually research does not focus on the cross-section created.

For this sample there are other complications. For instance, Si(111) was used as a substrate, which has different cleavage planes. Furthermore, a structure is present on the wafer. This could not only alter the preferred breaking lines across the sample, but it also reduces the possible cleavages that would be useful: as can be seen in figure 23, the area of interest, though repeating, is not very large, meaning a cleavage would have to be made exactly through one (or many, in the interest of redundancy) of these structures. This complicates the cleavage procedure, especially since the symmetry direction of the silicon substrate is at an angle with that of the structure.

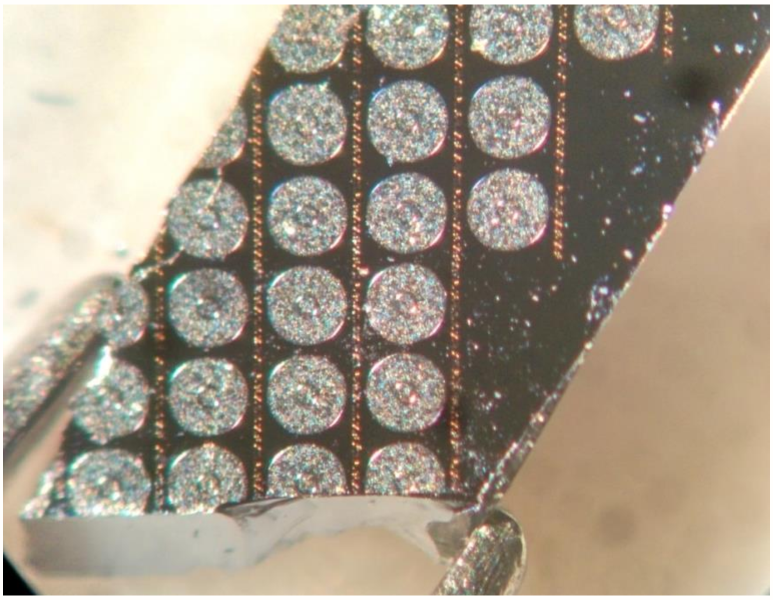


Figure 23. Optical microscope image of a cleaved sample with the preferred cleavage direction at an angle with the deposited structure.

In figure 24 the resultant cleavage can be observed through an optical microscope. As one can see the cleavage only crosses three ring structures barely, reducing the chance of finding a good cross-section of the whole transistor.

When looking at the cross-section of the wafer itself, it is found that the cleavage does not look particularly straight. Although the silicon, apart from some defects, broke fairly straight, where the cleavage comes near the surface where the structure is grown, the sample clearly did not break in a controlled manner. When one would perform STM or AFM on this sample, the chance of crashing the tip would be too great, and there is a good chance the structure would be deteriorated as well, and therefore badly visible.

Multiple cleavages have been performed, at various angles and breaking from both the back and the front of the wafer. None, however, provided a cross-section that one can expect detailed and reproducible results from. It was therefore decided to fabricate a cross-section by sawing.

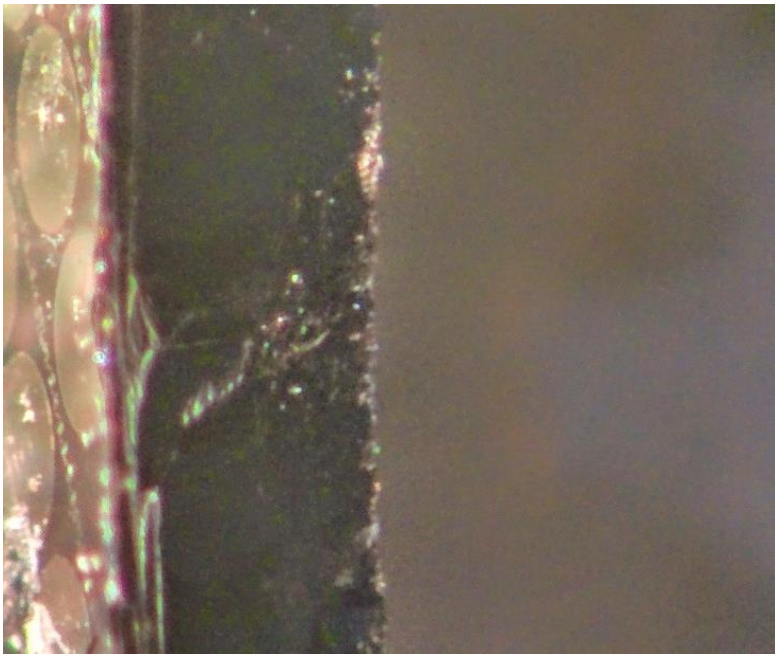


Figure 24. Optical microscope image of the cross-section created by cleaving. As can be seen, the structure is not cleaved straight, and this worsens in the last part towards the surface of the structure where the GaN is grown.

4.1.2 Sawing

When sawing a sample, one has to be particularly careful not to disturb the structure on top of the wafer in any way. In order to protect the top of this structure, a glass plate is glued on top of the structure, encapsulating the stack between the silicon substrate and the glass. The wafer and glass are then sawed in the preferred direction (along the centreline of a line of ring structures). Because the structure is sawed, there is a chance that this process damages the stack or the structure of the different layers, thereby not giving a representative view of what this structure looks like in the actual bulk. This is a serious disadvantage of this method, which will have to be assessed in due time, when the extent of the damage is quantified. To make sure the cross-section is smooth enough to measure on, the kerf needed to be polished. To this end, a polishing silica film was used to polish the sample in 4 hours. The polishing step was done on two samples, where different grades of film were used. One was polished with a 1 µm silica film, the other with a .02 µm film.

The reason for this double approach is the possible disadvantage of the finer silica film. Since the particles on this film are only 20 nm in diameter, this polishing step deposits silica spheres from the film on the sample, which can damage AFM tips and distort SEM or HIM images. On the other hand, the coarser film might leave grooves on the sample which distorts the image as well. As its beforehand unknown which effect will be least disruptive, both polishing films are used on different samples.

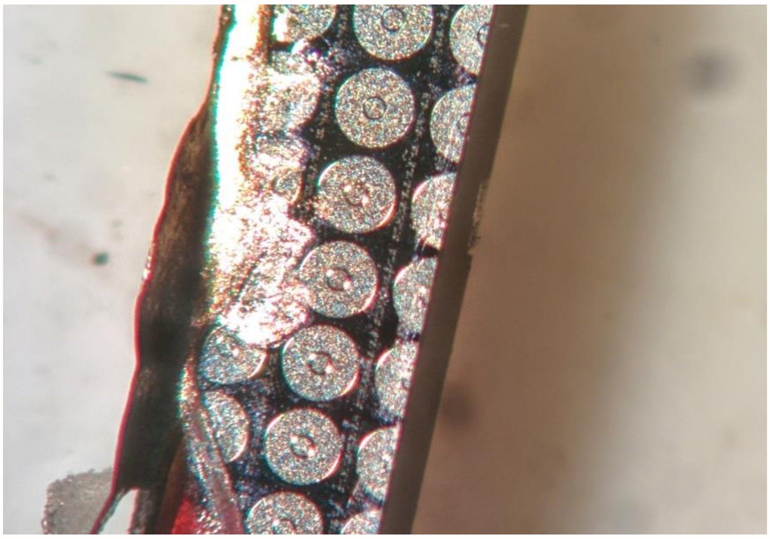


Figure 25. Optical microscope image of the sawed sample. As can be seen the cross-section runs through many different transistors.

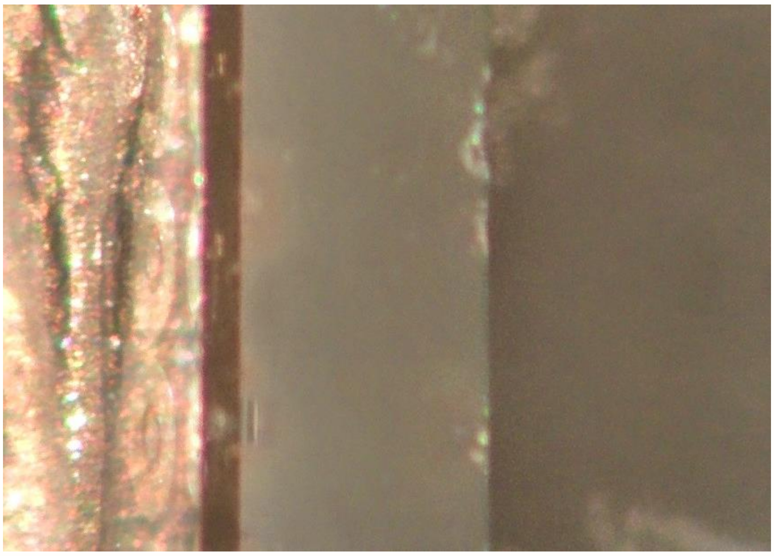


Figure 26. Optical microscope image of the cross-section created by sawing. The result is much more smooth then the cleaved sample. The dark (50 um) layer on the left side of the image is the glass plate glued on the top of the structure to protect the deposited stack.

In figure 25 a top view image of the sawed sample is shown. Clearly this cross-section has many more structures that possibly show a good interior view of the transistor stack. The cross-section itself (figure 26) also shows a much better result, with a smooth cross-section and little defects. Another benefit of the glass plate is that the area of interest is no longer on the edge of the cross-section, but well inside, on the boundary of the glass and the sample. This is fortuitous with AFM and STM, since the feedback loop would cause them to walk over the edge of the sample if the scanning range would reach that far.

4.2 Sample holder

After the sample is prepared it needs to be put into a sample holder that fits many different microscopes. Many AFM and STM setups have a very limited operational height, where they can measure samples. This makes it quite difficult to measure a cross-section, as turning the sample on its side, small as it may be, makes it over 1.4 mm high, (the width of the sample in figure 25). The sample needs to be conducting as well when it is further modified using the FIB and when measuring with STM and current sensing AFM. For this reason the sample is put on a small metallic plate using conducting silver paste.

This does not only increase the height of the sample by another millimetre, but it also is particularly difficult task because of the dimensions of the sample, and its aversion to standing on its side to let the cross-section face upward. Furthermore, it proved difficult to get the resistivity of the contact between the sample and the metallic plate to acceptable values[12]. This has an unfortunate side effect, as it has proven inevitable that some paste gets stuck underneath the sample, skewering the end result. Fortunately all scanning microscope techniques used can handle the apparent angle. In figure 27 the end result can be seen.

Figure 27. Image of the sample plate and the sample on top, slightly skewered by the amount of silver paste, of which the excess can be seen around the sample.

4.3 FIB grown contacts

For additional research into the behaviour of the devices and the 2DEG, a proposition was made to do various measurements while the transistor was operational. In this case, this meant running a current through the 2DEG while measuring with AFM or STM, in order to better determine the quality and size of the 2DEG. For this, a macroscopic contact had to be made to the transistor’s source and drain (the ring and circle in the pattern respectively). However, because of the glass plate, it is impossible to make use of these existing macroscopic contacts. Therefore, additional contacts were grown on the cross-section, that run from the existing titanium over the glass to contact pads (a schematic representation is shown in figure 28, where wires could be bonded to connect the source and drain. This way a current could be run through the 2DEG while measurements are taking place. In figure 29 an optical microscopy image is shown of one set of contact pads.

---

[12] This was eventually accomplished by using quite exorbitant amounts of silver paste to cover the entire bottom part of the sample

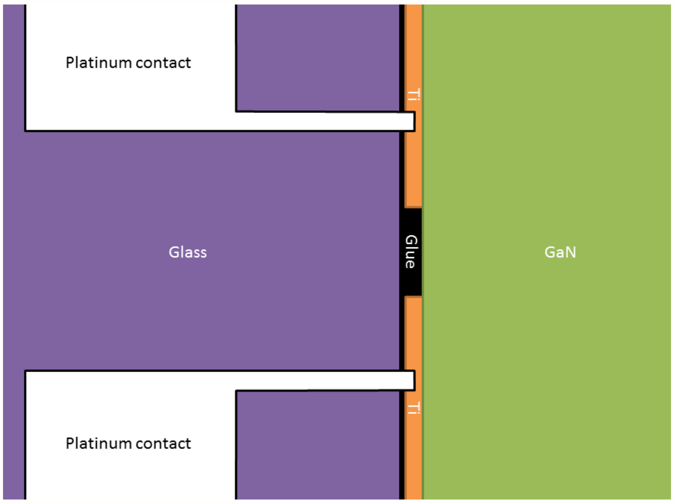


Figure 28. Schematic representation of the contacts grown using FIB (compontents not to scale). In reality the contact pads (white) extend far onto the glass (left) and away from each other (top and bottom).

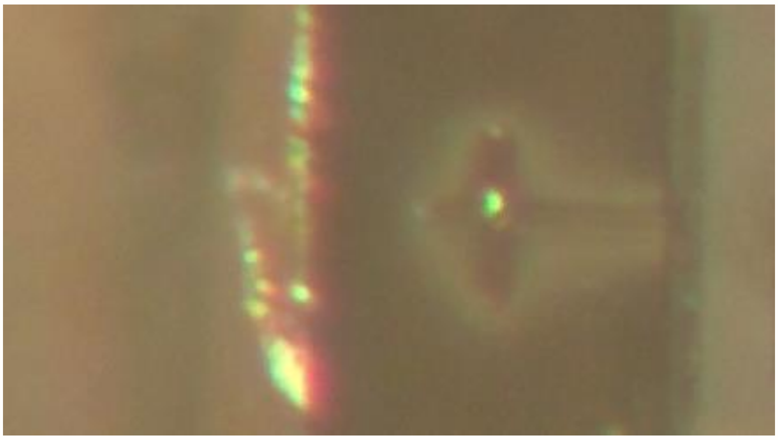


Figure 29. Optical microscope image of one of the contact pad pairs. The dark layer on which it rests is the glass plate on the topside of the structure.

Chapter 5.

Results

In this chapter, the results of the various measurements are presented. Student AFM will be discussed first, since this device provided a good look at the samples topography. These are followed by followed by the measurements done with current sensing AFM on the microscope of the MTP group. Although chronologically made first, the measurements done by the built-in SEM of the FIB are presented third, since their interpretation requires certain knowledge of the samples topography. After this the more detailed HIM images will be discussed. In the aftermath of this research, it was thought to be fruitful to measure on a unpatterned sample of GaN as well. This was done in UHV STM and these results are discussed last.

5.1 Student AFM

The first measurements done on the Student AFM were over quite a large area and at low force. This was done to determine whether the smoothness of the sample was good enough after the polishing step mentioned in chapter 4. In figure 30 one of the first images on the 1 µm polished sample can be seen[13]. The grooves that were predicted have a very disruptive effect on the measurement, as no detail from the sample can be distinguished.

---

[13] As a consequence of the way the samples are polished, the grooves always run perpendicular to the samples top edge (where it is glued to the glass). In this image however, they are running at an angle. This is done by skewing the scanning direction, as to make sure the stripes are in fact grooves and not a defect in the AFM. Such a defect would almost always show up in one line of the scan, forming a stripe laterally, or as a defect on the same point each line, forming a stripe vertically. Diagonal errors, though, are virtually impossible, and thus these stripes have to be related to the samples topography.

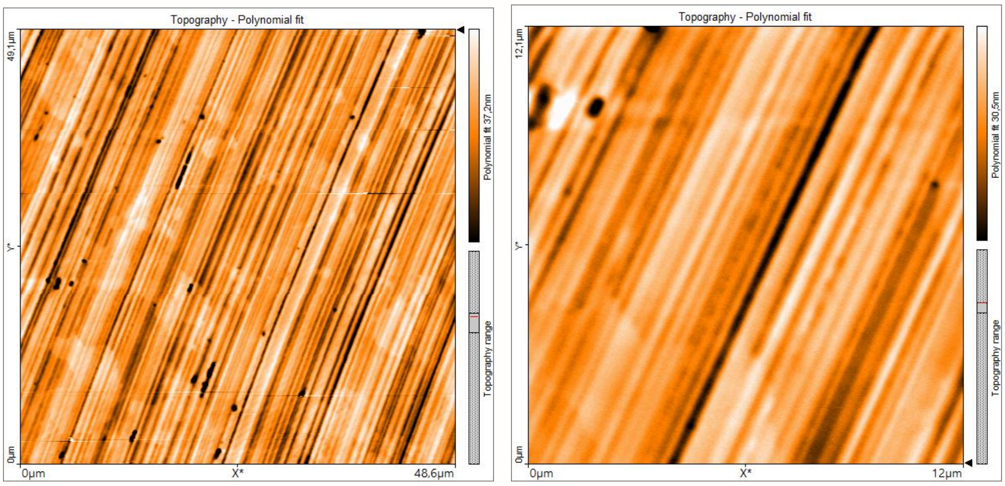


Figure 30. AFM image of the 1 µm polished sample (right: zoomed in). The polishing grooves distort the image too much for detail to be seen.

Focus was therefore put on the .02 µm polished sample, a first image of which can be seen in figure 31. The grooves in this sample are, as can be expected, are much less distorting, and details of the sample can be seen much more clearly. The high feature running vertically over the image for instance is the glue layer between the glass (on the right) and the sample (on the left). The glue is approximately 3 µm wide and varies in height across the sample (from well over 600 nm above the surrounding sample, to 300 nm deeper).

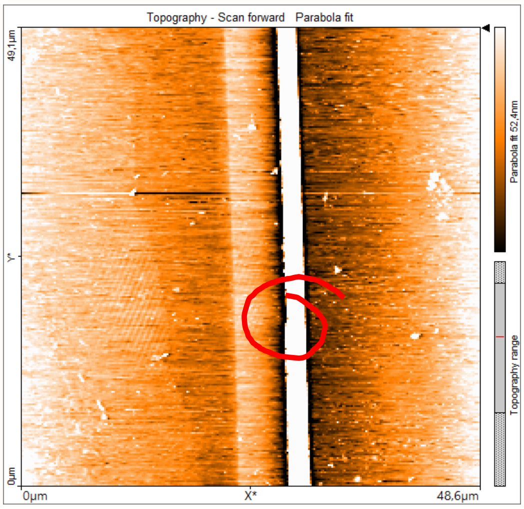


Figure 31. AFM image of the .02 µm polished sample. The grooves on this sample are much less distorting. The vertical white stripe is the glue between the glass (right) and the sample (left). Furthermore, the boundary between the GaN and Si can be seen as a difference in measured height, about 5 µm from the glue. The possible edge of one of the contacts is highlighted.

What also can be seen is the difference in hardness, as mentioned in chapter 2, between GaN and Si. This difference makes the AFM tip bump up, reading as a difference in height between the two materials; this can be seen as a slightly whiter pattern, forming consequently in every line scan, 5 µm from the edge of the glue on the sample side.

A final feature can be seen, which is highlighted in figure 31. It is a bump in the height of the sample, which can be seen by a reduction of the width of the glue. This is thought to be the start of one of the Ti contact pads, and therefore is the region of interest in this image.

Figure 32 shows the same region zoomed in, as well as the corresponding reflection image[14]. A deflection image does not show the topograph created by the software, but a direct image of the laser spots deflection as a function of position. In essence this is the derivative of the topography, meaning slopes and *changes* in height are visualised much better and, more importantly, constant features, like the grooves that remain, are filtered out of the image. In the latter image, the boundary between the contact and the substrate can be seen quite clearly, as well as the previously mentioned GaN-Si interface, 5 µm from the glue.

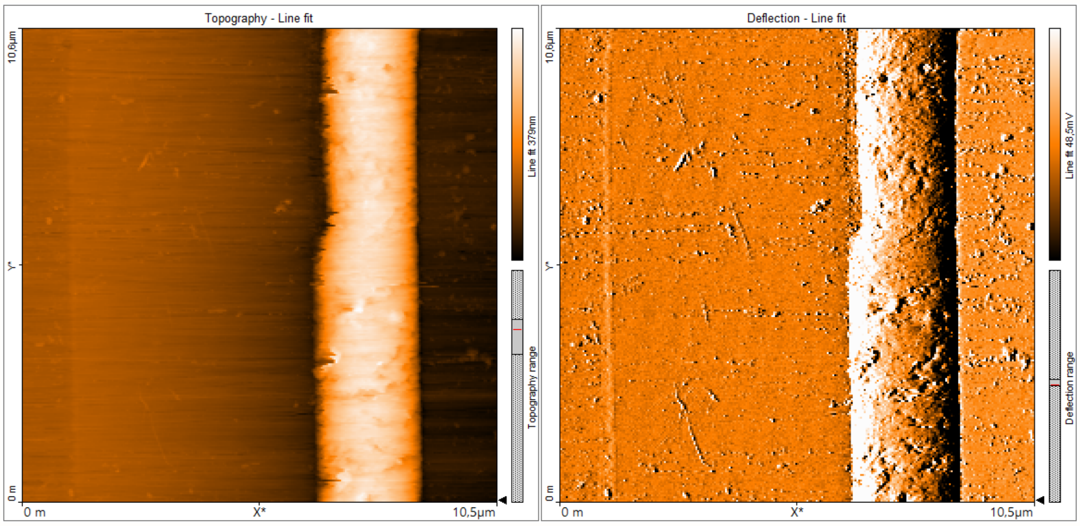


Figure 32. AFM topography image (left) and AFM deflection image (right) of the sample. The edge of one of the titanium contacts can clearly be seen, as well as the earlier seen interface between the GaN and Si.

---

[14] To a layman, one could see these images as a satellite photograph taken with the sun close to the horizon. Slopes are well lit (white) or facing away from the sun (dark).

It is worthy of note that the interface between the GaN buffer layer and the crystalline GaN cannot be seen with AFM. From this it can be concluded that these layers do not differ much in mechanical properties, contrary to electrical conduction, as will be shown in the following paragraphs. The remaining features of the stack, i.e. the AlGaN layer and GaN-Ti interface, are too small to be imaged here.

> ***Note:***
> *Given the distortion of the grooves and the subtleness of the features under investigation, as well as the limiting resolution of this AFM, using deflection images provides a better view. Therefore the remainder of the images shown in this paragraph will be deflection images.*

A further zoomed in image is shown in figure 33. This image shows the metal contact surface to be much rougher that the crystalline GaN to the left. This is thought to be a consequence of the lower hardness of the metal. The contact is shown to be 300 nm +/- 50 nm wide, which fits the growth parameters provided. On the subject of metal spikes not much can be concluded from this image. The AlGaN layer cannot be identified visually, nor can it be distinguished from the metal. It is however clear that the region where the AlGaN layer is similar in roughness to the metal contact, suggesting metal could be present there.

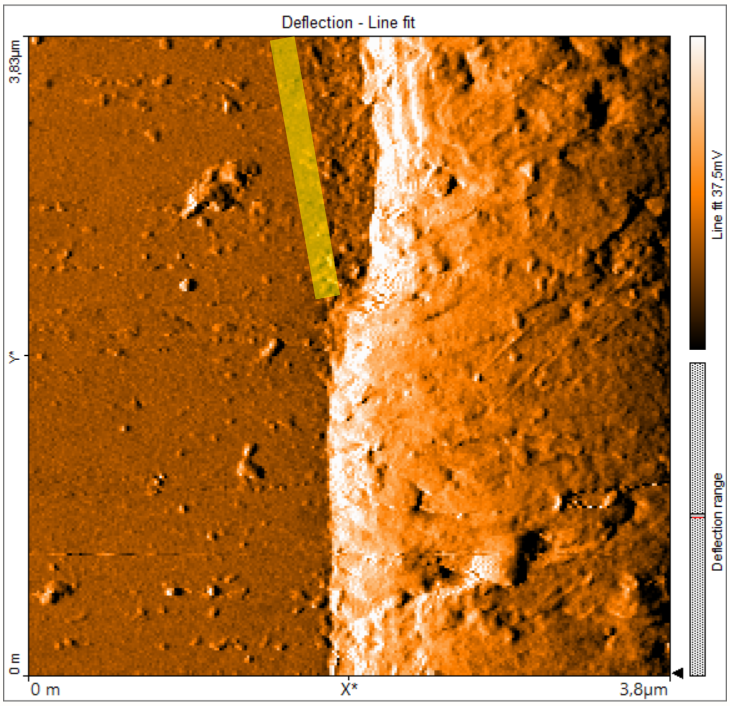


Figure 33. AFM deflection image of the contact and GaN-AlGaN stack (not visually identifiable) highlighted in yellow.

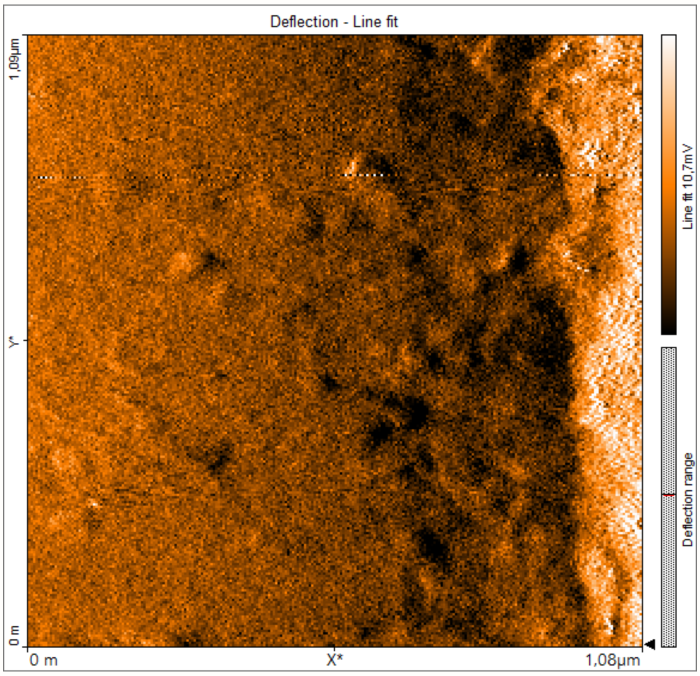


Figure 34. AFM deflection image at the limiting resolution of the Student AFM. The Ti/Al contact can clearly be distinguished from the GaN layer, but the AlGaN layer cannot be seen.

At the limiting resolution of the Student AFM the image does not improve significantly, as can be seen in figure 34. The AlGaN layer cannot be distinguished, nor can it be definitively determined whether metal penetrates the layer or not.

5.2 Bruker Icon AFM

The Icon AFM is significantly more precise than the Student AFM. As can be seen in the deflection image in figure 35. The region around the contact edge can be imaged perfectly. Naturally the same limitations apply as with the Student AFM: not much can be said about the electronic properties of this sample. However, the current sensing option of this AFM can provide such information.

The current sensing AFM images provided by this AFM were made without a lock-in amplifier. Because of this the original images look something like the image in figure 36. This is of course not very usable and the images have to be filtered manually using a Fourier transform to filter the original frequencies of the input current.

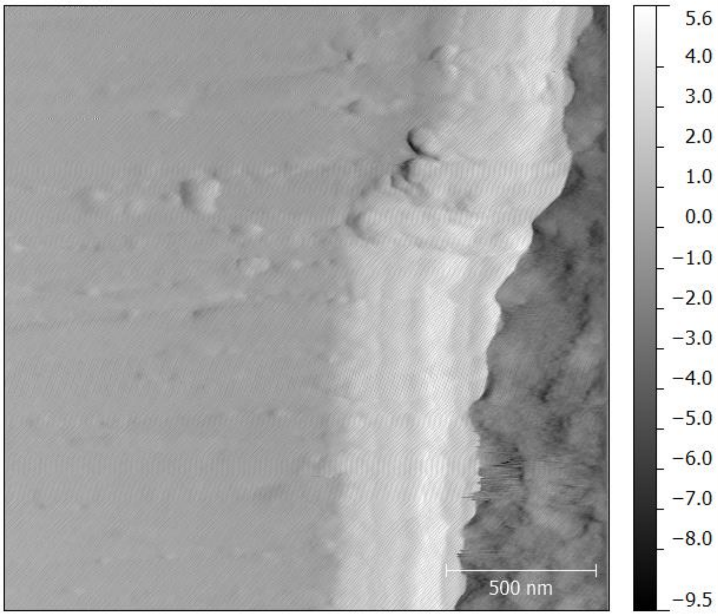


Figure 35. AFM deflection image of the contact edge. The image is of a much higher quality than that of the student AFM. Scale in arbitrary units.

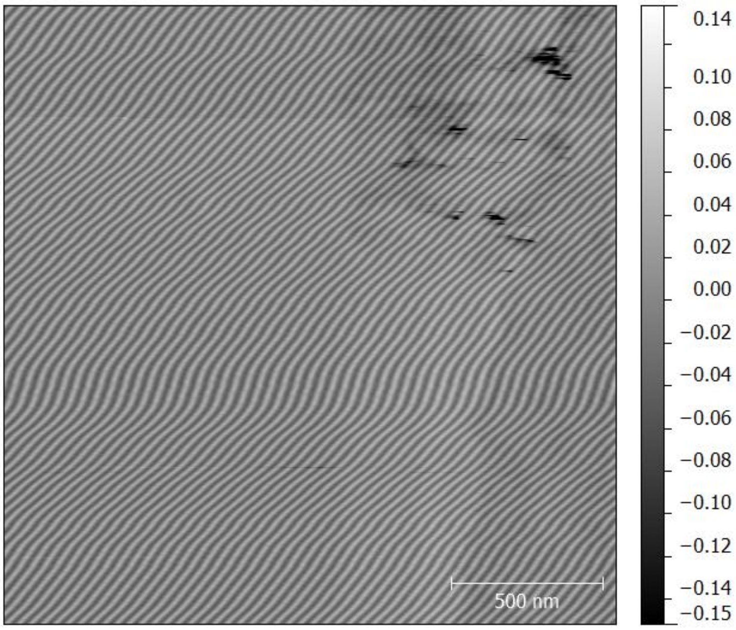


Figure 36. Unfiltered I-sensing AFM image. A Fourier transform is needed to remove the ghost image of the input frequency of the current.

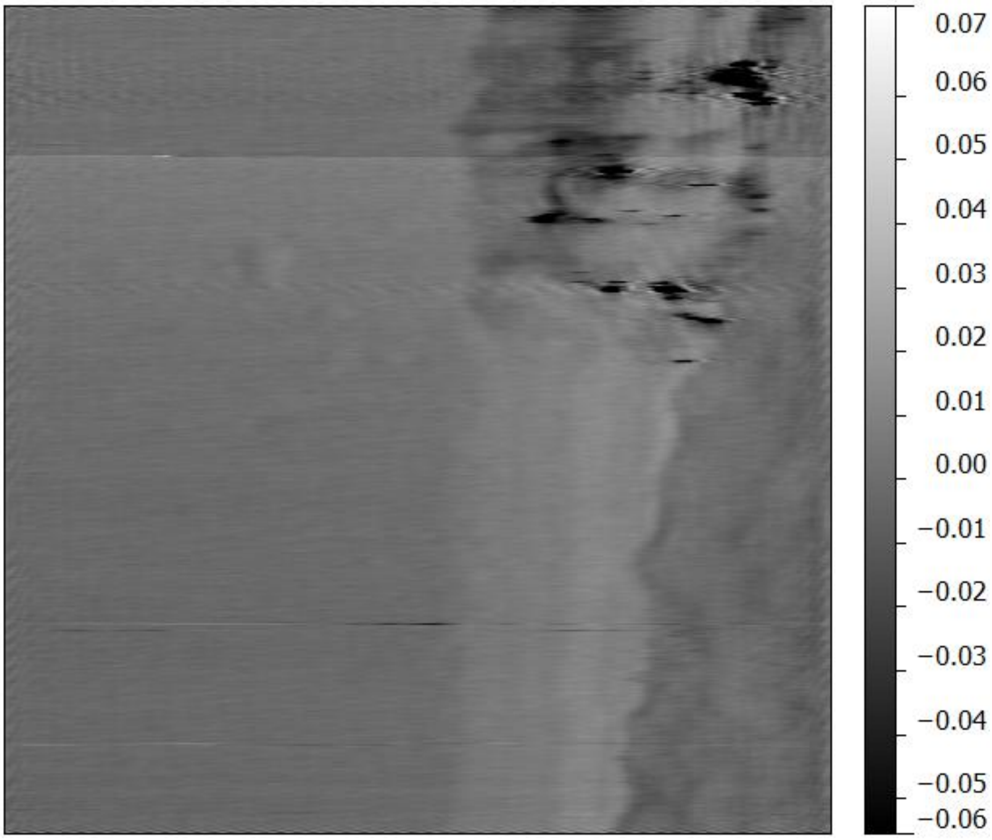


Figure 37. Filtered I-sensing AFM image of the contact region. Most of the region is insulating, apart from the darker spots on the Ti/Al contacts. Scale in arbitrary units.

The resulting filtered image is shown in figure 37. In this image a mapping of the conductivity is shown. Dark regions are metallic, while lighter regions are resistive. The supreme right part of the image is the glue, which is known to be insulating from the SEM measurements (next paragraph), and thus gives a gauge to the conductivity of the rest of the sample. The SiN layer, adjacent to the glue, shows to be insulating as well, but the remainder of the sample, the AlGaN, the GaN and the 2DEG all seem to be equally resistive. Only on top of the metal there are a few dark metallic spots.

The spots continue well over 500 nm into the sample, meaning metal is found inside the crystalline GaN. This could be evidence that metal does penetrate the AlGaN layer.

However, more importantly, it is found that the entire sample is far too resistive for the materials it should be made of. This, in all likelihood, means the sample has an oxide layer on top of it, obscuring the view of the cross-section.

5.3 SEM

As mentioned before, during the growing of the additional contacts using FIB, some images were made using the built-in SEM of the former device. These images showed a not-yet-seen detail of the electronic structure of the region of interest, as well as a few interesting unexplained phenomena. Therefore, the images are discussed here.

In figure 38, one of the test contacts is shown, but not the structure, rather its surroundings area is of interest. The sample is rotated 90 degrees in comparison with the Student AFM measurements, with the glass now top-side and the wafer at the bottom. What one can see is the differentiation between the Silicon, GaN buffer layer, and Crystalline GaN. It is also possible to distinguish the AlGaN layer and the Ti contacts, because they are darker and brighter respectively than their surroundings. On the bottom side of the AlGaN, another thin bright layer is visible. This is the 2DEG, which, for the electrons used to make this image, behaves similarly to a metal. What is not visible in this image, but did show in the live view was charge buildup in the glue (as it is an insulator) but also in the GaN and on the titanium.

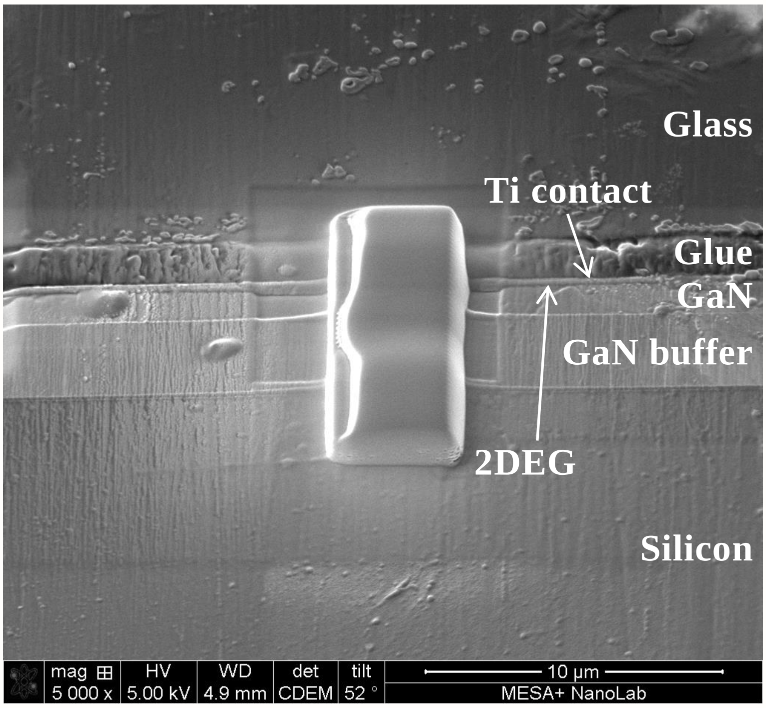


Figure 38. SEM image of a test-contact pad. What is interesting is the well-defined differentiation between the different layers in the stack. The AlGaN layer can be distinguished here as well, as can be the Ti and the 2DEG.

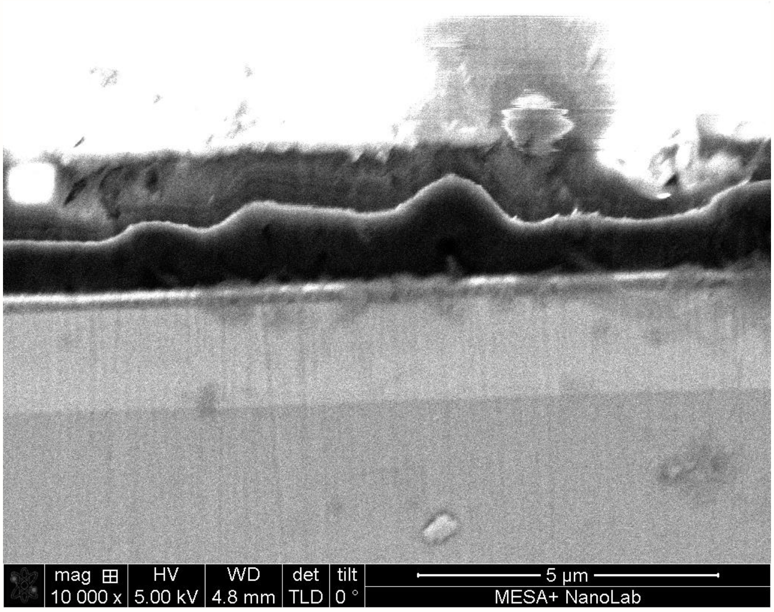


Figure 39. SEM image of charge buildup in the glue. The GaN layer can be seen as a brighter layer than the GaN buffer at the bottom of the image. There are also darker stains on top of the Ti contacts.

It was tried to capture a longer exposure to visualise the charging. the resulting image is given in figure 39. It is clear that there is charge buildup in the glue. But in this long exposure various spots can be seen on the titanium as well. As these spots are not shown with shorter exposures, it cannot be dirt, but it is unclear whether this is a side effect of some electronic property, or another form of charge build up in a material covering the contacts (for instance, an oxide layer). It is also worthy of note that in this longer exposure shows an even greater difference in luminosity between the crystalline GaN and the buffer layer.

A final image was taken from another interesting region. The region in question shows a smooth flat surface in the Student AFM, but in the SEM something else is observed. Figure 40 shows the silicon substrate (bottom) and the crystalline GaN (top), as it is across the wafer. However, the GaN buffer layer in between is not smooth at all. It has various densities and resistivities, and shows a curious mountainous pattern at the interface with the silicon. Again, it is unclear why this layer is showing up in this way under an SEM.

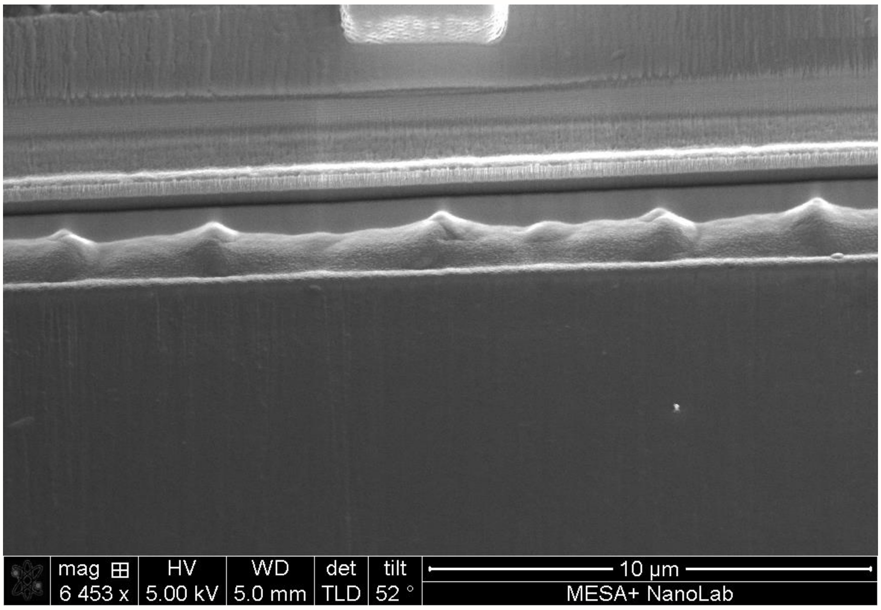


Figure 40. SEM image of the GaN Buffer layer, that should be smooth and homogeneous (as it is in most places) but shows up here having various density and resistivity layers, as well as a strange pattern at the Si interface.

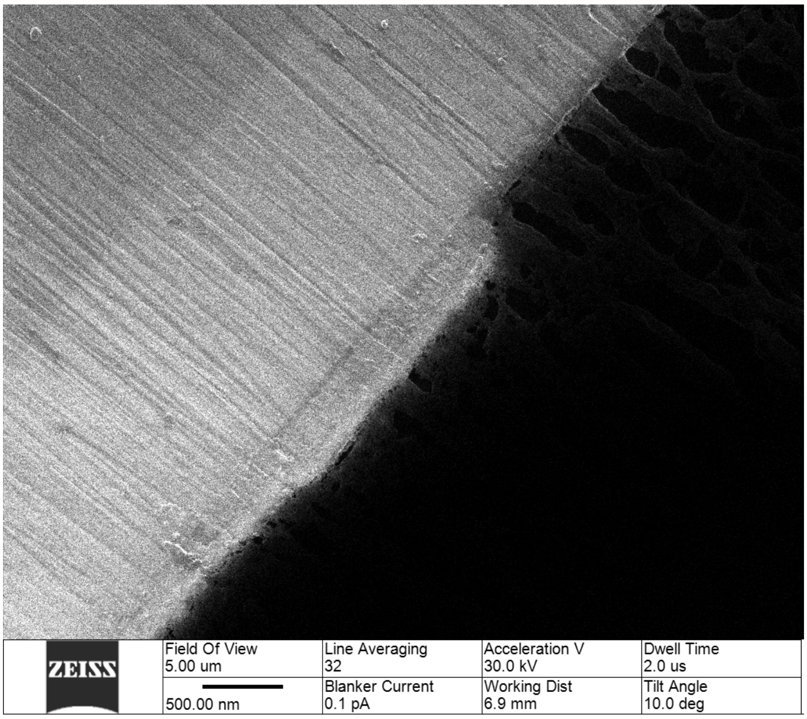


Figure 41. HIM image of the contact region. The distinction between the crystalline GaN and the buffer layer is again quite clear, the contacts are also clearly visible, as is the AlGaN layer.

5.4 HIM

Using the HIM to image the regions that were imaged with the SEM could give more detail because of the slightly higher resolution of the HIM over the SEM. In figure 41 a first view of the sample with the HIM is shown. The images are again rotated; the glue is the dark region in the bottom right corner, while the sample goes up to the top left. What is interesting to see, is that the HIM shows much more detail at the edge with the glue. This is partly due to the higher resolution of the HIM, but also because the glue has been drying for several months, and has receded slightly into the crevice. The contact is slightly thicker at its edges, due to the overlapping SiN/SiO_2 layer.

Figure 42 shows a zoomed in view of another contact. Here we can clearly see the contact and the SiN/SiO_2 layer overlapping with the contacts. There is also a layer of 20 +/-5 nm wide that should be the AlGaN layer. The 2DEG does not light up like it did in the SEM because the helium ions cannot register it as a metal. Neither is the 3 nm GaN toplayer visible, as its size is near the limit of the resolution of the HIM. Although the 0.02 μm polished sample is used, the grooves still distort a lot of the layers in this image. Because of this it is difficult to definitively determine whether this image shows a continuous AlGaN layer, or one that is disrupted by other materials. It is proposed that sputtering the sample after polishing could reduce this problem and restore the visibility of the stack.

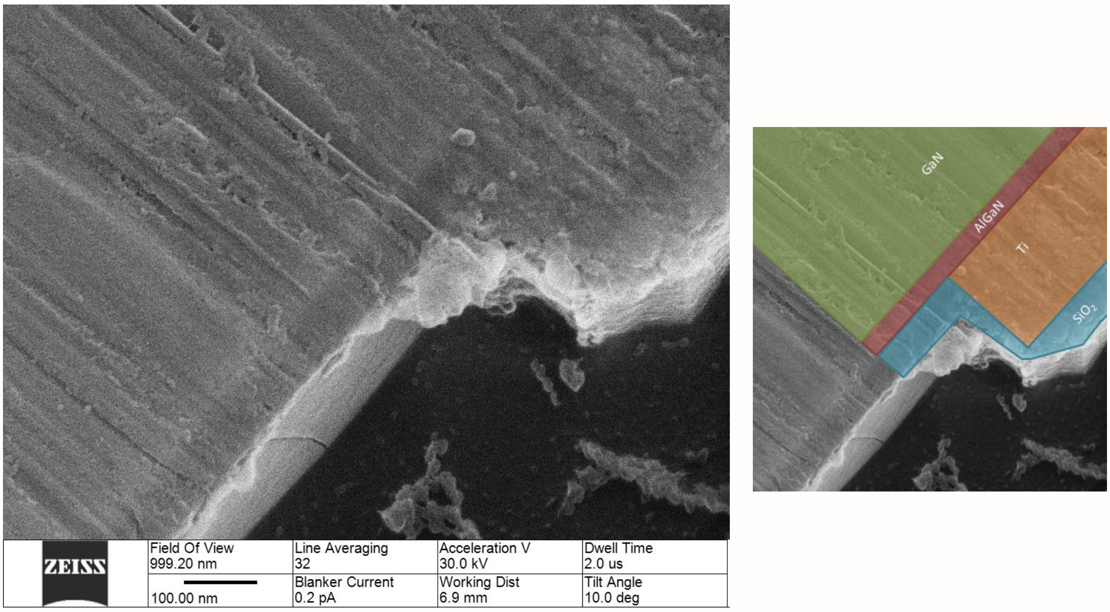


Figure 42. HIM image of a contact edge showing many different layers in the stack. The 2DEG and the 3 nm GaN toplayer are not visible. Explanatory image on the right.

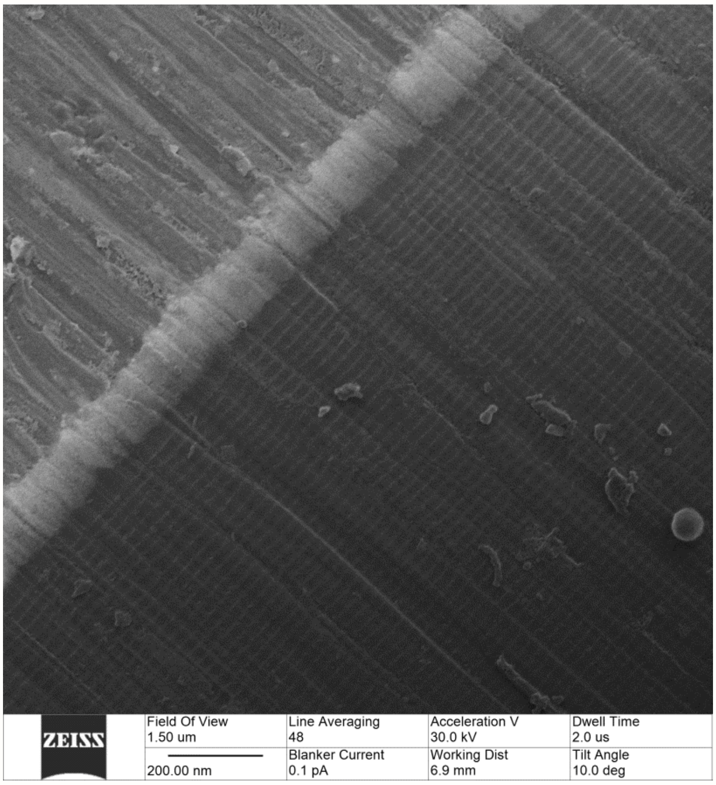


Figure 43. HIM image of the interface between the Si (top left) and the GaN buffer layer (bottom right).

A final image that needs to be examined is the buffer stack near the strange region seen in the SEM. Figure 43 shows a zoomed in part of this region. The top left corner is Si, while the buffer stack works its way down to the bottom right. This is the first image to show the actual stack of layers, meant to slowly ease the crystal structure from the silicon structure to the crystalline GaN, in detail: The different epitaxial layers can be seen quite clearly. But the strange artefacts that were seen in the SEM are not visible with this HIM measurement. The aforementioned artefacts must therefore be electron related, given they only show up when electrons are used for imaging. It is still unclear, though, what is causing these ghost images.

It is also interesting to note that the difference in hardness between Si and GaN can clearly be seen in this image, as the silicon contains significantly more grooves than the GaN, due to the fact the silica dug deeper into the silicon than it did in the latter material.

5.5 GaN sample in UHV STM

Given the various unexplained results for the patterned samples discussed before, it was put forward to do some additional measurements on unpatternl GaN samples. These are silicon wafers with a GaN-AlGaN-GaN system on them, but no contacts have been grown. Measurements were made on the top of the substrate, instead of a cross-section, as the behaviour of the GaN was of interest in the case.

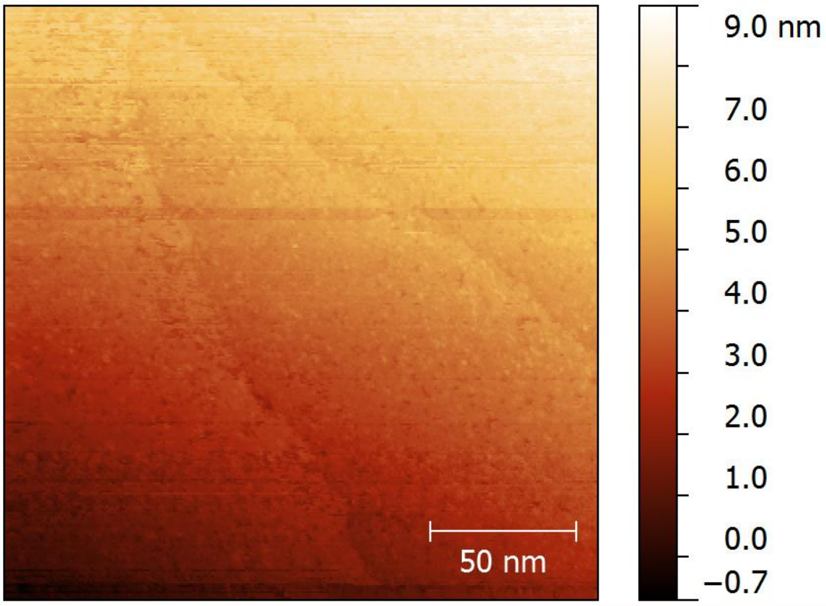


Figure 44. STM image of steps on a GaN-on-Si wafer after annealing.

Measurements were made in the UHV STM of the PIN group. On initial trials the sample proved to be non-conducting. This was thought to be caused by oxides on the surface, as it had been exposed to atmosphere. Therefore the sample was annealed to a temperature of 800 K. The first image after annealing is shown in figure 44. As can be seen, the sample still has various steps, despite annealing at such a high temperature. Also, a strange pattern of dots is shown on the surface. When zooming in and measuring at higher bias voltage these spots become more distinct, as is shown in figure 45. Since the sample has been annealed, the surfaces of the different steps are expected to be completely level. It could be possible that when annealing the GaN layer interacts with the released oxygen from the oxide layer, as it is expelled from the surface, as is seen in silicon. It remains unclear, though, if this is really the phenomenon that causes them.

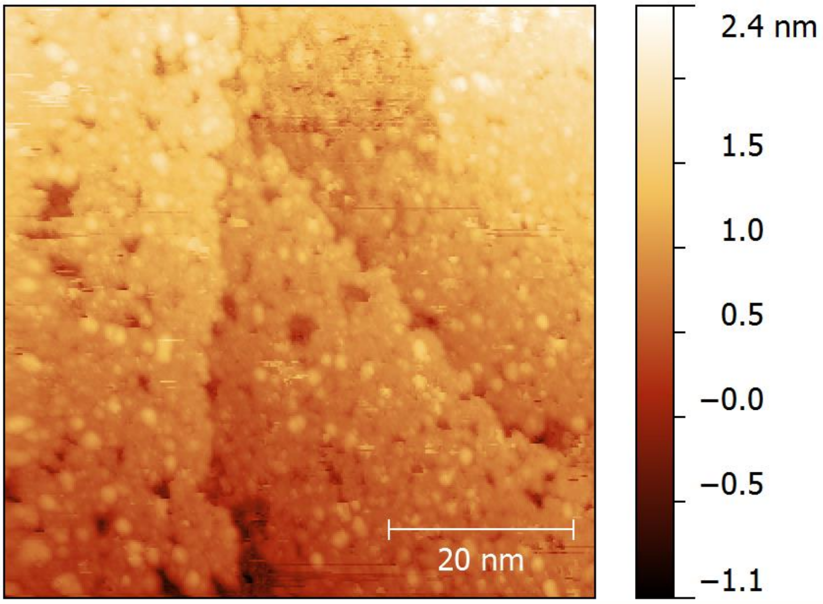


Figure 45. STM image zoomed in on the unexplained spot-pattern on the GaN surface.

On several places on the sample, contamination is found in the form of large regions of the surface varying in height with the rest of the surface. These regions appear to be roughly 400 nm in diameter and appear quite frequently across the sample (i.e. every ~1µm). One of these regions is shown in figure 46. It is uncertain what these regions are: A few possibilities are put forward.

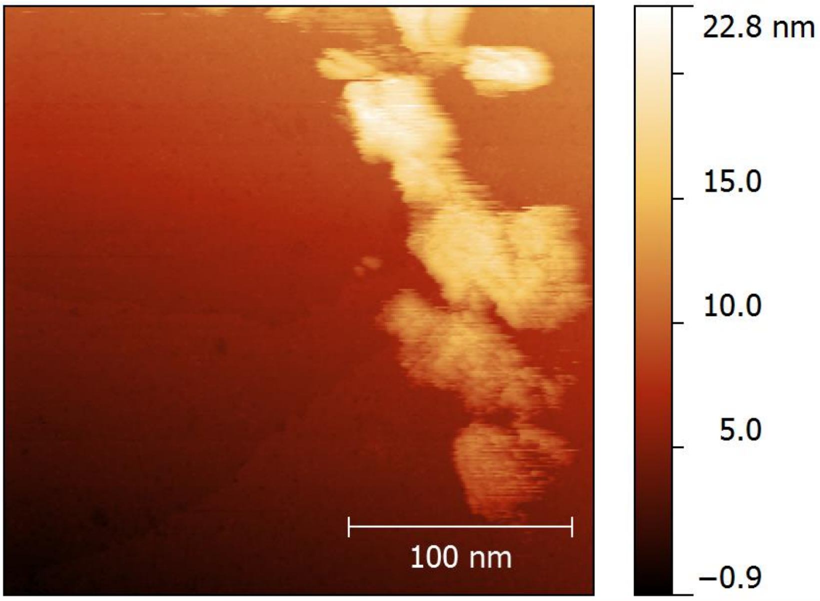


Figure 46. STM image of contamination on the surface of the GaN.

A first possibility is that the regions are contamination from the chamber. In a Ultra High Vacuum system, such as this, this is very unlikely though. STS measurements have been done on both the surface of the GaN as well as these regions. The results of this are shown in figure 47.

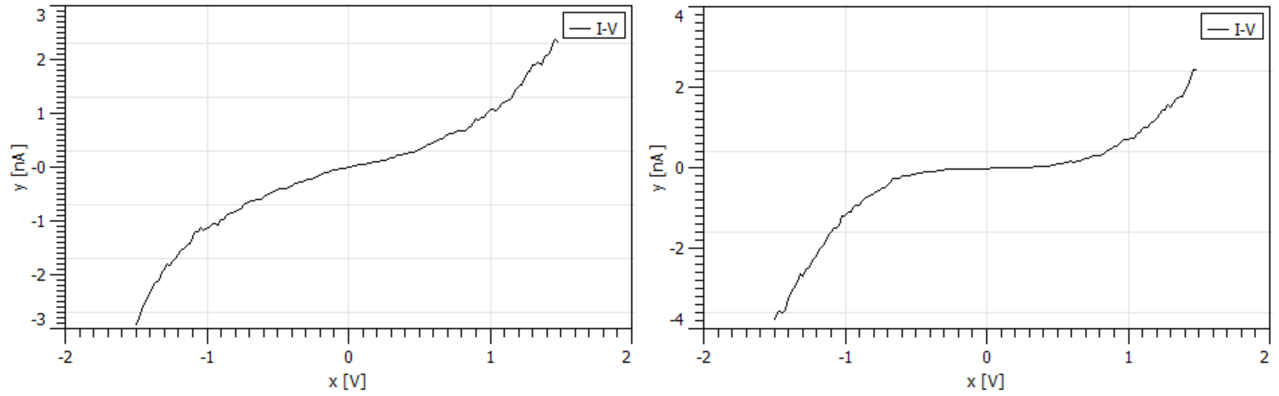


Figure 47. I-V curves from the STS measurements on the contaminants (left) and the GaN surface (right).

It should be noted that these graphs do not allow one to determine the bandgap of the material. The explanation for this is that the IV curve starts at a well-defined setpoint, here 1.4 V and a tunnel current of 2.3 nA. For GaN this is within the band gap of the material(!) meaning that the first part of the curve, from the setpoint onwards, does not represent the properties of the material, but rather the system lapsing to the correct values. In the PIN group it is preferred to take a voltage set point in the band gap rather than in the conduction or valence band because (1) the current sensitivity near the Fermi level is much higher and (2) a too high set point voltage might lead electric field induced changes.

For the sake of these measurements, however, it is not necessary to correctly determine the bandgap. Rather, the behaviour of the material around zero bias (determining if the material is semiconducting or metallic see paragraph 2.1.1.4) is of importance.

As can be seen in these graphs, the surface (right graph) shows the expected semiconducting behaviour. The ones measured on the contamination regions on the surface, though, show a metallic behaviour, having a nonzero slope around zero bias. It is unlikely that metallic material would have been deposited on the sample during preparation or measurement in the UHV system, which suggests that, if these are surface contaminants, the material was already on the sample, likely deposited there during fabrication. Temperatures during the fabrication process get up to 1100 K, meaning metallic contamination could be a possible cause for these regions.

Since the regions measure as metallic, another possibility is depletion of nitrogen from the regions in question. This would result in spots of pure gallium, which would show as metallic. However, it is uncertain why these nitrogen-depleted zones would form.

It is also feasible that, in the annealing step, the temperature is not sufficient to expel the oxide layer from the surface, but merely the oxygen from this layer. If this would be the case, pure gallium would be left on the surface, as a remainder from the oxide.

These metallic spots on the surface of the GaN substrate are unprecedented in any previous research. It is clear that the GaN substrate requires further research, in order to determine its state and behaviour when transistors are grown on them.

Chapter 6.

Conclusions and Recommendations

AFM and current sensing AFM images were made of the cross-section of a GaN-AlGaN HEMT as to make a conclusive judgement what caused the resistivity of the contacts to decline when annealing. Three hypotheses were put forward: metal spikes penetrating the AlGaN layer and contacting the 2DEG directly; N-vacancies forming in the AlGaN layer acting as a donor-layer for electrons from the contact to the 2DEG; or TiN formation, reducing the Schottky barrier height between the titanium and the AlGaN.

Neither of the three hypotheses can be discarded or proven definitively, since measurements done did not prove fruitful due to oxidation of the cross-section. Metal compounds were found, though, well beyond the AlGaN layer into the GaN, suggesting the metal-spike hypothesis to be at least part of the answer.

Further measurements were done using SEM and HIM, providing a more detailed image of the electronic structure of the HEMT. These showed various interesting phenomena, such as charge buildup in the contact region as well as strange artefacts in the GaN buffer layer. Again conclusive results could not be given as the preparation of the sample distorts the surface of the cross-section too much for any details to remain.

Additional measurements with STM on a unpatterned GaN substrate show spots of varying height on the surface as well as clear steps, despite annealing at high temperatures. STS measurements show some regions to be semiconducting while others are behaving like a metal. This is unprecedented in research of this substrate.

It is unclear where this behaviour arises from; however, it has significant consequences for the way the material could behave in any of its applications. Therefore it is recommended that further measurements be done on the system as to answer where the various phenomena found in this research arise from. More STS measurements should be done both on the surface as well as the cross-section of unpatterned GaN and structured samples. SEM or HIM measurements could prove fruitful as well, as these might give an insight in the anomalous behaviour of the GaN buffer layer. Current sensing AFM within a vacuum system could also prove fruitful, but to do this in the PIN research group, the RHK AFM needs to be operational first.

On the subject of sample fabrication, it should be noted that, when making cross-sections, one should prefer to do this in situ at low vacuum to prevent any oxides forming on the cross-section. This will inevitably mean cleaving such samples, as vacuum saws are not readily available, meaning a procedure needs to be devised to do this cleanly and straight, contrary to the samples made in this assignment. Furthermore, a way should be devised to create a conducting sample holder in situ, where the cleaved sample can be put on after it is finished.

If sawing maintains to be the preferred method, samples should be polished using a grade of 0.02 µm or smaller, and should be put in vacuum as soon as possible. It should also be considered to sputter the sample to remove the top layer of material damaged by the polishing step.

Chapter 7.

Acknowledgements

For the readers that know me, it should come as no surprise that I would like to use the opportunity of finishing this thesis, and with it my master's programme, to also look back on all the other things that contributed to my education at the University of Twente. To anyone else, I like to make clear that my time here started out as just the programme of Applied Physics, but ended up being a wild quest for life experience, that has been my reason to get up in the morning, every single day I spend here.

In those years I had fun with many of my peers, doing what we all loved to do: physics. But I also had fun doing many of the other things that this university offers, colloquially called "activism". Through it, I have made many friends, and oh so many memories. For the most part, this relates to study association Arago, which, as an institution, I would like to thank dearly, for giving me so many opportunities and experiences; despite ever decreasing support from the university, many more people should be allowed to do what I did being there. It truly is an enrichment to any that will.

Notes like these are too short to thank all the people I worked with, though thankful I am. Therefore, I'd like to thank some people in particular. To start with, my fellow board members at Arago: I think I sum it up quite nicely by saying I learned how different people can be, and I feel I can celebrate the fact we succeeded together, still, at our frequent dinners together; long may they last. I also like to thank the Mātauranga committee, with whom I got to organise a trip to the other side of the planet. I will not forget all the wonderful things we achieved and did there, or the crazy stories about all that went wrong, and I hope to see you all, April 15th 2021 at Mr. Bun in Wellington. I thank many of the honorary members and members of merit of Arago, through which I got to know them and above all, the many things us physicists have in common. It gave me more inspiration for my own future than any career program could have achieved. My thoughts go out to the late Mrs. de Winter, whose love for all of her children I will never forget, and Henk and Thijs, for their effort to continue her legacy to this day; I'll hope to see you at the next BBQ! I thank my fellow editors of the KLAT, with whom, aside from all the nonsense, I learned better writing than any Focus editor could ever hope to learn. And finally, on behalf of my mother I thank the people of Objectief, not so much for our work together, but for making a picture of me too sometimes.

In the case of this research itself, thanks has to go out to Harold Zandvliet, not just as the chairman of my graduation committee, but as the auctor intellectualis of this assignment, a mentor and even a hands-on experimentalist in the labs. I'd also like to thank the other members of my graduation committee: Herbert Wormeester, with whom I spent weeks at, on top of, under and inside the RHK AFM, trying to fix it and thinking of alternative ways to get our measurements done; Dirk Gravesteijn, from NXP, who not only provided the samples but was also a frequent discussion partner concerning the subject; and Marc Dhallé, the external member of the committee, a professor on this university who has been involved in my progress since I started here, and who, I feel, got to know me quite well over the years. Furthermore, all of this research would not have been possible without the contributions of Marcin Hajlasz, from NXP. I'd like to especially thank Martin Siekman, Gregor Hlawacek, Peter Schon, Lije Zhang, Rico Keim and Henk van Wolferen for their expertise, without which many of the preparations and measurements would not have been possible. And finally I thank all members of the PIN group for having me in their midst the last year.

Speaking to my family and friends, far away as some of you may be, I'd like to say, I would not have been able to finish this work, this study programme, or this part of my life without the joy you gave me. I hope to thank each of you in person in the way I know best – cooking you a far too elaborate dinner – somewhere in the near future.

Having had most of you, I dedicate this last paragraph to a special person, without whom I would never have been who I am now. She gave me my interest in science, the curiosity to want to learn new things, and the means to keep exploring. She never ceases to support me in every way she can, even when she couldn't. I do not thank her enough for everything she does, and feel I never will be able to do so adequately. More so am I proud to be calling her my mother.

Explorer
Always taking a new path,
Destinations I left behind me.
While I wander, I wonder:
What will be next?

I am, and will always be, in your debt.

Bibliography

[1] Y. Arakawa, “"Progress in GaN-based quantum dots for optoelectronics applications",” *IEEE Journal of Selected Topics in Quantum Electronics 8 (4),* p. 823–832, 2002.

[2] F. M. I. A. L. Wang, “Differences in the reaction kinetics and contact formation mechanisms of annealed Ti/Al/Mo/Au,” *Jour. of App. Phys.,* vol. 101, p. 013702, 2007.

[3] J. Millman, Electronic devices and circuits, ISBN 0-07-085505-6: McGraw-Hill International, 1985.

[4] K. R. P. Park, “Ohmic contact formation in AlGaN/GaN heterostructure via graphene insertion,” *App. Phys. Lett.,* vol. 102, p. 153501, 2013.

[5] P. H. J. E. S. Miller, “Ohmic contacts to n-type GaN,” *Jour. Elect. Mat.,* no. 25, p. 1709, 1996.

[6] F. M. I. A. L. Wang, “Dislocation-induced nonuniform interfacial reactions of Ti/Al/Mo/Au ohmic contacts on AlGaN/GaN Heterostructure,” *App. Phys. Lett.,* vol. 87, p. 141915, 2005.

[7] S. M. R. G. Qian W., “Structural defects and their relationship to nucleation of GaN thin films. in III-Nitride, SiC, and Diamond Materials for Electronic Devices.,” *Material Research Society Symposium Proceedings,* vol. 423, pp. 475-486, 1996.

[8] H. R. G. Binnig, “Scanning Tunneling Microscopy - From birth to adolescence,” *Rev. Mod. Phys.,* vol. 59, p. 615, 1987.

[9] J. Chen, Introduction to Scanning Tunneling Microscopy, ISBN 0-19-507150-6: Oxford University Press., 1993.

[10] J. Simmons, “Generalized Formula for the Electric Tunneling Effect between Similar Electrodes Seperated by a thin Insulating Film,” *J. Appl. Phys.,* vol. 34, p. 1793, 1963.

[11] M. Schmid, Artist, *Schematic diagram of a scanning tunnelling microscope.* [Art]. Wikimedia commons, 2005.

[12] F. Q. C. G. G. Binnig, “Atomic Force Microscope,” *App. Phys. Lett,* vol. 56, p. 930, 1986.

[13] P. W. P. Eaton, Atomic Force Microscopy, ISBN: 978-0199570454: Oxford University Press., 2010.

[14] O. Twisp, Artist, [Art]. Wikimedia commons, 2009.

[15] D. H. K. Lang, “Conducting atomic force microscopy for nanoscale tunnel barrier characterization,” *Rev. of Sci. Inst.,* vol. 75, p. 2726, 2004.

[16] J. Burgess, Under the Microscope: A Hidden World Revealed, ISBN 0521399408: CUP, 1987.

[17] M. Tolo, Artist, *Schema MEB.* [Art]. Wikimedia Commons, 2010.

[18] Zeiss, “Zeiss Orion Helium Ion Microscope Technical Data,” 23 06 2014. [Online]. Available: http://www.smt.zeiss.com/C1256E4600307C70/EmbedTitelIntern/ORIONEssentialSpecificationPDF/$File/oriondata.pdf. [Accessed 23 06 2014].

[19] M. U. L. S. J. Orloff, High Resolution Focused Ion Beams: FIB and Its Applications, ISBN 0-306-47350-X: Springer Press., 2003.

[20] Kriegor27, Artist, *Schematic and real FIB.* [Art]. Wikimedia Commons, 2009.

[21] Nanosurf, Nanosurf NaioSTM manual, 2013.

[22] Nanosurf, Nanosurf NaioAFM manual, 2013.

[23] Bruker, “Bruker Icon AFM technical data,” 23 06 2014. [Online]. Available: http://www.bruker.com/products/surface-analysis/atomic-force-microscopy/dimension-icon/learn-more.html. [Accessed 23 06 2014].

Appendix A.

Basics of semiconductor physics

What is a semiconductor?

For most people the difference between conducting and non-conducting materials is evident[15]. Electric conduction takes place through metals fairly easily, whereas insulators do not conduct electricity, and convert the energy into other forms, like heat or light. From a physical point of view, the reason why a material is either conductive or not is rather more complicated.

In fact, every material in existence, from metal to plastic and from rocks to grapefruits, contains an unimaginable number of electrons. Whether or not these materials also conduct electricity is dependent on whether or not these electrons are also able to move. In general they are bound to the atoms that make up the material, and they cannot just leave. Each element has a specific amount of electrons around it's atoms (called the atomic number), and each of those electrons has a specific amount of energy that binds it to said atom. The binding energy of the atom, or the outermost electrons of an atom, is called the Fermi energy (E_F)[16].

---

[15] That is to say, most people have a feeling for what materials to use for either conduction or insulation; a copper wire and a rubber boot for instance. The nomenclature of the scientific world, however, can be far from 'evident' for much of the general public, especially when scientists invariably keep claiming it to be so. On behalf of all of us, I humbly apologize.

[16] One could compare this to a harbour full of boats; some are very close to the shore and don't feel anything of the wind and current outside the harbour, while for those farther out on the end of the pier one can already hear the wind howl through their rigging. The boat nearest to the shore won't easily move out of the harbour, it is said to have a high binding energy. Similarly, the one on the end of the pier has a low binding energy (it is easily blown out of the harbour).

What happens when an electric field is applied to the material, for instance by putting a potential difference between two of its extremities? All the electrons in the material feel this as a pull away from their atom, most of the electrons don't get enough energy to leave the atom, but since the electrons near E_F are only slightly bound, they are pulled free, and they can roam the material without getting stuck to other atoms[17]. They can even flow in and out of the material. These electrons form the current we see in conducting materials, and they are what makes a material conducting or not.

For all general purposes then, electrons with an energy below E_F, stuck to the atoms, are said to be in the 'valence band', and all the electrons above E_F, conducting potential currents, are said to be in the 'conduction band'. Usually this is visualised by the schematic representation shown in figure A - 1.

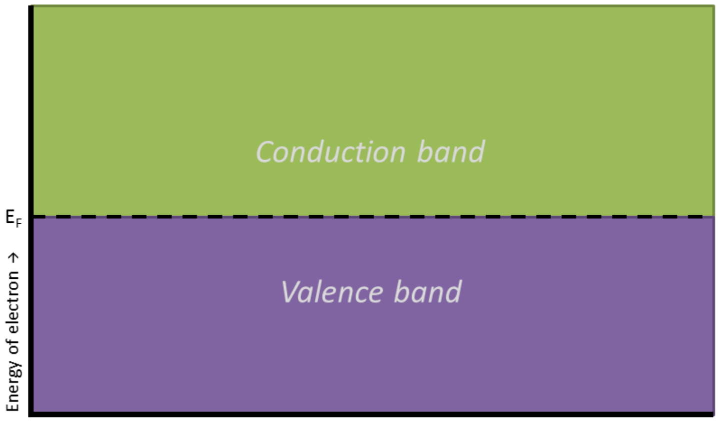


Figure A - 1.Visual representation of "bands" of electron energy in a material.

But what about insulators? A problem arises when quantum physics[18] is taken into account: according to quantum physics, the 'bands' described earlier cannot just be filled with electrons that are above or below E_F; one cannot put any electrons within the supposed energy range in them. Instead they consist of a limited amount of discrete states with specific energies, where in each state only one electron can

---

[17] A current of electricity could be compared to a storm in the harbour: all boats will feel the wind, but the ones furthest to the outside will feel it the most. They can even cut loose and leave the harbour altogether.

[18] Although an explanation of quantum physics could be useful, no physicist in the history of the world has succeeded in making one that is both brief, as well as…explanatory. I decided it wasn't necessary to try my luck here. Later on, however, I did need to explain some quantum mechanics for the further explanation of tunnelling; hence one can now read Appendix B.

reside. In other words, an electron needs to get a specific energy from an electric field to get into a state in the conduction band, if a different amount of energy is added, the electron will simply emit it again as heat or light[19].

This is where the physical difference between a metal and an insulator is made. Metals are materials where the amount of energy levels around E_F is very high. In other words, electrons just below E_F have a lot of states to go to just above E_F, which makes it easy for electrons to start conducting. Insulators, on the other hand, have no states close to E_F; electrons will have to get an enormous amount of energy from the field in order to move away from the atom. So much energy, in fact, that in daily life these materials never become conducting.

However, there is a third possibility: a small range of materials exists (silicon, germanium, etc.), which do not have energy levels very near to E_F. The distance between the end of the highest valence band levels and the start of the lowest conduction band levels, though, is small enough to overcome with a small potential (i.e. 1 to 10 Volts). These materials do not conduct, until a high enough potential is applied. Beyond that, however, they behave like a metal. Since these materials sometimes conduct and sometimes don't, they are applicably called semiconductors (figure A - 2).

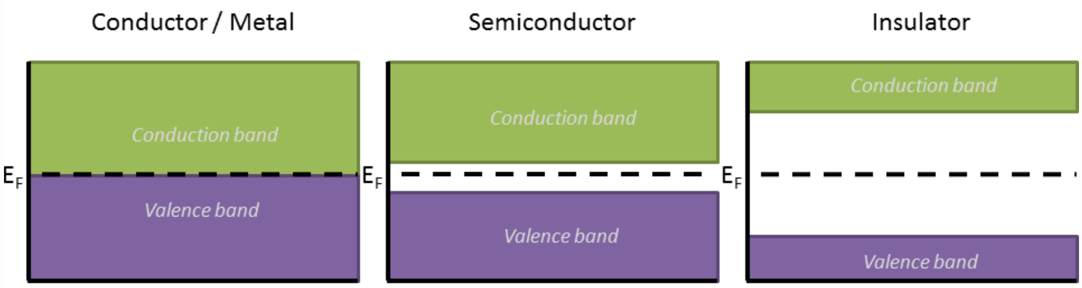


Figure A - 2. The different band(gap)s of a conductor, semiconductor and insulator.

---

[19] One could see this as if the ships in our harbour are only cast loose and sail to sea for specific wind speeds, for instance, 3 or 5 bft, but not 4 or 6 bft. It is here, I think, that this analogy with boats has become severely unrealistic.

The gap between the valence and conduction bands, equally aptly named “bandgap”, is usually designated by the energy needed to cross it in elementary units: electronvolts (eV). This is done because the number of electronvolts a gap is wide, is equal to the number of volts one needs to use to create an electric field strong enough to get electrons from the top of the valence band to jump to the bottom of the conduction band.

It is important to note, though, that the energy needed to cross this band need not come from an electric field; light can be used as well[20], or even heat can cause electrons to start conducting. In fact, there are many materials that officially could be called a semiconductor, but have such a small bandgap that merely being at room temperature[21] makes them conducting

One can imagine that scientists have thought up many inventive ways of using these materials’ properties. Indeed, the semiconductor industry is huge and brought forth, among other things, radio amplification (FM radio, radar), the microchip (computers), photovoltaics (solar cells) and Light Emitting Diodes (LED’s),

Transistors

The most widely used application of semiconductors has always been the transistor. Basically, this is an electronic apparatus which acts like a switch: you can turn the flow of a large current on or off by applying a small current or not. The semiconductor comes in because it acts as the limiter: by applying a small current with a potential greater or equal to the size of the bandgap, the electrons in the semiconductor move to the conduction band where they can then be used to channel a large current through the semiconductor.

The most widely used version of the transistor is a so called Field Effect Transistor (FET). The field effect transistor, as the name implies, uses the electric field applied by a ‘gate’ to create a conductive channel in a semiconductor, which then conducts a current flowing from a ‘source’ to a ‘drain’. The advantage of this transistor over other designs is that it’s physical layout can be really small and can be put together

---

[20] Einstein even won his Nobel price, not for his famous relativity theory, but for his proof of the coupling between electrons and light, called the photoelectric effect.

[21] Room temperature, in physics, is a really, really warm place. Absolute zero – the temperature at which atoms and electrons do not have any energy left – is at -273.25 °C, So room temperature, at, say, 295 degrees above that, already gives a lot of energy to electrons in a material.

layer by layer, which even makes it possible to print or etch it. The only step needed towards building a microchip is downsizing each transistor, so as many as possible can be put on a small surface. Today's computer processors contain several billion transistors on a piece of silicon the size of a fingernail.

MOSFET's

In most cases (as in figure A - 3) a Metal-Oxide-Semiconductor FET (MOSFET) is used, because of its simplicity and durability. Here a potential is applied to a gate metal and an enhanced electric field is created by the insulating oxide layer in between the metal and the semiconductor. The field created then starts pulling charge carriers (i.e. free electrons) to the interface of the oxide and the semiconductor, creating a conducting channel from the source to the drain.

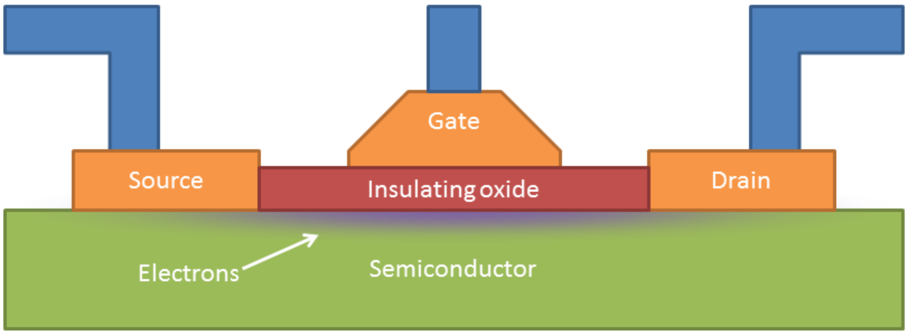


Figure A - 3. A *very* rudimentary Field Effect Transistor.

This structure, basically a planar capacitor with a semiconductor as one of the electrodes, changes the electronic structure of the semiconductor when a voltage is applied to the metal. The semiconductor in question is doped using impurity atoms with a different valence (amount of electrons) than the semiconductor atoms, freeing electrons or holes[22] for respectively negatively doped (n-type) and positively doped (p-type) material. These impurities are called donors and acceptors respectively. Doping a semiconductor effectively reduces the bandgap width, since the presence of

---

[22] The concept of a hole being a charge can sound curious. One can think of it as a line of cars parked in a street with one open space in the middle. If all cars to the left of the space move one space to the right one by one, this could be seen as a consecutive movement of the hole to the left. The same is the case for an atom having one electron less than its neighbours. If electrons from the neighbouring atoms start hopping around, one could also see this as a movement of the hole.

additional free charge could be seen as the addition of a donor/acceptor band inside the bandgap close to the conduction or valence band. This means that the semiconductor in essence becomes a metal.

In a MOSFET the semiconductor is doped in both ways, n-type under the source and drain, while the bulk is p-type[23], as is illustrated in figure 1. Effectively, a small potential could make the entire semiconductor conductive, however, because the free charge carriers are different for the different regions of doping (holes vs. electrons) a current cannot flow.

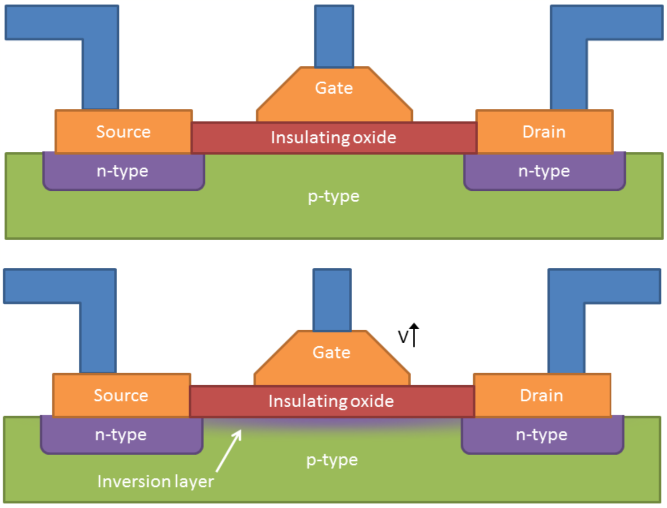


Figure A - 4. A MOSFET structure with a doped semionductor layer, both with zero potential (top) and nonzero potential (bottom).

Suppose another potential is put on the gate electrode. If this potential is positive, an electric field will drive the positively charged holes away from the interface, leaving a carrier-free region of negatively charged acceptor ions, called a depletion layer. The electric field also has an effect on the electrons in the semiconductor, namely increasing their potential energy, effectively lowering E_F towards the conduction band of the semiconductor. If the potential is high enough, the conduction band will cross E_F, which in turn creates available states for electrons in the semiconductor. a high concentration of negative charge carriers will need to form close to the semiconductor-oxide interface in order to compensate for the growing field; the inversion layer. This inversion layer needs to get these electrons from somewhere,

---

[23] Although these two can be inverted without a difference in operation, apart from a difference in the applicable potential, this is the most used version.

and in the MOSFET this source generally is the negatively doped source and drain terminals of the transistor. For a current to then flow through the inversion layer, only a small potential between these two terminals is needed, which is the raison d’etre of the transistor.

Appendix B.

Quantum Mechanics

To understand some of the physics involved in this thesis, a basic understanding of quantum mechanics is useful. Especially given the electronic transport mechanism it implies: tunnelling.

Particle-wave duality

Quantum mechanics treats matter as having both properties of waves as well as particles, meaning that everything in the universe can be described by both particle interactions as well as wave mechanics. The wavelength of a particle is given by the de Broglie formula:

$$\lambda = \frac{h}{p}$$

where λ and p are the wavelength and momentum of a particle, and h is a constant known as Planck's constant. In other words: the wavelength used to describe the physical behaviour using wave mechanics is related to the description in particle physics by the momentum of that particle, times a given constant. Momentum, for macroscopic matter is given by its mass times its speed, so the bigger its mass or speed, the smaller it's wavelength. The reason one doesn't see the world behave like waves and particles is the nature of the third part of the formula: Planck's constant.

$$h = 6{,}626 \cdot 10^{-34}\ m^2 kg/s$$

In other words: really really small. This means the momentum has to be really really small as well for the wavelength to have any significance. In general this means that quantum mechanics only becomes important at an atomic scale, for elementary particles like electrons, protons and neutrons, where the wavelength becomes larger than the size of the particle it relates to.

But how does something behave both as a particle and a wave at the same time? An experiment was thought up to determine this in the early 20th century. Imagine a screen that can detect the impact of a wave or particle (this can be as simple as a wall, for light for instance), in front of which is a plate with two slits in it. When a plane wave passes through the two slits scattering causes the waves to become spherical after the plate, and interference of these wave fronts causes a specific pattern to form on the screen. This interference pattern is seen for all sorts of waves, from sound waves to waves on the sea.

However, when a beam of particles (water, let’s say) passes through the slits, one would expect only two straight beams remain, and thus two straight lines on the screen, like someone shot two holes in the side of a barrel.

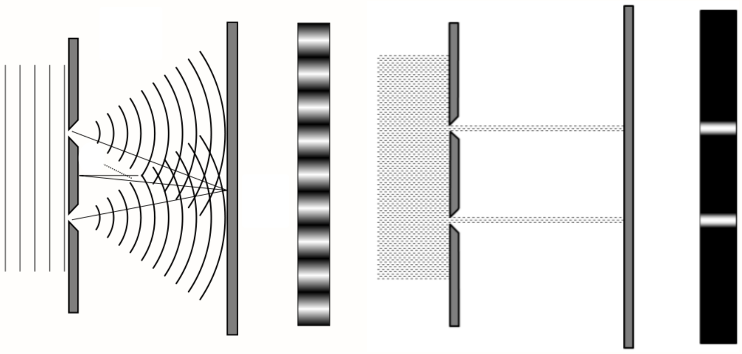


Figure B - 1. Waves (left) and particles (right) passing through a double slit, and their resultant detector patterns.

How can a particle behave like both phenomena at the same time? An experiment was done with a beam of electrons passing through the two slits. Electrons are so small that they truly behave both like a wave and a particle at the same time. The results are shown in figure B - 2. What can be seen is that individual electrons indeed leave a single dot (not a pattern) on the screen, ergo, they behave as particles. But they do not all end up at the same spot; instead they are scattered across the screen at, what seems like, random locations. When enough electrons have passed through the slits one can see that a distribution arises, that of an interference pattern of waves.

This experiment shows that, in quantum mechanics, particles have a specific location, but what the probability is to find this location to be at a certain point, is determined by the wave behaviour of the same phenomenon.

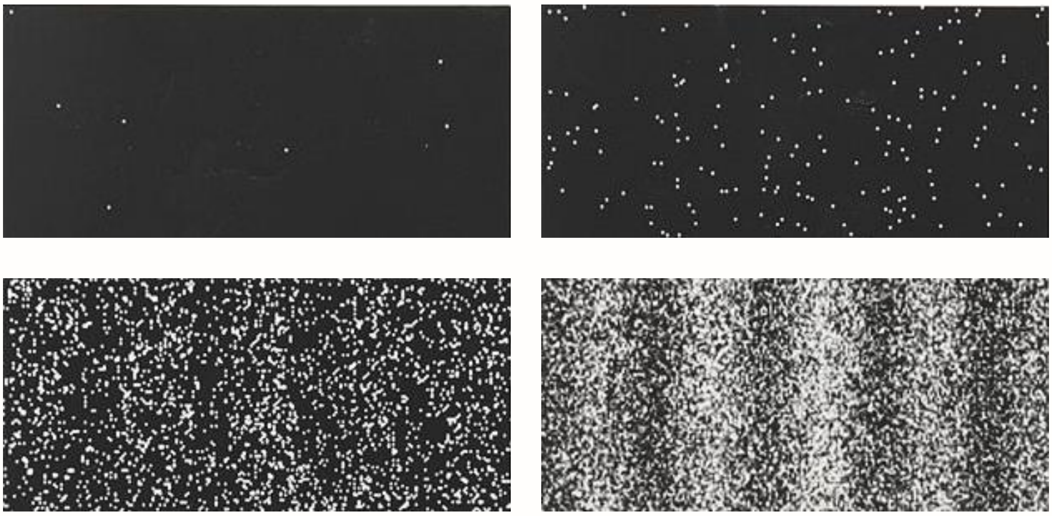


Figure B - 2. Double slit experiment with electrons at different times of exposure.

Uncertainty principle

From the derivation of the de Broglie formula, it can be mathematically proven that physics on an elementary particle level is far from exact. This has to do with the fact that a particle is no longer at a specific position when its wavelength derived from the de Broglie formula is larger than its own size.

Try and imagine you are looking for the position of, for instance, four balls in the yard. You could build a device that measures the height of the yard at various positions, and where you find a significant difference (a bump) with the rest of the yard, you know a ball must be there; their position is quite clearly defined.

This is no longer the case when the balls would have a wavelength larger than their own diameter (i.e. really small balls). Instead of being in a specific place, the balls are now waves of matter flowing over the yard, and you cannot be certain a ball is at a specific location; there is a high probability that they are at some places, and a very low one at other locations, but you will never know exactly. You could compare this, quite literally, with looking through a defocused lens: everything is fuzzy and you can’t see where the balls really are.

Like the electrons in the previous paragraph, you would find specific places for the balls if you were to measure the entire yard, but the coupling of location and momentum by the de Broglie formula makes it impossible to both know the exact speed and the exact location of any particle in the universe.

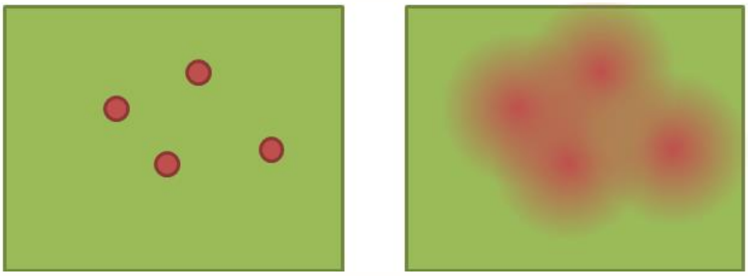


Figure B - 3. Balls in a yard behaving as particles (left) and waves (right).

In other words: you either know exactly where a particle is, but not how fast it is going, or you know the speed of a particle, but not where it's at. This principle, aptly called the uncertainty principle[24], is put forward by Heisenberg:

$$\Delta x \cdot \Delta p_x \geq \frac{\hbar}{2}$$

where ħ is the reduced planck constant or $h/2\pi$.

Schrödinger equation

Naturally, when confronted with the problem of describing a particle as a wave and vice versa, physicists found a way to convert their calculations from the one form of mechanics to the other. The equation which does this is called the Schrödinger equation:

$$-\frac{\hbar}{2m}\frac{\partial^2 \psi_n}{\partial z^2} + U(z)\psi_n(z) = E\psi_n$$

This equation can be solved for given energy and mass, to give a solution ψ_n in the form of a wave, or vice versa. With this, one can determine the wave function that fits an electron travelling through a material. But like the balls in the yard, the electrons location is a bit fuzzy; there is a small probability to find the electron outside of the material, visualised by the wave continuing across the material boundary in figure B - 4.

---

[24] It has taken quite some time for the physics community, which is based on *exact* science, to accept something as definitively inexact as the uncertainty principle. A famous opponent of quantum physics, Albert Einstein, once said: "Gott würfelt nicht" – God does not play dice – as to say it cannot be true that physics at an atomic level is dependent on probability and uncertainty.

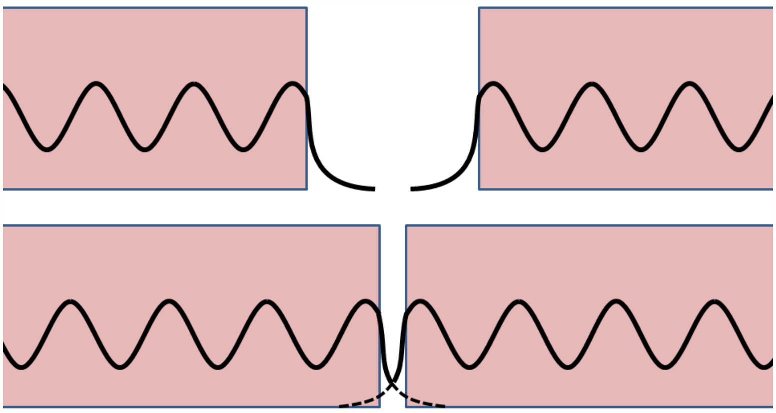


Figure B - 4.The concept of tunnelling: the wave function of an electron in two pieces of conducting material. If the two pieces are brought close enough to each other the wave function will have a nonzero value in both materials.

If there is another conducting material in the vicinity, the electron would have the same wave function if it was in that material, but again, outside of the material, the probability would quickly go to zero. If the two pieces are brought close together, however, both wave functions would overlap into the other material, as can be seen in the bottom part of figure B - 4[25].

As with the electrons and the double slit, the wave function exists on both sides of the gap between the two pieces of metal, but this exhibits itself by getting an electron with this wave function, sometimes on one side, sometimes on the other. This will effectively enable electrons to fly towards the gap from one side, and appear on the other without even truly having been inside the material[26]. This is the concept of tunnelling; the mathematical derivation and further implication is discussed in chapter 2.

---

[25] People with an understanding of quantum mechanics will notice the wave functions are lacking a pre-factor in this image, which is done for the sake of comprehension to the intended audience of this appendix. I implore you not to linger on this fact for too long.

[26] As if it went through a secret *tunnel*. Aha! Physicists are so clever in naming things.

UNIVERSITEIT TWENTE

PIN group

NXP Semiconductors