Growth, Physics, and Device Applications of InAs-based Nanowires

Doctoral Thesis

Linus Fröberg

Thesis supervisor:
Prof. Lars Samuelson

Faculty opponent:
Dr. habil. Walter Riess
IBM, Zurich Research Laboratory

LUND UNIVERSITY

Department of Physics
Lund, Sweden 2008

For Louise
Sjömannen ber inte om medvind, han lär sig segla.
Abstract

This thesis is based on three different projects: 1) the epitaxial growth of nanowires using chemical beam epitaxy, 2) the study of electron transport through quantum dots and multiple quantum dots in nanowires at low temperature, and 3) the development of wrap gated nanowire field effect transistors.

In the first part, a method of studying the diffusion of the source material on the substrate surface was developed. Nanowires were positioned in a triangular pattern on the substrate and, depending on the density of the nanowires in the arrays, the growth rate changes due to competition for available source material on the surface. A model was developed that could be fitted to experimental data using the diffusion length as one of the fitting parameters. The growth rate dependence on nanowire diameter was also studied and was satisfactorily explained by a model that takes both substrate diffusion and the Gibbs–Thomson effect into account. Nanowire heterostructures in the InAs/InP system were studied and the importance of seed particle alloying was demonstrated. The nanowires were grown from Au seed particles which alloy with indium, forming a Au–In seed particle. The composition of the particle is different during InAs and InP growth, and for each heterostructure interface the seed particle composition has to change. This affects the initial growth of InAs and InP segments.

By growing two thin segments of InP in an InAs nanowire, a quantum dot was formed between the InP tunnel barriers. At low temperature (4.2 K), the electron transport through these quantum dots is governed by Coulomb interactions. In a thin quantum dot, the energy levels are raised up in energy, which allows transport through the lowest level. This transport was investigated by scanning gate microscopy to map out the electrostatic environment of the quantum dot. Furthermore, double quantum dots were studied, where the characteristics of two single quantum dots were combined to form a more complex transport system.

The final part of the thesis is devoted to the development of vertical nanowire transistors with a gate that wraps around the nanowire channel.
This allows enhanced control of the potential in the channel compared to conventional planar devices. Three different devices were studied: 1) an InAs device with a \( \sim 1 \) \( \mu \)m gate length, 2) an InAs device with 50 nm gate length, and 3) a heterostructure device with an InAsP segment in the channel. The InAs devices showed good current saturation and subthreshold characteristics for drive voltages around 0.5 V. The InAsP device showed a reduced off current and lower inverse subthreshold slope in comparison to similarly processed InAs reference devices.
This thesis is based on the following papers, referred to in the text by their roman numerals.

I. Role of surface diffusion in chemical beam epitaxy of InAs nanowires
I did the planning, processing and growth of the samples and wrote the paper with Mikael Björk.

II. Surface diffusion effects on growth of nanowires by chemical beam epitaxy
A. Persson, L. E. Fröberg, S. Jeppesen, M. T. Björk, L. Samuelson
I did the processing and some of the growths and developed the model. I took part in writing the paper.

III. Diameter dependent growth rate of InAs nanowires
L. E. Fröberg, W. Seifert, and J. Johansson
I did the growths and analysis. Jonas Johansson developed the model and I wrote the main part of the paper.

IV. Transients in the formation of nanowire heterostructures
*Accepted for publication in Nano Lett.*
I planned the experiments together with Brent Wacaser, and I performed the growths and developed the model and wrote most of the paper. Brent
Wacaser and Jakob Wagner did the TEM investigations and the compositional analysis.

V. Few-electron quantum dots in nanowires
M. T. Björk, C. Thelander, A. E. Hansen, L. E. Jensen, M. W. Larsson, L. R. Wallenberg, and L. Samuelson
I took part in the growth and some of the measurements.

VI. Imaging a one-electron InAs quantum dot in an InAs/InP nanowire
I did the growth and processing of the sample. The measurements were done together with Ania Bleszynski-Jayich.

VII. Few electron double quantum dots in InAs/InP nanowire heterostructures
I developed the growth and did some of the measurements.

VIII. Vertical high-mobility wrap-gated InAs nanowire transistor
T. Bryllert, L.-E. Wernersson, L. E. Fröberg, and L. Samuelson
I did the growth.

IX. Vertical enhancement-mode InAs nanowire field-effect transistor with 50-nm wrap gate
C. Thelander, L. E. Fröberg, C. Rehnstedt, L. Samuelson, L.-E. Wernersson
I did the growth and together with Claes Thelander and Carl Rehnstedt developed the processing and did the measurements.
X. Heterostructure barriers in wrap gated nanowire FETs
L. E. Fröberg, C. Rehnstedt, C. Thelander, E. Lind, L.-E. Wernersson and L. Samuelson,
I did the growth and together with Claes Thelander and Carl Rehnstedt developed the processing. I performed most of the measurements.

Other papers have also been published but are not included due to overlapping content or because they deal with subjects beyond the scope of this thesis.

xi. Nanowire single-electron memory
C. Thelander, H. A. Nilsson, L. E. Jensen, and L. Samuelson

xii. Tunable effective g factor in InAs nanowire quantum dots

xiii. Fabrication, optical characterization and modeling of strained core–shell nanowires
Z. Zanolli, L. E. Fröberg, M. T. Björk, M.-E. Pistol and L. Samuelson

xiv. Nanowire-based multiple quantum dot memory

xv. Detection of charge states in nanowire quantum dots using a quantum point contact
D. Wallin, A. Fuhrer, L. E. Fröberg, L. Samuelson, and H. Q. Xu

xvi. Quantum-confinement effects in InAs–InP core–shell nanowires
Z. Zanolli, M.-E. Pistol, L. E. Fröberg and L. Samuelson
xvii. Shear stress measurements on InAs nanowires by AFM manipulation
M. Bordag, A. Ribayrol, G. Conache, L. E. Fröberg, S. Gray, L. Samuelson, L. Montelius, H. Pettersson
Small 3, 1398–1401 (2007)

xviii. Strain and shape of epitaxial InAs/InP nanowire superlattice measured by grazing incidence X-ray techniques
J. Eymery, F. Rieutord, V. Favre-Nicolin, O. Robach, Y.-M. Niquet, L. E. Fröberg, T. Mårtensson, and L. Samuelson

xix. Strain mapping in free-standing heterostructured wurtzite InAs/InP nanowires
Nanotechnology 18, 015504 (2007)

xx. AFM-based manipulation of InAs nanowires

xxi. Drive current and threshold voltage control in vertical InAs wrap-gate transistors

xxii. Electrical properties of self-assembled branched InAs nanowire junctions
D. B. Suyatin, J. Sun, A. Fuhrer, D. Wallin, L. E. Fröberg, L. S. Karlsson, I. Maximov, L. R. Wallenberg, L. Samuelson, and H. Q. Xu
Nano Lett. 8, 1100–1104 (2008)
The research presented in this thesis was performed during a period of 5 years at the Department of Solid State Physics. I started working with nanowires in a two-week project, with Magnus Borgström as my supervisor, during one of the final courses in my undergraduate studies. This project really caught my interest, and after the course I started working on my Master’s thesis dealing with embedding vertical nanowires in an insulating material, in order to place a top contact. This work became unexpectedly relevant in the later part of my studies and was interesting enough to convince me that I wanted to deepen my knowledge through postgraduate studies. So, I started my PhD in October 2003 and spent my time growing nanowires and studying electron transport through them at low temperatures. Half-way through my studies I was given the opportunity to work part-time for Qumat/Qunano on the development of a nanowire transistor. All this work has been brought together in this thesis. During my time as a PhD student I also had the opportunity to travel to conferences and meet with researchers to discuss physics. My contributions were mainly talks, sometimes even invited talks, which have given me valuable experience.

This work was made possible by the help of many people. First I would like to thank my main supervisor, Prof. Lars Samuelson, whose never-ending enthusiasm and inspiration for research motivated me during my work. This was especially important to me in difficult times. Furthermore, he allowed me to collaborate with scientists from around the world, which I really appreciated, and I have learnt a great deal from those collaborations. I’m also grateful to my assistant supervisors, Prof. emeritus Werner Seifert and Dr. Jonas Ohlsson. Werner has taught me, in principle, all I know about crystal growth and epitaxy. Jonas, who was my supervisor during my Master’s project, is one of the reasons I continued and started my PhD studies (it must have been the wine). Although you were away for a while, most of the time you were deeply involved in the discussions about nanowire growth.

Thanks to Sören, who always keeps “Bettan” in shape and for all the
comments and ideas, and the sailing. I will never forget sailing _Sjælland rundt_! Thank you Ann for handing over Bettan to me. I really enjoyed working with you. Thanks to Claes and Mikael who introduced me to the world of low-temperature electron transport, and to Andreas for teaching me to understand the measurements. I also enjoyed working with “trial and success”-Claes and “tjuvlyssnat”-Calle during the transistor work, I think we really did some nice work there. Thanks also to Philippe who supplied us with excellent wires, and to Erik and Lars-Erik for helping me understand the physics of devices.

I would also like to thank Prof. Robert Westervelt, Dr. Ania Bleszynski-Jayich, Halvar Trodahl and Erin Boyd for the collaborative efforts, both in Lund and at Harvard. Good luck with the new Dewar!

Thanks to Ladon, Jakob and Reine at the Department of Materials Chemistry for supplying me with amazing images, and for good collaboration.

The nanolab staff, Ivan, Mariusz, Lena, David, Peter, and Aline, for doing a great job and keeping the lab running. Thank you for all your instruction and help.

I would also like to thank the Epitaxy Group for valuable discussions during our Friday morning meetings, especially Jonas for helping me to actually understand epitaxy. Thanks to Mats-Erik for stimulating my mind with all kinds of problems. Thanks to Lars-Åke and Bo for being such good bosses, and for giving me the opportunity to catch a glimpse of the commercial world. Thanks to everyone at Qunano that I’ve worked with and gain knowledge from. Thanks to Kimberly and Brent, roommates who were always up for a lively discussion. Erik, Claes, and Ann, thank you for proofreading various parts of this thesis. Thanks to the innebandy team with Knut as our enthusiastic leader, for many hard battles.

Thanks to the staff at the Department of Solid State Physics, for making my time at the department so pleasant, and I apologize to anyone who should have been included in the above list, but that I forgot.

Finally, I would like to thank my family and friends for sharing the important things in life with me.

Louise, thank you for your endless love and support during these years, in both good times and bad. You make me a better person.

_Linus, September 2008, Lund_
1 Introduction

1.1 Crystalline materials and semiconductors 2
1.2 Energy structures 4
1.3 Heterostructures 5

2 Growth of nanowires

2.1 Crystal growth 9
  2.1.1 Nanowire growth 13
2.2 Chemical beam epitaxy 15
2.3 Time evolution of nanowire length 17
2.4 Temperature effects on growth 18
2.5 A diffusion model for nanowire growth 20
  2.5.1 Diffusion effects on growth 21
  2.5.2 Temperature effects on diffusion 22
2.6 Diameter-dependent growth rate 24
2.7 InAs/InP heterostructure growth 28

3 The physics of nanowires

3.1 Device processing 33
3.2 Constant interaction model 34
3.3 Single quantum dots in nanowires 37
3.4 Imaging single quantum dots 39
3.5 Double quantum dots in nanowires 41
4 Device applications of nanowires
   4.1 Field effect transistors .................................. 45
   4.2 Device processing ........................................... 49
   4.3 Long-channel transistors .................................. 51
   4.4 Short-channel transistors .................................. 53
   4.5 Heterostructure transistors ............................... 55

Future outlook ..................................................... 59

Populärvetenskaplig sammanfattning ................................ 61

References ...................................................................... 63
Nanoscience involves studying structures with sizes of 1–100 nm, a length scale where physics, chemistry, and biology meet. For example, the width of a DNA molecule is 2 nm [1], the thickness of a cell membrane is 5 nm [2], and the thickness of the thinnest layer in modern commercial transistors is 1.2 nm [3]. From the physics point of view, much of the work devoted to producing structures on the nanoscale has been stimulated by the increasing demand from the consumer electronics market to include more functionality in integrated circuits. In 1965, one of the co-founders of Intel, Gordon Moore, published a paper in which he predicted an exponential increase in the number of transistors that could be placed on an integrated circuit without increasing the cost per function [4]. This rate was first thought to be a doubling every year, but was later adjusted to a doubling every second year, and this is now known as Moore’s law. In his paper he states: “Integrated circuits will lead to such wonders as personal computers—or at least terminals connected to a central computer—automatic controls for automobiles, and personal portable communications equipment.”. These are things we take for granted today, and the word wonders reflects the incredible development in this area over the past fifty years.

Miniaturization is expected to continue for another decade or so [5], but the down-scaling of commercial transistors is reaching fundamental limits, for example, the finite size of an atom. Much research has therefore been
directed towards making devices with alternative materials and designs, which should still be compatible with traditional ones. This thesis presents one alternative, namely, vertical nanowire transistors made of indium arsenide (InAs).

The electronic revolution that Gordon Moore predicted would not have happened were it not for the enormous amount of basic research in materials science and physics. The fabrication of high-quality materials and a good understanding of the relevant physical processes are essential for the manufacture of competitive devices. The area covered by this thesis ranges from the fabrication of the material (nanowires), to studies of the physical properties and device characterization of nanowires. The outline is as follows:

**Chapter 1** briefly introduces the field of solid state physics and semiconductors.

**Chapter 2** is dedicated to nanowire epitaxy. Experiments and models used to explain and improve our understanding of nanowire growth are presented.

**Chapter 3** focuses on the physics of electron transport in nanowire quantum dots at low temperature.

**Chapter 4** describes the fabrication of vertical transistors and presents characteristics of long- and short-channel devices, as well as devices with heterostructure barriers in the channel.

**Chapter 5** gives a brief outlook on future developments of the field.

### 1.1 Crystalline materials and semiconductors

Solid materials can be categorized as crystalline or amorphous, depending on the arrangement of the atoms within the material. Many of the material-specific properties originate from the atomic structure, and crystalline and amorphous materials thus exhibit different properties. Crystalline materials can be categorized further, and the two most important crystal structures for semiconductors are zinc blende (or diamond) and wurtzite.

Atomistic views of these crystals can be seen in Fig. 1.1(a) for zinc blende, and (b) for wurtzite. The tetrahedral bonds are indicated by gray lines and the base vectors are labeled with the corresponding Miller indices, defining the different directions [6]. The \( \langle 111 \rangle \) direction in the zinc blende (red arrow in Fig. 1.1(a)) and the \( \langle 0001 \rangle \) direction in the wurtzite
1.1. Crystalline materials and semiconductors

crystal (red arrow in Fig. 1.1(b)) are very similar, and these are the directions in which nanowires tend to grow. All the nanowires described in this thesis had the wurtzite structure and were grown in the $\langle 0001 \rangle$ direction, exhibiting hexagonal cross section. The use of 4 base vectors to describe a three-dimensional world may seem strange, but these base vectors are conveniently related to the crystal symmetry. In Fig. 1.1 the crystals are made up of two kinds of atoms, each representing an atom from group III and from group V of the periodic table. These semiconductors are called III-Vs, and include InAs, InP, GaAs, and AlAs. The most well-known semiconductor is composed of Si atoms in a diamond structure. This is the same crystal structure as zinc blende but with only one type of atom. The name is taken from diamonds, which exhibit this structure.

![Figure 1.1: Atomistic view of (a) zinc blende, and (b) wurtzite crystal structures. In both cases the crystal consists of two types of atoms bound by tetrahedral bonds.](image)

Crystal surfaces are also denoted by Miller indices, which are enclosed in parentheses, where the direction with the same indices is the normal. The shaded area in Fig. 1.1(a) is a $\langle 111 \rangle$ surface, and if the crystal is viewed in a direction parallel to this plane, the crystal can be seen to be composed of alternating layers of “black” and “white” atoms, stacked in the $\langle 111 \rangle$ direction. Each monolayer consists of two bilayers, one containing group-III atoms and one containing group-V atoms. The starting point for all nanowires described in this thesis is an InAs($111$)B substrate, where “B” indicates that the upper bilayer contains As atoms.
1.2 Energy structures

The atomic periodicity in crystals determines the electronic structure of the material. The energy levels in the individual atoms are combined into bands of densely spaced energy levels that extend over the entire crystal. These energy bands are separated by energy gaps, or bandgaps, where no energy levels exist. In a perfectly pure III-V crystal, all the valence electrons form covalent bonds, which in an energy band picture, would mean that all the energy levels in the valence band are full, while no energy levels are occupied in the higher-lying band, the conduction band, see Fig. 1.2(a).

Figure 1.2: (a) Energy bands in a semiconductor showing an empty conduction band (white), except for the added electrons (dots) and a full valence band (shaded), except for the holes (rings). (b) Energy bands in an insulator with no available carriers, and (c) a metal with half-filled bands or overlapping bands.

Semiconductors typically have bandgaps of 0.5 eV to 3.0 eV, while the bandgaps in insulators are > 8 eV, and metals have half-filled bands or overlapping energy bands, as can be seen in Fig. 1.2 [7]. The width of the bandgap is related to the conductivity of the material. Metals, which have small or no bandgap, conduct well, while insulators, with large bandgaps, conduct poorly. Semiconductors are neither good conductors nor good insulators. However, their conductivity can be varied and controlled by adding foreign atoms, called dopants, to the crystal. Such doping can either donate an electron or cause a hole in the energy bands, creating available energy levels into which electrons can flow. If an atom with one extra electron is added to a crystal, this atom can donate the extra electron to the conduc-
tion band and still create 4 covalent bonds to the neighboring atoms. The material is then called n-type: n denoting negatively charged carriers. If, on the other hand, an atom with less than 4 electrons is added, an electron in the valence band can become trapped at this atom, enabling the 4 covalent bonds to be made, and leaving an empty energy level behind. Another electron in the valence band can now move into this empty level, effectively creating a positively charged hole. This material is called p-type: p denoting positively charged carriers. The local addition of electrons and holes is perhaps the single most important step in semiconductor processing.

At equilibrium, the electrons in a band moving in one direction are balanced by the same number moving in the opposite direction, thus there is no net current. In order for a current to flow through the crystal, a shift of this balance has to be created. This is done by exciting electrons to different energy levels within the band, for example, by applying a bias. However, in a fully occupied band, this shift is not possible, since all the levels are already occupied, and hence no current can flow. Thus, in order for the semiconductor to be conductive, dopants, either n- or p-dopants, need to be introduced.

1.3 Heterostructures

Different semiconductor materials have their valence bands and conduction bands at different energies relative to the vacuum level. If two materials are combined in one crystal, discontinuities will occur in the valence and conduction bands. An epitaxial combination of two materials is called a heterostructure. In Fig. 1.3(a), the band structure of a heterostructure is shown, where a segment of a wide-bandgap material has been inserted into a narrow-bandgap material. When electrons (black dots) in the narrow-bandgap material arrive at the interface with the wide-bandgap material, no more energy levels are available for the electrons, since this is the bandgap of the wide-bandgap material. The insertion of the wide-bandgap material results in an energy barrier for electrons. One way for electrons to cross this barrier is to gain enough thermal energy to overcome the barrier and enter the conduction band of the wide-bandgap material, as indicated by the solid arrow in Fig. 1.3(a). If the thermal energy is low relative to the barrier height and the barrier is narrow enough, another possibility exists. The electrons can quantum mechanically tunnel through the barrier, as shown by the dotted arrow in Fig. 1.3(a). For example, if the two narrow-
bandgap materials are contacted and a small bias is applied, the drifting electrons, measured as a current, must tunnel through the barrier if the thermal energy is low. This phenomenon is related to the wave nature of electrons. In principle, it means that an electron on the left-hand side of the barrier has a finite probability of being found on the right-hand side. Typical widths of tunnel barriers range from 1 to 5 nm, which corresponds to roughly 4–20 atomic layers. The above description of thermal excitation and tunneling is also applicable to holes in the valence band.

If tunneling were to take place on a macroscopic scale, strange things would happen. For example, imagine holding a bottle of beer in your hand, when it suddenly starts to tunnel through your hand, ending up on the other side of your hand, and falling to the floor!

![Figure 1.3:](image)

Figure 1.3: (a) Energy bands for a heterostructure barrier showing thermal excitation and tunneling as paths across the barrier. (b) Energy bands for a quantum dot showing energy level structure.

In Fig. 1.3(b), the inverse structure is shown; i.e. a thin segment of a narrow-bandgap material inserted into a material with a wide bandgap. In their constant effort to minimize their energy, the electrons in the conduction band of the wide-bandgap material will fall into the narrow-bandgap material. If the segment is thin enough another quantum mechanical effect will be observed. The spacing of the densely lying energy levels in the conduction and valence bands will increase, forming well-separated energy levels. The spacing between energy levels is dependent on the width of the segment. The energy levels are indicated in Fig. 1.3(b) by dotted lines,
and electrons are only allowed to exist at these specific energies. This energy level structure can be compared to climbing a ladder. You can only stand on the rungs. You can stand on two rungs at the same time, but if you try to place your foot in between two rungs, it will fall down to the closest lying rung below. As you climb up you gain potential energy and when you climb down you lower your potential energy. The same applies to confined electrons. They can only exist on an energy level or in a combination (superposition) of two or more energy levels. This is also true for holes in the valence band, with the difference that holes lower their energy by floating upwards. In Fig. 1.3(b) two electrons and a hole are trapped in the narrow-bandgap material. An electron and a hole can recombine leading to the emission of light. This is the principle behind the majority of light-emitting diodes (LEDs) and semiconductor lasers, which are found, for example, in laser pointers and DVD players.

In Fig. 1.3(b), only one spatial dimension is shown, but if electrons in the narrow-bandgap material are confined in all three dimensions the structure is called a quantum dot. The typical dimensions of a quantum dot (if one can speak of dimensions of a dot) are 5–15 nm.

In Fig. 1.3 the energy bands are all horizontal and the interfaces are vertical. In reality, charge redistribution and strain will create bending of the bands and a non-ideal heterostructure interface will create a smooth transition from one material to the other.
This chapter focuses on the growth experiments performed and the models used in order to broaden our knowledge and understanding of nanowire growth. Both basic growth studies, and more sophisticated studies, aimed at determining growth parameters, will be presented. This chapter is based on Papers I–IV.

2.1 Crystal growth

The word epitaxy originates from the Greek words ‘epi’, meaning above and ‘taxis’ meaning in ordered manner [6]. It thus describes the addition of atoms in a periodic arrangement forming a crystal lattice. The atoms constituting the crystal also give the crystal its specific material properties, such as electron mobility, bandgap, electron effective mass, etc. To fully understand why crystals grow under certain conditions, both thermodynamic effects and kinetic effects must be considered. This is far too complex to be dealt with fully here, instead we consider a simple box model.

Each box can be considered as one piece of growth material, usually called a monomer, cf. Fig. 2.1, which may consist of single atoms or more complex clusters. During growth, several individual processes occur, such as;

- Adsorption
Chapter 2. Growth of nanowires

- Diffusion
- Incorporation
- Desorption

The material is often supplied to the surface as a vapor or a molecular beam. The first step is the adsorption of growth material onto the crystal surface. Only a certain amount of the incoming material will stick to the surface, and the sticking coefficient is defined as the ratio of the adsorbed material to the incoming material. The sticking coefficient depends on the material supplied and the surface onto which it is adsorbed.

Once adsorbed, the monomer or adatom will diffuse over the crystal surface and find a suitable position to become incorporated into the crystal. The ease with which monomers diffuse over a surface is determined by the rate of the hopping from one surface site to a neighboring one, \( \nu_{\text{hop}} \), and the length of each hopping event, \( a \). These are combined to give a diffusion coefficient:

\[
D = a^2 \nu_{\text{hop}} = a^2 \nu_{\text{v}} e^{-E_s/kT},
\]

where \( \nu_{\text{v}} \) is the lateral vibration frequency of the atoms in the lattice [6]. The diffusion coefficient relates the average time a monomer remains on the surface to the mean distance traveled:

\[
\lambda^2 = D\tau,
\]

where \( \lambda \) is the diffusion length and \( \tau \) is the average time spent on the surface. \( \tau \) is limited either by desorption (\( \tau_{\text{des}} \)) or by incorporation (\( \tau_{\text{inc}} \)).

Assuming a constant monomer concentration and no interaction between the monomers on the surface, the diffusion process will be totally random and can be modeled as a ‘random walk’, where each step is independent of the previous one. Assume that a monomer adsorbs at \( x_0 \) on a one-dimensional (1D) axis. After an average time \( t_{\text{step}} = \nu_{\text{hop}}^{-1} \) the monomer will move a distance \( a \) in either positive or negative direction with equal probability. After \( n \) steps (where \( n \) is a large number), taking a total time \( \tau = n \cdot t_{\text{step}} \), the probability of finding the monomer at a position \( x \) is described by a Gaussian (normal) distribution:

\[
P(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(x-x_0)^2/2\sigma^2},
\]
2.1. Crystal growth

where the standard deviation $\sigma = a\sqrt{n}$ [8]. From Eqs. 2.1, 2.2 and $n = \tau_{\text{hop}}\nu$, we obtain $\sqrt{n} = \lambda/a$ and $\sigma = \lambda$, that is, the diffusion length equals the standard deviation. Thus:

$$P(x) = \frac{1}{\sqrt{2\pi\lambda}} e^{-\frac{(x-x_0)^2}{2\lambda^2}}.$$  (2.4)

This will be used to describe how monomers diffuse from the substrate surface to the growth front of a nanowire.

Monomers diffusing over a crystal surface will always have a probability of getting re-evaporated, or desorbed, from the surface. The process is temperature dependent and is governed by the activation energy of desorption, $E_d$. The rate of desorption can be described by:

$$\frac{1}{\tau_{\text{des}}} = \nu_\perp e^{-E_d/kT},$$  (2.5)

where $\nu_\perp$ is the vertical vibration frequency of the atoms in the lattice. The desorption rate will vary depending on the surface structure and the type of monomer.

Combining Eqs. 2.1, 2.2 and 2.5 and assuming $\nu_\perp = \nu_\parallel$ yields:

$$\lambda = ae^{(E_d-E_s)/2kT}.$$  (2.6)

Eq. 2.6 relates the diffusion length to the distance between surface sites, and since we expect $\lambda > a$, $E_d$ must be greater than $E_s$ [6]. Note that Eq. 2.6 is only correct when the growth is desorption limited. In this limit the diffusion length decreases as the temperature increases due to desorption. Desorption-limited growth is dominant at high growth temperatures and low material fluxes.

At lower temperatures, or when desorption can be neglected, the average time a monomer remains on the surface is limited by incorporation and $\tau$ is inversely proportional to the incorporation rate.

Diffusion allows the monomer to find a position at which it can bind and be incorporated into the crystal lattice. The more bonds it can form to the lattice, the lower its energy becomes. The positions that are most attractive for the diffusing monomers can be identified by simply counting the number of faces of a box in Fig. 2.1 that face (bind to) the original crystal. Incorporation onto a flat surface only creates one bond to the crystal (purple boxes). Such a position is least favorable for a monomer, and the bond is probably not stable. A certain number of neighboring

---

11
monomers on a flat surface is usually required in order for such a 2D nucleus to be stable, as indicated in the right in Fig. 2.1 (yellow boxes). A monomer that diffuses to, and is incorporated at, a step has two of its faces towards the crystal, and this position is thus more favorable (green box). Such a monomer can continue to diffuse along the step to a kink, where three of its faces are bound (red boxes), and so on.

So far, only kinetic effects have been considered, but in order for incorporation to occur, a thermodynamic driving force must exist. This driving force is called supersaturation, and is defined as the difference in chemical potential between the vapor phase and the solid, $\Delta \mu = \mu_{\text{vapor}} - \mu_{\text{solid}}$. In the case of condensation of a vapor the supersaturation can be expressed as:

$$\Delta \mu = kT \ln \left( \frac{p}{p_0} \right),$$

(2.7)

where $p_0$ is the equilibrium pressure and $p$ the actual pressure. If $\Delta \mu$ is positive the vapor condenses, and if it is negative it evaporates/sublimes.

For example, imagine a sealed container containing water and water vapor at an elevated temperature. The vapor is at its equilibrium pressure, which means that an equal number of water molecules are being evaporated from the liquid as condensed from the vapor. If water vapor is added to the system, keeping other parameters constant, the vapor will be supersaturated and water droplets will form. This will restore the equilibrium.

This example is similar to a crystal surface in contact with its vapor phase. If the system is maintained at equilibrium, an equal number of monomers will condense on and evaporate from the surface. However, if
material is added in the vapor phase, the equilibrium will be disturbed, and more monomers will condense than evaporate, i.e. $\Delta \mu > 0$, and growth is possible.

2.1.1 Nanowire growth

The first solid model to describe nanowire growth was proposed by Wagner and Ellis in the 1960s, who studied the growth of micron-sized Si wires from vapor [9]. They noted some important features: (a) Si wires did not contain a screw dislocation, which excluded a previously suggested growth mechanism, and (b) a small metal particle was present on the top of the wire. The information in (b) and other ideas led to the vapor liquid solid (VLS) growth model. The essential feature of this model is that the metal particle forms a liquid alloy with the growth material, exhibiting a low melting point. The gas phase molecules preferably condense on the impurity and alloy with the impurity. The alloy eventually becomes supersaturated and wire growth occurs through precipitation of the growing material from the impurity. In the case of Si wire growth from Au particles, the Au–Si forms a eutectic alloy with a melting temperature of 363°C, which should be compared to the individual melting temperatures of 1063°C and 1412°C of Au and Si, respectively. This model has been widely referenced and used to explain all types of nanowire growth. It has recently become clear that other mechanisms might be responsible for the growth of nanowires, when the metal particle is in its solid phase [10, 11].

This thesis describes studies of nanowires containing In, As and P grown from Au particles. Studies have shown that In alloys with Au, and that this alloy persists throughout growth [10]. It is also known that As has a very low solubility in Au and is not expected to alloy with Au [12]. Furthermore, P could form a stable alloy with Au and In, but post-growth measurements in a transmission electron microscope showed no traces of either As or P in the Au particle (see Paper IV). The measurements were performed using energy dispersive X–ray spectroscopy (EDS) and electron energy loss spectroscopy (EELS). Indium is thus the only element alloying with Au. The Au–In material system has a eutectic point at 454°C [13], which is above all the growth temperatures used in this work. This suggests that the Au–In seed particle is solid during growth, but our knowledge of nanoscale phase diagrams is not sufficient for this to be a reliable argument.

The group-V materials have extremely short diffusion lengths in molecular beam epitaxy and chemical beam epitaxy [14], and can only contribute
Figure 2.2: (a) Schematic view of nanowire growth. The group-V material contributes to nanowire growth only if it is adsorbed at or close to the Au particle. (b) The fraction of the material collected from the surface decreases, according to a Gaussian function, with the distance from the nanowire. Assuming a constant monomer concentration, the available material at a distance $R$ from the nanowire is proportional to the area $dA = 2\pi R \cdot dR$. 

to nanowire growth if the molecules impinge directly on the seed particle. Group-V molecules are always supplied in excess, and therefore seldom limit the growth. The group-V molecules adsorbing on the substrate surface only contribute to planar 2D growth or surface reconstruction.

On the other hand, the group-III molecules have a longer diffusion length and can contribute to nanowire growth either through direct impingement on the seed particle or via diffusion from the substrate surface, up the sides of the nanowire, as illustrated in Fig. 2.2(a). This diffusion is characterized by two diffusion lengths, one on the substrate surface, $\lambda_s$, and one on the nanowire side facets, $\lambda_{nw}$. As will be shown in Section 2.3, $\lambda_{nw}$ is always longer than the nanowires in these studies, and hence never limits the total diffusion length. Due to the vertical impingement of the material beams in chemical beam epitaxy, a negligible amount of growth material will be adsorbed onto the nanowire side facets. The amount of growth material that desorbs from the nanowire side facets is also negligible. Thus, all the molecules that diffuse from the substrate to the nanowire side facets will reach the nanowire growth front, that is, the seed particle–nanowire interface. Direct impingement and diffusion are denoted $J_{direct}$ and $J_{diff}$, respectively, see dashed circles in Fig. 2.2(a).
2.2 Chemical beam epitaxy

Chemical beam epitaxy (CBE) [15] is a high-vacuum technique, similar to molecular beam epitaxy (MBE) [16], the main difference being that vapor or liquid precursors are used in CBE. The technique was developed at Bell Labs in the early 1980s, stimulated by the interest in heterostructure lasers in the Ga$_x$In$_{1-x}$As$_y$P$_{1-y}$ system. The use of vapor sources for group-V materials was expected to yield more stable beam fluxes and better control of the As/P ratio than the elemental sources used in MBE. The use of metalorganic sources for group-III materials was motivated by their long-term stability and easy handling when changing sources.

The present setup is partly home-built and the precursors used are tertiary-butyl-arsine (TBAs) and tertiary-butyl-phosphine (TBP) as group-V sources, and tri-methyl-indium (TMIn) as the group-III source. The source bottles are stored in a water bath, which is maintained at 10°C. At this temperature the TBAs and the TBP are both liquid, and the TMIn is a porous solid. The vapor pressure in the source bottles is sufficient to transfer the material from the bottles to the growth chamber, i.e. no carrier gas is needed.

![Figure 2.3: Schematic of the chemical beam epitaxy system.](image)

Fig. 2.3 shows a schematic of the CBE system. The source gases from
Chapter 2. Growth of nanowires

the bottles can be directed to the growth chamber or to the vent chamber. It is important that the two chambers have the same pressure, otherwise pressure waves will be created as a source is redirected from one chamber to the other, causing inhomogeneities in the growth rates. The pressure on the vent side can be adjusted in order to fulfill this requirement. The TMIn passes a constriction when entering the growth chamber, at which point the partial pressure of the TMIn decreases, and the molecules are transported in a cone-shaped beam. The density of molecules will vary over the cross section of the beam but, over the 2-inch diameter sample holder in the system used, the variations are expected to be less than 10%. When growing nanowires the sample sizes were usually less than 1x1 cm$^2$ and variation in beam pressure should therefore be negligible.

The group-V gases are thermally cracked upon entering the growth chamber. The crackers are situated before the constriction, which is in fact inside the growth chamber since it is placed after the valve, see Fig. 2.3. A spiral-shaped molybdenum filament is placed inside the TBAs line. A large current is applied, causing the filament to reach about 1000$^\circ$C. After the filament, there is a similar constriction to that for TMIn in order to produce a beam. The reaction products important for growth are AsH$_3$, As$_2$ and As$_4$ [17]. The TBP is cracked using a molybdenum filament in the wall of the line. The reason for having the filament in the wall is that TBP, or its reaction products, form a crust on the filament if it is situated in the gas flow, which makes the filament less stable. The main reaction product from the cracking of TBP, which contributes to growth, is P$_2$ [17]. Other products include H$_2$ and various hydrocarbons. Hydrogen alone is the main reason for the increased pressure in the growth chamber during growth, and it can reach a pressure up to $10^{-4}$ mbar. Gases passing to the vent chamber are not cracked and thus the vent chamber requires less pumping capacity. The base pressure, that is the pressure in the growth chamber without source gases, is $10^{-8}$ mbar.

Before the gases are switched to either the growth chamber or the vent chamber, they pass a needle valve and a pressure gauge. The measured pressure is fed back to the pressure control unit which adjusts the needle valve. This results in stable pressures in the lines leading to the growth chamber. The pressures reported in this thesis are the regulated pressures in the lines.

The areas close to the sample holder in the growth chamber and the interior of the vent chamber are both flushed with liquid nitrogen (N$_2$)
2.3. Time evolution of nanowire length

during growth, see Fig. 2.3. This enhances the condensation of material on the walls, which lowers the pressure in the chambers. It also minimizes the risk of non-beam material, such as partly decomposed source molecules, condensing on the sample.

The sample is indium-bonded to a 2-inch molybdenum disc, which is loaded into the growth chamber through a load-lock, (not shown in Fig. 2.3). The disc is attached to a rotatable manipulator about 1 mm from the heating filament, and the sample is heated from behind, see Fig. 2.3. The filament is spirally shaped to achieve homogeneous heating of the molybdenum disc, and the heater is shielded to minimize the risk of source material cracking directly on the heater.

The system is also equipped with a reflection high-energy electron diffraction (RHEED) system. The RHEED system accelerates electrons to 20 kV and directs the beam towards the sample surface at a grazing angle. The reflected and diffracted electrons are visualized on a phosphorus screen opposite the electron gun. The resulting pattern depends on the surface characteristics. An amorphous surface, such as an oxide, would not give any diffraction pattern. A flat regular surface would yield a series of diffraction lines, the separation of which is related to the unit cell on the surface. This is used to characterize the surface during deoxidation of the sample and nucleation of the nanowires.

2.3 Time evolution of nanowire length

Studying the time evolution of nanowire length can provide important information on the nucleation and growth of these structures. In Fig. 2.4 the nanowire length is plotted as a function of growth time for wire diameters of 55, 80, and 120 nm. The start of the growth time is defined as the time at which TMIn is introduced into the chamber. The lines are linear fits to the data and extrapolate to zero length for zero growth time. This shows that the nanowires start to grow as soon as TMIn is available. It is known from elemental analysis of the seed particle that it is a Au–In alloy [10] and this suggests that alloying occurs before the TMIn is introduced, probably during deoxidation. This is further supported by studying samples grown for short times, where pits in the substrate are seen around the seed particle. The pits are probably the result of the Au collecting In from the substrate.

Comparing the different slopes it can clearly be seen that thicker nanowires
Chapter 2. Growth of nanowires

Figure 2.4: Nanowire length as a function of growth time for three different nanowire diameters. Lines are linear fits to the data and the error bars correspond to one standard deviation.

have a slower growth rate. This is discussed in more detail in Section 2.6.

As the nanowires increase in length the In-species on the substrate surface have to diffuse further up the nanowire side facets in order to reach the growth front. If the length of the nanowire becomes greater than the diffusion length on the nanowire side facets, then the amount of In reaching the growth front would decrease drastically, as would the growth rate. In Fig. 2.4 the growth rate is constant, with no sign of a decrease, indicating that the diffusion length on the nanowire side facets is greater than the nanowire length.

2.4 Temperature effects on growth

The growth temperature has a considerable impact on both the chemical reactions and the kinetic effects on the substrate surface. Growth was performed at various temperatures, but with the same pressures (TMIn: 0.15 mbar, TBAs: 1.5 mbar) and the results are presented in Fig. 2.5.

The nanowire growth rate, indicated by squares, peaks at 415°C indicating optimal conditions for nanowire growth regarding the collection of material and incorporation efficiency. It is important to note that the growth of nanowires is dependent on the conditions on the substrate surface, since the substrate acts as a collector of growth material. For growth
2.4. Temperature effects on growth

![Graph](image)

**Figure 2.5:** Nanowire growth rate (squares) and substrate growth rate (triangles) as a function of growth temperature. Error bars correspond to one standard deviation.

Temperatures below 415°C nanowire growth is believed to be limited by incomplete decomposition of TMIn. To investigate this the substrate growth rate was extracted from RHEED oscillations [18] performed under the same growth conditions. The results are shown as triangles in Fig. 2.5. Clearly, the growth rate increases as the temperature is raised, and the activation energy was extracted by fitting the equation:

\[ R = A e^{-E_a/kT}. \] (2.8)

The value obtained for \( E_a \) was 0.94 eV, which is close to the value of 1.03 eV, obtained by Verschuren et al. [19] for TMIn decomposition on InP(001).

Therefore, it was concluded that in the temperature range where nanowires were grown the substrate growth rate is limited by the decomposition of TMIn. The slope of the line fitted to the substrate growth rate is similar to the slope of the low-temperature side of the nanowire growth rate, and it can be concluded, as stated above, that the nanowire growth rate is limited by incomplete decomposition of TMIn at temperatures below 415°C.

The decrease in the nanowire growth rate at higher temperatures could
have several explanations, (i) In could start to desorb from the surface which, according to Eq. 2.6, would shorten the diffusion length and lead to less material diffusing from the substrate surface to the growth front. (ii) As the growth temperature increases, the 2D nucleation, and 2D growth would increase, leading to less material reaching the growth front. According to previous studies of the growth of InP by CBE using TMIn, the desorption of In is only important for growth temperatures above 535°C [19, 20]. These studies were carried out on InP(001), and not the InAs surface in the present experiments, but in the absence of better data, this must suffice as a first approximation. Therefore, the possibility of In desorption being the limiting factor for growth rates in the temperature range 415°C–445°C was ruled out. The reason for the decreasing nanowire growth rate is believed to be related to a decrease in the diffusion length, as will be discussed in Section 2.5.2.

2.5 A diffusion model for nanowire growth

Classically, the diffusion length is defined as in Eq. 2.6 [6], with desorption limiting the diffusion length. However, the incorporation of material into the substrate should also be considered as a possible limiting factor for diffusion in the present case. Thus, an adatom can diffuse over the substrate surface for a certain time, \( \tau_{\text{total}} \), which is limited by desorption or incorporation into the substrate;

\[
\frac{1}{\tau_{\text{total}}} = \frac{1}{\tau_{\text{des}}} + \frac{1}{\tau_{\text{inc}}}.
\]  

(2.9)

Which of the two processes limits the material transport to the nanowire growth front depends on growth conditions such as temperature and source pressures. The diffusion length is defined as the distance traveled from adsorption to desorption or incorporation into the substrate.

To be able to extract quantitative results from the growth experiments performed, a simple model for the diffusion of material from the substrate surface was developed. As discussed in Section 2.1, a monomer follows a random path which, after many steps, results in a Gaussian probability distribution around the site of adsorption. From the nanowire point of view, the probability of a monomer reaching the nanowire base via diffusion follows a Gaussian distribution centered at the nanowire, see Eq. 2.4. As pointed out above, the diffusion length, \( \lambda_s \), equals the width (standard
deviation $\sigma$) of the distribution, see Fig. 2.2(b). The available material at a distance $r$ from the wire is proportional to the area of an infinitesimally thin circle, $dA = 2\pi r \cdot dr$, centered at the nanowire. Multiplying the two parts yields the amount of material, $M$, collected around the nanowire base:

$$M \propto \frac{1}{\sqrt{2\pi \lambda_s}} e^{-\frac{r^2}{2\lambda_s^2}} \cdot dA = \frac{1}{\sqrt{2\pi \lambda_s}} e^{-\frac{r^2}{2\lambda_s^2}} \cdot 2\pi r \cdot dr$$  \hspace{1cm} (2.10)$$

Integrating Eq. 2.10 from the nanowire radius, $r_w$, to a certain distance from the nanowire, $R$, and adding two constants yields the nanowire growth rate.

$$\frac{dL}{dt}(R) = c_1 \cdot \sqrt{2\pi \lambda_s}(e^{-\frac{r_w^2}{2\lambda_s^2}} - e^{-\frac{R^2}{2\lambda_s^2}}) + c_2$$  \hspace{1cm} (2.11)$$

The constant $c_1$ relates the material that is collected to that which actually becomes incorporated, and $c_2$ accounts for the growth rate due to the direct impingement of material on the Au particle. This expression is used to fit a curve to the experimental data using $c_1$, $c_2$ and $\lambda_s$ as fitting parameters.

Note that this is by no means a complete model of diffusion, but rather a simple model that relates the growth rate of a nanowire to the available substrate surface from which growth material may be collected. The model is used in the following sections to systematically extract diffusion lengths from experimental data.

### 2.5.1 Diffusion effects on growth

To study the effects of material diffusion on the growth of nanowires, the nanowire growth rate was measured for varying inter–wire distance. The Au seeds were positioned in a triangular pattern using electron beam lithography (EBL), evaporation of 15 nm Au and subsequent lift–off, see Fig. 2.6(a). The remaining seed particles are disc-shaped and after heating in the growth chamber they reshape into a hemisphere. Each nanowire then has a hexagonal area on the substrate surface from which it can collect material. This is called the available surface collection area, and is shown in Fig. 2.6(a).

As discussed in the previous section, the incoming material adsorbed on the substrate surface will diffuse and, depending on the distance from the site of adsorption to a nanowire, the diffusing species will have a certain probability of reaching the growth front, cf. Eq. 2.11. To investigate this, different areas on the same sample were prepared with inter–wire distances
Figure 2.6: (a) Top-down view of Au seeds positioned in a triangular pattern. The dashed white lines indicate the hexagonal available collection area for three adjacent nanowires. (b)–(d) Nanowires grown with an inter–wire distance of 0.5, 0.75, and 1 µm respectively. All scale bars are 1 µm.

varying from 250 nm to 8 µm. Thus, all the different areas went through the same process steps and were grown simultaneously, making the comparison between different areas reliable. In Fig. 2.6(b)–(d) nanowires grown with inter–wire distances of 0.5, 0.75, and 1 µm, respectively, are shown. It can clearly be seen that the nanowire length increases as the inter–wire distance increases. This shows that nanowires close to each other compete for the surface-diffusing In–species on the substrate surface. When the inter–wire distance is longer than the diffusion length of the In–species, this dependence disappears, and the growth rate does not increase with larger available collection areas.

2.5.2 Temperature effects on diffusion

The method presented in the previous section was used to study samples grown at temperatures ranging from 404°C to 445°C. The results are shown in Fig. 2.7(a)–(c). The squares indicate the experimentally determined growth rates and the solid lines are the least-squares fits of Eq. 2.11 to the data using $c_1$, $c_2$ and $\lambda_s$ as fitting parameters. The diffusion lengths for different growth conditions, such as temperature and source pressures can then be compared.

The diffusion lengths obtained are shown in Fig. 2.8. As can be seen, the diffusion length decreases as the temperature increases. Again, Eq. 2.6 could explain this behavior but, as concluded in Section 2.4, the temper-
2.5. A diffusion model for nanowire growth

Figure 2.7: Nanowire growth rate as a function of available collection area at (a) 404°C, (b) 425°C, and (c) 445°C. Error bars correspond to one standard deviation and the solid lines are the results of fitting the model described above to the data.

ature is not high enough for desorption to be important. The decrease in diffusion length as the temperature increases is instead attributed to the increased 2D incorporation rate on the substrate surface. This reduces the lifetime of the adatoms on the substrate surface, thus decreasing the diffusion length according to Eq. 2.2.

Figure 2.8: Diffusion length as a function of inverse temperature.
2.6 Diameter-dependent growth rate

The dependence of growth rate on the diameter of the nanowire has been discussed since the 1970s [21]. Reported dependencies, as the nanowire diameter decreases, include monotonically increasing [22], monotonically decreasing [21], independent [23], first increasing then decreasing [24], and first decreasing then increasing [25]. Here, experimental results are presented that show first an increase and then a decrease of the growth rate as the diameter is decreased. A model was developed and fitted to the experimental results, allowing the extraction of fundamental growth parameters.

The diffusion model presented and used in the previous sections is limited to conditions of constant adatom concentration. This is not a very realistic situation and in this section another model will be presented, which allows the surface concentration of adatoms to vary with the distance from the nanowire. As was explained earlier, the two ways in which In can reach the growth front are by direct impingement and diffusion from the substrate surface. The contribution to the growth rate from direct impingement is assumed to be:

\[
\left( \frac{dL}{dt} \right)_{\text{direct}} = R \Omega_{\text{InAs}},
\]

where \( R \) is the beam flux and \( \Omega_{\text{InAs}} \) is the molecular volume of InAs. Note that in this simple form, Eq. 2.12 is independent of nanowire diameter. The beam flux was measured from the results of bulk growth experiments and found to be 1 nm\(^{-2}\) s\(^{-1}\).

The adatom concentration on the substrate surface, \( n_s \), is governed by the diffusion equation:

\[
\frac{\partial n_s}{\partial t} = D \nabla^2 n_s - \frac{n_s}{\tau} + R,
\]

where \( \tau \) is the lifetime of adatoms on the substrate surface. If no nanowires are present the adatom concentration would be homogeneous and equal to \( R \tau \). The steady-state condition, \( \partial n_s / \partial t = 0 \), has previously been solved using the boundary condition \( n_s(r_w) = 0 \), where \( r_w \) is the nanowire radius [26]. However, in this work, the boundary condition was set to \( n_s(r_w) = \gamma R \tau \), where \( 0 \leq \gamma \leq 1 \). The dimensionless parameter \( \gamma \) allows the adatom concentration at the base of the nanowire to vary between 0 and \( R \tau \). One solution to this problem is:
2.6. Diameter-dependent growth rate

\[ n_s(r) = R\tau \left( 1 - \frac{K_0(r/\lambda_s)}{K_0(r_w/\lambda_s)} + \gamma \frac{K_0(r/\lambda_s)}{K_0(r_w/\lambda_s)} \right), \quad r \geq r_w, \quad (2.14) \]

where \( K_0(x) \) is the modified Bessel function of the second kind. Now that \( n_s \) is known, the flux of diffusing adatoms from the substrate onto the nanowire side facets can be expressed as:

\[ J_{\text{diff}} = D \frac{\partial n_s}{\partial r} \bigg|_{r=r_w} = R\lambda_s(1 - \gamma) \frac{K_1(r_w/\lambda_s)}{K_0(r_w/\lambda_s)}, \quad (2.15) \]

where \( K_1'(x) = -K_1(x) \) has been used. This flux is the number of adatoms that cross the interface between the substrate and the nanowire side facets per nm per s. As mentioned earlier, this equals the flux at the nanowire–seed particle interface. If \( J_{\text{diff}} \) is multiplied by \( 2\pi r_w \Omega_{\text{InAs}} / \pi r_w \), an expression is obtained for the growth rate supported by In atoms diffusing from the substrate surface. The total growth rate, including the part due to direct impingement, Eq. 2.12, is thus:

\[ \frac{dL}{dt} = R\lambda_s(1 - \gamma) \frac{2\Omega_{\text{InAs}}}{r_w} \frac{K_1(r_w/\lambda_s)}{K_0(r_w/\lambda_s)} + R\Omega_{\text{InAs}}. \quad (2.16) \]

In Eq. 2.16, \( \lambda_s \) and \( \gamma \) are unknowns and \( \lambda_s \) is assumed to be constant. However, if \( \gamma \) is also assumed to be constant, \( dL/dt \) is a continuously increasing function with decreasing nanowire radius. This second assumption is not consistent with the experimental data. In Fig. 2.9 the growth rate is plotted as a function of diameter and the dependence exhibits a maximum at a diameter of about 25 nm. Thus, the parameter \( \gamma \), which relates the adatom concentration at the nanowire base to the concentration far away from the nanowire, must be a function of the radius. If the nanowire growth front is ideal, then all atoms reaching the nanowire base will be incorporated into the nanowire, and \( \gamma = 0 \). If, on the other hand, some non-ideality is present, the incorporation rate will decrease, leading to a build-up of adatoms at the nanowire base, hence, \( \gamma \neq 0 \). It is reasonable to expect that \( \gamma \) is a function of \( r_w \).

The decreasing growth rate observed for small nanowire diameters is interpreted as the Gibbs–Thomson effect. That is, as the seed particle diameter decreases, the equilibrium pressure of In atoms in the seed particle, \( p_0(r_w) \), increases as:

\[ p_0(r_w) = p_\infty e^{2\sigma\Omega_{\text{Au}}/r_w kT}, \quad (2.17) \]
where $p_\infty$ is the equilibrium pressure in an infinitely large seed particle, and $\sigma$ is the surface energy density of the seed particle. Here it is assumed that the nanowire and the seed particle have the same radius. This leads to lower supersaturation, that is, a lower driving force of In atoms diffusing from the surroundings into the seed particles and consequently also into the nanowire, according to Eq. 2.7 and:

$$\Delta \mu (r_w) = kT \ln \left( \frac{p}{p_0(r_w)} \right) = kT \ln \left( \frac{p}{p_\infty} \right) - \frac{2\sigma \Omega_{Au}}{r_w} = \Delta \mu_\infty - \frac{2\sigma \Omega_{Au}}{r_w}. \quad (2.18)$$

Here, the supersaturation for an infinitely large seed particle is defined as $\Delta \mu_\infty$. This expression for the decrease in supersaturation is the same as that presented by Givargizov [21]. In the absence of values for the surface energy density for a Au–In alloy, the value for pure Au was used, 1.2 J/m$^2$, and $\Omega_{Au}$ is assumed to be the same as for a Au crystal in the face-centered cubic structure, 0.017 nm$^3$. In order to include the Gibbs–Thomson effect on supersaturation in the diffusion model, it is expressed as a dimensionless parameter that varies between 0 and 1, and is related to $\gamma(r_w)$,
2.6. Diameter-dependent growth rate

\[ \gamma(r_w) = 1 - \frac{\Delta \mu(r_w)}{\Delta \mu_\infty} = \frac{2 \sigma \Omega_{Au}}{r_w \Delta \mu_\infty}. \]  

(2.19)

Inserting this into Eq. 2.16 yields the complete expression for the growth rate:

\[ \frac{dL}{dt}(r_w) = R \lambda_s \left( 1 - \frac{2 \sigma \Omega_{Au}}{r_w \Delta \mu_\infty} \right) \frac{2 \Omega_{InAs} K_1(r_w/\lambda_s)}{K_0(r_w/\lambda_s)} + R \Omega_{InAs}. \]  

(2.20)

In Eq. 2.20 the only unknowns are \( \lambda_s \) and \( \Delta \mu_\infty \), and these are used as parameters when fitting Eq. 2.20 to experimental data. The result can be seen as the solid line in Fig. 2.9, giving values of \( \lambda_s = 130 \) nm and \( \Delta \mu_\infty/kT = 0.54 \). For nanowire diameters greater than 25 nm, the growth rate is limited by diffusion, whereas for smaller diameters the Gibbs–Thomson effect limits growth, eventually causing the nanowire to stop growing at the critical diameter, \( d_c \).

A similar diameter dependence was seen for different In supply pressures, which allows the diffusion length and the critical diameter to be studied as a function of supersaturation. The curve in Fig. 2.9 shifts to the upper left with increasing supersaturation and to the lower right with decreasing supersaturation. The values extracted for the critical diameter and the diffusion length are presented in Fig. 2.10(a) and (b), respectively. In this range of In supply pressures the critical diameter varies between 11 nm and 18 nm, and the diffusion length between 130 nm and 180 nm. An expression for the critical diameter can be obtained by setting \( dL/dt = 0 \) in Eq. 2.20, neglecting the term for direct impingement and solving for \( r_w \):

\[ d_c = 2r_c = 2 \frac{2 \sigma \Omega_{Au}}{\Delta \mu_\infty} \frac{4 \sigma \Omega_{Au}}{kT \ln \left( \frac{p}{p_\infty} \right)}. \]  

(2.21)

The dependence of the critical diameter on the supply pressure of In in Fig. 2.10(a) is consistent with Eq. 2.21. From Eq. 2.21 it is also clear that as the temperature is increased the critical diameter decreases, which is consistent with previous experimental results [27]. Furthermore, increasing the supersaturation will also decrease the critical diameter.

The two methods of determining the diffusion length of In species presented in the present chapter give similar results, around 100–200 nm.
Chapter 2. Growth of nanowires

2.7 InAs/InP heterostructure growth

The epitaxial combination of different semiconductor materials has led to the study of new physical phenomena [28, 29] and new devices [30]. The only constraint is that the crystal periodicity of the two materials must match, otherwise strain will build up in successive crystal layers, eventually causing interface and crystal defects. This can be illustrated in a box model, similar to Fig. 2.1, but using differently sized boxes. In order to fit a top layer (adlayer), these differently sized boxes have to be resized in a plane parallel to the bottom layer. In Fig. 2.11 this is shown schematically for a situation where a layer of orange boxes (large boxes) is epitaxially grown on a crystal made up of small gray boxes. As the orange boxes bind to the underlying crystal they become compressed in the horizontal directions and stretched in the vertical, and are shown red depicting strain. If the strain in the red layer is too great, crystal defects will form, as is illustrated by the yellow boxes, which have the same dimensions as the orange boxes, and two orange boxes will bind to three underlying gray boxes. The strain can also cause the surface to reshape, forming small pyramids or islands, so-called Stranski–Krastanow islands [6]. In this growth mode a small number of strained layers will grow before reconfiguration of the surface occurs.

In the case of nanowires, where the cross-sectional area is small, the nanowire can accommodate the strain induced in a heterostructure by changing its diameter (Paper xix). This allows for defect-free formation of heterostructures from a much broader range of materials than in bulk heterostructure growth. Many material systems have been shown to produce nanowire heterostructures, including Si/Ge [31], InAs/InP [32] GaAs/GaP
2.7. InAs/InP heterostructure growth

Figure 2.11: Block heteroepitaxy where orange boxes are grown on gray boxes. The adlayer adapts to the underlying crystal resulting in a strained layer, illustrated by a color change from orange to red. Too much strain can lead to crystal and interface defects, as shown by the two yellow boxes which are bound to three gray boxes.

[33] and GaN/InGaN [34], but few are sharp on the atomic scale. One reason for the smoothness of heterostructures in nanowires is that growth material alloys with the metallic seed particle. When switching materials there may be a delay due to emptying/filling of material from/into the seed particle, causing a gradient in the heterostructure [35] or transients in the growth rate. For example, in the combination of InAs and GaAs it is obvious that In and Ga will alloy differently with the seed particle and that this can cause irregularities. That which is presented here is a more subtle effect which occurs with the combination of InAs and InP.

In Fig. 2.12(a) a high-resolution transmission electron microscope (TEM) image of an InAs/InP nanowire heterostructure is shown. The crystal structure is wurtzite, both in the InAs and the InP domains, and very few crystal defects can be seen. In Fig. 2.12(b) a scanning TEM (STEM) image is shown where an atomically abrupt interface can be seen. In this system, In and Au alloy and growth proceeds from this alloy. Unfortunately, it was not possible to study the composition of the seed particle during growth, and it was necessary to rely on post-growth measurements. Compositional analysis was performed on nanowires whose growth was stopped abruptly, by closing both the As/P and In source simultaneously, and then cooling down. The growth temperature, 390°C, is probably low enough to ensure that the nanowire remains intact. This abrupt growth termination is as close as one can get to quenching the seed particle. The In composition in the seed particle was determined from EDS measurements to be 34% (atomic) for an InAs nanowire, and 44% (atomic) for an InP nanowire after abrupt growth.

29
termination. These are assumed to be the steady-state In compositions of the seed particle during the growth of InAs and InP. Thus, when switching between growing InAs and InP the seed particle composition must adjust before steady state can be attained. Switching from growing InAs to InP will cause a delay in the initial growth when the In composition in the seed particle increases from 34\% to 44\%. Switching from growing InP to InAs results in excess In in the seed particle and no additional In source is needed to initiate growth. This was studied experimentally by growing short segments of InP in InAs nanowires and short segments of InAs sandwiched between two InP segments. The short InAs segments were grown from the excess In in the seed particle and only As was supplied externally. In Fig. 2.13(a) the temporal evolution of InP segments can be seen. For growth times longer than 20 s the length increases linearly, and a steady state is reached with a constant growth rate of 0.2 nm/s. In Fig. 2.13(b) the temporal evolution of short InAs segments grown from excess In in the seed particle is shown. Initially, the segment grows at a constant growth rate of 1 nm/s, but after 10 s the excess In in the seed particle is consumed and growth comes to a halt. The segment length is then 10.5 nm. This is qualitatively consistent with the model.

A quantitative analysis was performed in order to validate the model. Let us assume that the number of Au atoms in the hemispherically shaped seed particle is constant and denote the particle volume $V_{Au}$ when it con-
2.7. InAs/InP heterostructure growth

Figure 2.13: Initial temporal evolution of short InAs and InP heterosegments. (a) Temporal evolution of short InP segments in InAs nanowires. (b) Temporal evolution of short InAs segments sandwiched between two InP segments. The excess In in the seed particle allows for 10.5 nm growth of InAs without the supply of external In.

tains 44% In. The average volume of each atom in the particle is \( \Omega_{\text{Au}} = a_{\text{Au}}^3 / 4 \), assuming the particle has a face-centered cubic crystal structure with a lattice constant \( a_{\text{Au}} = 4.1 \) Å and In atoms replacing Au atoms. The total number of atoms in the particle is \( V_{\text{Au}} / \Omega_{\text{Au}} \). As the In content decreases to 34%, the number of In atoms (\( \Delta n_{\text{In}} \)) will change by \( 2.2 \times 10^5 \) atoms, but the radius will only decrease by 1.2 nm. It is believed that the seed particle does not shrink, but reshapes slightly maintaining the radius. As the In atoms crystallize in the nanowire, the same number of As atoms are incorporated into the nanowire. Thus, the In atoms leaving the seed particle can be converted into molecular volume elements, \( \Omega_{\text{InAs}} \), of In–As molecules in a wurtzite crystal with lattice constants \( a = 4.3 \) Å and \( c = 7.0 \) Å (Paper xix). The resulting length of an InAs cylinder with a hexagonal cross section is then given by:

\[
L = \frac{\Delta n \cdot \Omega_{\text{InAs}}}{A_{\text{nw}}}. \tag{2.22}
\]

For a 46 nm diameter nanowire and a change in In content from 44% to 34%, the resulting length is 10.2 nm. This agrees well with the experimentally determined length of InAs growth from excess In in a 45 nm diameter seed particle (10.5 nm, see Fig. 2.13(b)) and suggests that the model can fully describe the initial InAs growth.
Similarly, the time needed to increase the In content from 34% to 44% when switching from InAs to InP can be calculated. This involves the arrival rate of In atoms at the seed particle, from both direct impingement of In atoms in the beam, \( \Gamma_{\text{direct}} \), and from surface diffusion of In, \( \Gamma_{\text{diffusion}} \). The beam flux, \( R \), was estimated to be 1 nm\(^{-2}\)s\(^{-1}\) from bulk growth experiments, resulting in an In arrival rate due to direct impingement of \( R \cdot A_{\text{nw}} = 1.83 \cdot 10^3 \) s\(^{-1}\). It was found in previous experiments that direct impingement accounted for only 20% of the total In arrival rate at the seed particle (Paper I), thus, the total In arrival rate, \( \Gamma_{\text{total}} \), is 9.16 \cdot 10^3 \) s\(^{-1}\). The time required to increase the In concentration is given by:

\[
t = \frac{\Delta n}{\Gamma_{\text{total}}},
\]

which gives 24 s for an increase from 34% to 44% In. This is close to the experimentally determined time (20 s), needed to reach steady-state growth of InP, and again the model fits the experimental results well.
The physics of nanowires

Physics on the nanometer scale involves both quantum physics and classical physics. In this chapter, the use of electron transport through nanowires, measured at low temperature (4.2 K), is described. This is used to probe the energetics and electrostatics of quantum dots in nanowires is described. This chapter is based on Papers V–VII.

3.1 Device processing

After nanowire growth, the nanowires are transferred to an oxidized Si substrate, which is degenerately doped so as not to lose its conductivity at 4.2 K. The thickness of the thermally grown oxide layer is typically 100 nm. The substrate is pre-patterned with a coordinate system in order to be able to locate individual nanowires and define ohmic contacts to the ends. Prior to contact evaporation, the sample is exposed to NH$_4$S$_x$ to remove the oxide and passivate the surface with S atoms [36]. An InAs surface can create surface states in the conduction band, pinning the Fermi level above the conduction band edge [37]. Thus, electrons accumulate on the surface, which further simplifies ohmic contact formation and results in n-type conductance also for nominally undoped nanowires. The metals in the contacts are Ni/Au or Ti/Au, and the structure is illustrated in Fig. 3.1(a).
3.2 Constant interaction model

The two green segments in the nanowire illustrated in Fig. 3.1(a) indicate InP tunnel barriers separating an InAs quantum dot and the InAs source and drain leads. The quantum dot circuit is schematically shown in Fig. 3.1(b), including capacitances and tunnel barriers. Consider the situation where a small source–drain bias, $V_{ds}$, is applied and a small current is detected at low temperature. Electrons are tunneling from the source into the quantum dot and then out to the drain. If energy quantization is neglected there are many available states on the quantum dot for the electrons in the source to tunnel into. If the quantum dot is large, then several electrons can tunnel simultaneously without much interaction. But the smaller the quantum dot, the greater the Coulomb repulsion between electrons on the quantum dot [38, 39]. For quantum dots in nanowires, with dimensions of 10–100 nm and a diameter of 50 nm, the repulsion is so strong that when one electron has tunneled into the quantum dot no other electron can enter, at least at small biases and low temperatures. Thus, the electron on the quantum dot has to tunnel out of the quantum dot before the next one can tunnel in, one electron at a time. The above device is thus called a single-electron transistor (SET) [40]. When an electron is trapped on the quantum dot, it is said to be in a Coulomb blockade [41].

In the constant interaction model, the interaction between electrons is assumed to be independent of the number of electrons on the quantum dot, hence the name [42]. The interaction is described by capacitances, one for
3.2. Constant interaction model

each electrode, where the total capacitance of the isolated quantum dot
is called the self-capacitance, \( C_\Sigma = C_s + C_g + C_d \). This self-capacitance
determines the energy scale of the system through the charging energy

\[
E_c = \frac{e^2}{C_\Sigma}.
\]

For Coulomb interactions to be visible in measurements the charging energy
must be large compared to other energies, for example, the thermal energy,
\( E_c \gg kT \). For the devices described in this chapter, it is necessary to
cool them down to \(< 20 \, \text{K}\) to see this effect. Therefore, measurements are
usually performed at the temperature of liquid He, 4.2 K.

The energetics of a quantum dot is often described in terms of elec-
trochemical potential, \( \mu \). This can be visualized in a band diagram such
as that in Fig. 3.2(a)–(c), where the shaded areas indicate states that are
occupied by electrons. The electrochemical potentials of the source, drain,
and quantum dot are denoted \( \mu_s \), \( \mu_d \), and \( \mu_{QD}(N) \) when the quantum dot
contains \( N \) electrons. In the constant interaction model [42] the electro-
chemical potential of the quantum dot is given by:

\[
\mu_{QD}(N) = (N - 1/2) \frac{e^2}{C_\Sigma} - e \frac{C_g}{C_\Sigma} V_{gs}.
\]

The source–drain bias is related to the electrochemical potentials through
\( eV_{ds} = \mu_s - \mu_d \). In Fig. 3.2(a) the quantum dot contains \( N \) electrons
and the electrochemical potential that must be overcome to add one more
electron to the quantum dot, \( \mu_{QD}(N+1) \), denoted \((N+1)\) in the figure, is
higher than both \( \mu_s \) and \( \mu_d \). Thus, no electron can enter the quantum dot
and no current flows through the system. The quantum dot is said to be
in Coulomb blockade. Two ways of lifting the blockade are shown in Fig.
3.2(b) and (c). According to Eq. 3.2, \( \mu_{QD} \) can be varied continuously by
varying the gate voltage, and in Fig. 3.2(b) \( \mu_{QD}(N) \) has been moved into
the bias window, that is, it is between \( \mu_s \) and \( \mu_d \). In this situation one
electron on the quantum dot can tunnel into the drain and allow another
electron from the source to tunnel into the quantum dot. Thus, current
flows through the system even though \( V_{ds} \) is the same as in Fig. 3.2(a).
If the gate voltage is swept while maintaining \( V_{ds} \) small and constant, the
situations illustrated in Fig. 3.2(a) and (b) alternate with increasing or
decreasing \( N \), resulting in current peaks separated by regions of no current,
as illustrated in Fig. 3.2(d). The band diagrams of Fig. 3.2(a) and (b) are
indicated in Fig. 3.2(d). These current peaks are called Coulomb blockade peaks or Coulomb oscillations, and the separation between peaks is related to the gate capacitance through $\Delta V_{gs} = e/C_g$. Another way of lifting the blockade is by increasing $V_{ds}$, which eventually leads to the situation in Fig. 3.2(c). $\mu_{QD}(N + 1)$ is now aligned with $\mu_s$, and $\mu_{QD}(N)$ is aligned with $\mu_d$, and two electrons can tunnel through the quantum dot simultaneously.

Figure 3.2: (a) Band diagram for a quantum dot in Coulomb blockade with a small $V_{ds}$. (b) Band diagram for a quantum dot with an energy level in the small bias window, $(\mu_s - \mu_d)/e = V_{ds}$. (c) Band diagram for a quantum dot under a bias that exactly corresponds to the charging energy. This allows for transport through both levels simultaneously. (d) Sketch of Coulomb oscillations in the $I_d - V_{gs}$ plane. The spacing between peaks is $e/C_g$ and the band diagrams in (a) and (b) are indicated. (e) Sketch of stability diagram showing conductance in the $V_{ds} - V_{gs}$ plane. The shaded regions are regions where the quantum dot is blocked and the conductance is zero; in the white regions the conductance is non-zero. The band diagrams in (a), (b), and (c) are indicated by circles.

Quantum dots are often characterized by plotting the current or differential conductance as a function of $V_{gs}$ and $V_{ds}$, as illustrated in Fig. 3.2(e). These plots are called stability diagrams and display diamond shaped regions along the $V_{gs}$ axis (shaded areas), where the quantum dot is in Coulomb blockade and no current is allowed through. Outside these Coulomb diamonds, current can flow through the system. The band dia-
3.3. Single quantum dots in nanowires

grams in Fig. 3.2(a)–(c) are indicated by the circles.

For small semiconductor quantum dots, the effects of quantum confinement cannot be ignored. When the energy level spacing is of the same order as the charging energy, the energy level structure will influence the Coulomb oscillations and diamonds. According to the constant interaction model the electrochemical potential on the quantum dot can now be written:

\[ \mu_{QD}(N) = E_N + (N - 1/2) \frac{e^2}{C_{\Sigma}} - e \frac{C_g}{C_{\Sigma}} V_{gs} \]  (3.3)

The single-particle energy levels are simply added to the other energy contributions, which will lead to uneven spacing between the Coulomb peaks and differently sized Coulomb diamonds.

### 3.3 Single quantum dots in nanowires

Single InAs quantum dots were grown with lengths between 10 nm and 100 nm and InP tunnel barriers with thicknesses of 3–4 nm in 45 nm diameter InAs nanowires. The energy level spacing in the 100 nm long quantum dot is much smaller than the charging energy, and therefore the Coulomb peaks are periodic in \( V_{gs} \), as can be seen in Fig. 3.3(a). The level spacing is estimated to be less than 1 meV and the charging energy was measured and found to be 5 meV, while the gate capacitance was determined to be \( \sim 10 \text{ aF} \). This quantum dot has an unknown number of electrons, and if a more negative gate voltage is applied in order to empty the quantum dot, the leads of the nanowire will also be depleted of electrons, making it impossible to detect the last electron leaving the quantum dot. By growing a thinner quantum dot, the energy levels are raised making \( \mu_{QD}(1) \) higher in energy than \( \mu_s \) and \( \mu_d \). In Fig. 3.3(b) the Coulomb peaks for a 10 nm long disc-shaped quantum dot are shown. As the gate voltage is increased to \( V_{gs} = 1.3 \text{ V} \), the first state, which can contain two electrons, is filled. The distance between the second and third peak is larger than that between the two first peaks, indicating that the energy level spacing is of the same order as the charging energy. If the axial and the radial variables are assumed to be separable, the lowest energy levels are due to the radial quantization. The gate capacitance, determined from the distance between the peaks, is \( \sim 1 \text{ aF} \).

The interaction between energy level spacing and charging energy is
Chapter 3. The physics of nanowires

Figure 3.3: Coulomb oscillations for (a) a 100 nm long InAs quantum dot and (b) a 10 nm long InAs quantum dot. The 100 nm quantum dot exhibits strictly periodic oscillations due to charging energy, while the 10 nm quantum dot shows irregular spacing between peaks due to energy level separation and charging energy.

even more visible in a stability diagram such as that shown in Fig. 3.4 for the 10 nm quantum dot. For $V_{gs} < 1.3$ V the region of zero conductance widens and the stability diagram never closes again indicating that the quantum dot is completely depleted of electrons at $V_{gs} = 1.3$ V. This allows the number of electrons in each diamond to be determined, and in Fig. 3.4 the larger diamonds are numbered. A large diamond indicates that a specific electron state is full, and in order for the next electron to enter the quantum dot the charging energy plus the energy level spacing must be overcome. The energy level structure can be mapped out in this way. All the energy levels are spin degenerate, so filling an odd number only requires the charging energy and thus, all odd-numbered diamonds should be similar in size. The charging energy can be determined from the tip of an odd-numbered diamond, which is equivalent to the situation in Fig. 3.2(c); $E_c = \mu_s - \mu_d = eV_{ds} = 5$ meV. Assuming that the source and drain capacitances are the same this gives $C_s = C_d = 15$ aF.

Semiconductor quantum dots are sometimes called artificial atoms. The name originates from the fact that the energy level structure (shell structure) of a quantum dot can be resolved, just as in normal atoms. A small periodic table has even been suggested for a certain type of quantum dots [42]. As these artificial atoms are man-made they can be designed to have different energy level structures depending on their geometric shape. It is important to note that these artificial atoms usually consist of > 100 atoms, and the characteristics of artificial atoms are affected by the indi-
3.4 Imaging single quantum dots

In the previous section, it was described how the electrochemical potential of the quantum dot could be varied continuously by a global backgate. This not only affects the quantum dot but also the source and drain parts of the nanowire. Fabricating local gates to a pre-defined quantum dot in a nanowire is possible, but it is difficult due to the uncertainty in the exact position of the quantum dot. In this section the use of a scanning probe microscope, with a conducting tip in order to locally gate the quantum dot is described. The charge induced on the quantum dot by the tip is given by: \( q_{QD}(V_{tip}, r_{tip}) = C_{tip}(r_{tip}) \cdot V_{tip} \), where the capacitance between the tip and the quantum dot is dependent on the distance between them.

This device is similar to the one in Section 3.3, with the addition of a hovering tip, Fig. 3.5(a). The height of the tip over the substrate is set to a fixed value, greater than the nanowire diameter, in order to avoid direct contact between the tip and the nanowire. The tip is scanned above the quantum dot at constant \( V_{tip} \) and the conductance through the quantum dot is detected for a small bias as a function of tip position. The result is a set of concentric rings of high conductance centered on the quantum dot, and two such measurements are shown in Fig. 3.5(b) and (c). The red areas indicates the position of the nanowire, the width of which corresponds to the diameter of the nanowire (50 nm). In these measurements the backgate

\[ V_{ds} \text{ (mV)} \]
\[ V_{gs} \text{ (V)} \]

**Figure 3.4:** Stability diagram for a 10 nm long few-electron quantum dot, clearly showing shell structure with “full shells” for electron numbers 2, 6, 8, 12, 16, and 22. The first electron enters at \( V_{gs} = 1.3 \) V. The darker the color the higher the differential conductance, blue shows negative differential conductance, and white regions of Coulomb blockade.

individual atoms.

3.4 Imaging single quantum dots

In the previous section, it was described how the electrochemical potential of the quantum dot could be varied continuously by a global backgate. This not only affects the quantum dot but also the source and drain parts of the nanowire. Fabricating local gates to a pre-defined quantum dot in a nanowire is possible, but it is difficult due to the uncertainty in the exact position of the quantum dot. In this section the use of a scanning probe microscope, with a conducting tip in order to locally gate the quantum dot is described. The charge induced on the quantum dot by the tip is given by: \( q_{QD}(V_{tip}, r_{tip}) = C_{tip}(r_{tip}) \cdot V_{tip} \), where the capacitance between the tip and the quantum dot is dependent on the distance between them.

This device is similar to the one in Section 3.3, with the addition of a hovering tip, Fig. 3.5(a). The height of the tip over the substrate is set to a fixed value, greater than the nanowire diameter, in order to avoid direct contact between the tip and the nanowire. The tip is scanned above the quantum dot at constant \( V_{tip} \) and the conductance through the quantum dot is detected for a small bias as a function of tip position. The result is a set of concentric rings of high conductance centered on the quantum dot, and two such measurements are shown in Fig. 3.5(b) and (c). The red areas indicates the position of the nanowire, the width of which corresponds to the diameter of the nanowire (50 nm). In these measurements the backgate
and tip voltages are tuned to probe the zero-to-one-electron transition, as indicated by the numbers 0 and 1 in Fig. 3.5(b).

![Schematic view of a nanowire quantum dot with the scanning probe tip hovering above. (b) Scan of conductance rings for $V_{tip} = -2.5$ V and tip height of 100 nm above the substrate. Blue shows zero conductance, and yellow/brown high conductance. (c) As in (b) but with $V_{tip} = -2.0$ V. The scale in the two scans is the same.](image)

**Figure 3.5:** (a) Schematic view of a nanowire quantum dot with the scanning probe tip hovering above. (b) Scan of conductance rings for $V_{tip} = -2.5$ V and tip height of 100 nm above the substrate. Blue shows zero conductance, and yellow/brown high conductance. (c) As in (b) but with $V_{tip} = -2.0$ V. The scale in the two scans is the same.

The rings correspond to Coulomb peaks, now displayed as a function of tip position rather than $V_{gs}$. In Fig. 3.5(b), $V_{tip} = -2.5$ V, and as the tip is moved away from the quantum dot $\mu_{QD}$ decreases, allowing the first electron to enter the quantum dot. As $V_{tip}$ becomes less negative, $\mu_{QD}$ decreases, resulting in smaller rings, Fig. 3.5(c) ($V_{tip} = -2.0$ V). In the corners of the images the ring due to the second electron entering the quantum dot can be seen. If the tip is scanned along the ridge of a ring, $\mu_{QD}$ is constant, i.e. the rings are contours of constant tip–quantum dot interaction. The shape of the rings depends on the shape of the quantum dot. In this case, the quantum dot is 18 nm long, which is about the size of the tip, and the rings appear circular. For a longer dot the rings will be more elongated and could show structure due to the spatial electronic variations inside the quantum dot [43].

The tip can also be used to produce stability diagrams, similar to that shown Fig. 3.4. In Fig. 3.6 the tip was positioned 70 nm above the quantum dot and the stability diagram shows Coulomb diamonds similar to those in Fig. 3.4. The obvious difference between the two stability diagrams is that the clear shell structure seen in Fig. 3.4 seems to have vanished in Fig. 3.6. This is believed to be the result of the potential from the tip changing the confining potential of the quantum dot as $V_{tip}$ is changed. As a result, the symmetry of the quantum dot is broken and the shell structure disappears.

From the size of the Coulomb diamonds $C_{tip}$ was estimated to be $\sim 0.4$
3.5. *Double quantum dots in nanowires*

![Stability diagram of a single quantum dot using the tip of a scanning probe microscope as the gate.](image)

**Figure 3.6:** Stability diagram of a single quantum dot using the tip of a scanning probe microscope as the gate. The shell structure is not as clear as when the backgate is used, which is probably related to a change in the confining potential as $V_{\text{tip}}$ is varied.

$a_F$ and $E_c$ to be $\sim 5$ meV. The presence of the tip does not change the charging energy significantly since the capacitance between the tip and the quantum dot is small compared to the self-capacitance of the quantum dot when the tip is not present.

### 3.5 Double quantum dots in nanowires

If single quantum dots can be regarded as artificial atoms, then double quantum dots can be seen as artificial molecules, with energy levels extending over both dots [39]. The “bond strength” depends on the width of the middle barrier separating the two quantum dots [39]. A narrow barrier leads to strong coupling between the quantum dots while a wide barrier weakens the coupling. In double quantum dot systems where the barriers are defined by local gates [44], the coupling between the quantum dots can be varied continuously, and such systems are being considered for quantum computing applications [45]. In the double quantum dots discussed here, the inter-dot coupling is defined by the width of the middle InP tunnel barrier defined during growth, and is not tunable. On the other hand, these double quantum dots are more robust and may be interesting in resonant tunneling applications.

The nanowires are fabricated in the same way as the single quantum
dots, resulting in a global backgate controlling both quantum dots simultaneously, but not individually. A backgate sweep results in Coulomb peaks similar to those in Fig. 3.3, but much more irregular, since now both quantum dots must have energy levels more or less aligned in the bias window in order for electrons to flow. The gate sweeps are very difficult to interpret but some understanding can be obtained from stability diagrams.

The stability diagram in Fig. 3.7(a) was obtained for an asymmetric double quantum dot. A scanning TEM image of the double quantum dot is shown in Fig. 3.7(b). The lengths of these quantum dots were determined to be 9 nm and 18 nm for dot 1 and dot 2, respectively. Due to quantum confinement, the larger quantum dot (dot 2) is expected to contain more electrons and its energy levels to be lower than the smaller one (dot

\[ V_{G_S} \text{ (V)} \]

\[ V_{ds} \text{ (mV)} \]
3.5. Double quantum dots in nanowires

1). Changing \( V_{gs} \) will affect the electrochemical potential of the two quantum dots simultaneously, but differently, since the capacitance between the backgate and the quantum dot depends on the length of the dot. The stability diagram in Fig. 3.7(a) closes at \( V_{gs} = 0.9 \) V, that is, the conductance of the quantum dot is non-zero for \( V_{ds} = 0 \) V. This indicates that the lowest energy level in dot 1 aligns with a level in dot 2 and with \( \mu_s \) and \( \mu_d \) (see the top energy band diagram). At this point, dot 2 contains an unknown number of electrons. For more positive values of \( V_{gs} \) Coulomb diamonds are seen due to filling of dot 1 (not shown). These diamonds are ruptured along the borders of the diamonds due to the more frequent charging of dot 2. Here, the focus is on the lower part of the stability diagram \( (V_{gs} < 0.9 \) V). The green dotted lines indicate where the lowest level in dot 1 aligns with the electrochemical potential of the source. Electrons can then tunnel into dot 1, then into dot 2, and then into the drain. The tunneling between the quantum dots does not necessarily require perfect alignment due to thermal broadening of the energy levels and the fact that inelastic tunneling can also occur [39]. Towards the lower right of these “green” borders \( V_{ds} \) increases and \( V_{gs} \) becomes less positive in order to maintain the alignment between \( \mu_s \) and \( \mu_{dot1} \) (1). Eventually, \( \mu_d \) will be sufficiently low to allow one of the electrons stored in dot 2 to tunnel out into the drain. This de-charging of dot 2 will affect \( \mu_{dot1} \) since electrons tunneling into dot 1 now experience less Coulomb repulsion from the electrons stored in dot 2. Thus, \( \mu_{dot1} \) is shifted down in energy causing a kink in the borderline. To re-align \( \mu_s \) and \( \mu_{dot1} \) the voltages must be adjusted, for example, by moving along one of the borderlines indicated by a red dotted line. These lines indicate alignment between \( \mu_d \) and \( \mu_{dot2}(N) \). Thus, at every kink in the borderline one electron leaves dot 2 [46] and we can trace 13 such kinks before the stability diagram opens up without any more kinks. To the right of the stability diagram, three band diagrams are shown for three different positions in the stability diagram. The lowest diagram shows the band structure of the first kink, when the alignment of \( \mu_s \) and \( \mu_{dot1}(1) \), and of \( \mu_d \) and \( \mu_{dot2}(1) \) occurs simultaneously. The middle diagram shows the band structure when dot 2 is being charged with its 6th electron and the top diagram shows the band structure when dot 2 contains 13 electrons at \( V_{ds} = 0 \) V.

The distance between the green dotted lines reflects the inter–dot charging energy, whereas the distance between the red dotted lines depends on the charging energy of dot 2. Larger distances are seen between the red
lines for electron numbers 2, 6, and 10, which reflects the energy level structure of dot 2. This corresponds to larger Coulomb diamonds in a stability diagram for a single quantum dot, cf. Fig. 3.4.
4.1 Field effect transistors

The electronic evolution, from the first FET [47] to today’s circuits with hundreds of millions of transistors [3], has revolutionized the world in many ways. Most semiconductor electronics is based on transistors, connected in various ways to perform logic functions. An FET is a three-terminal device in which the current between two of the contacts (the source and the drain) is controlled by an intermediate capacitively coupled electrode (the gate) [47]. These transistors are fabricated on a wafer using processes such as UV-lithography, ion implantation, and metal deposition. A schematic cross section of an n-type metal oxide semiconductor FET (MOSFET) is shown in Fig. 4.1. The source and drain contacts are deposited directly on highly n-doped regions to minimize contact and series resistances. The area along the top surface separating the two regions is the channel. The gate stack is positioned between the source and drain contacts, directly
over the channel, and consists of the gate oxide and the gate metal. The oxide has traditionally been SiO$_2$, but recently Intel and IBM announced the introduction of hafnium-based oxides (HfO$_2$) into production [3, 48]. The reason is that miniaturization has led to a gate oxide thickness of only 1.2 nm [3], and extensive tunneling from the gate to the channel leads to heating and large leakage currents. By using HfO$_2$, which has a higher relative dielectric constant $\kappa$, the gate oxide can be made thicker while maintaining the capacitance to the channel. This leads to reduced tunneling through the gate oxide. The gate contact has traditionally been made of polysilicon, but as dimensions are shrinking, it is expected that metallic materials will be used in the future [3].

![Cross section of a planar n-type MOSFET.](image)

**Figure 4.1:** Cross section of a planar n-type MOSFET.

The MOSFET in Fig. 4.1 is fabricated on a p-type substrate in order to create a potential barrier between electrons in the n$^+$ regions. Electrons in the source region must overcome this barrier in order to reach the drain region, even at $V_{gs} = 0$ V. This transistor is therefore normally off, and is called an enhancement mode transistor. If the channel region had been n-type, a current would easily have flowed from the source to the drain at $V_{gs} = 0$ V, and the channel would normally have been on. This type of transistor is called a depletion mode transistor. Enhancement mode transistors are more convenient to work with from a circuit design perspective.

A vertical cross section through the gate stack and the channel, showing the energy band structure, is shown in Fig. 4.2(a). The conduction band edge, $E_C$, the valence band edge, $E_V$, the intrinsic Fermi level, $E_i$, and the Fermi level, $E_F$, are indicated in the illustration. The intrinsic Fermi level is almost halfway between $E_C$ and $E_V$, and the Fermi level is determined by the substrate doping. The situation in Fig. 4.2(a) is referred to as the flat band condition. In order to create a channel for electrons, a positive potential is applied to the gate. This will lower the gate potential and the
4.1. Field effect transistors

bands in the semiconductor will bend as in Fig. 4.2(b). When the difference in energy between \( E_i \) and \( E_F \) at the interface between the oxide and the semiconductor is the inverse of that in the bulk, the semiconductor changes locally from being p-type to n-type, and inversion takes place. Electrons are attracted from the bulk of the semiconductor and accumulate at the semiconductor–oxide interface. Thus, a channel of conducting electrons is created, and electrons can now easily flow from the source to the drain, which would be in the direction perpendicular to the paper.

![Figure 4.2: (a) Band diagram of a metal oxide semiconductor junction in the flat band condition. (b) Band diagram for the same junction with a positive potential on the gate. This lowers the gate potential, and the bands in the semiconductor bend downwards resulting in the accumulation of electrons at the semiconductor–oxide interface.](image)

The current from the source to the drain through a long-channel device in the on state \( (V_{gs} > V_{th}) \) is purely resistive for small \( V_{ds} \), and the current increases linearly with the bias applied according to:

\[
I_d = \frac{W \mu C_{ox} L}{V_{gs} - V_{th}} V_{ds},
\]

where \( W \) is the width of the device, \( L \) is the channel length, \( \mu \) is the channel mobility, and \( V_{th} \) is the threshold voltage, which defines whether the transistor is on or off [7]. Comparing Eq. 4.1 to Ohm’s law [49], \( I = 1/R \cdot U \), shows that a higher \( V_{gs} - V_{th} \) corresponds to a lower resistance. This is seen in the linear region (left part) of the output characteristics shown in Fig. 4.3(a).

As \( V_{ds} \) increases, the transistor eventually reaches saturation \( (V_{ds} > V_{sat}) \) and the current no longer increases, see Fig. 4.3(a). The reason for current saturation depends on the device geometry. For a long-channel device, the electric field from the drain depletes the carriers on the drain side of the channel, destroying the inversion. This situation is called pinch-off,
because the width of the inversion layer on the drain side of the channel is reduced to zero [47]. Any increase in the drain voltage will cause the pinch-off point to move towards the source, leading to an increase in the resistance between the pinch-off point and the drain. This will balance the higher voltage so that the current is constant or saturated. For short-channel transistors, saturation can also be caused by electron velocity saturation. At high electric fields along the transport direction, the scattering rate of electrons increases, resulting in reduced mobility and finally saturation of the velocity [47]. For even shorter channel lengths, where the mean free path of the electrons is comparable to the channel length, the electrons move ballistically across the channel [50]. The electron transport in the channel is then limited by the injection velocity from the source, and the drain current is dependent on how many modes are available for transport.

In Fig. 4.3(b), both the linear and logarithmic transfer characteristics of a MOSFET are shown. The I–V characteristics above threshold (\(V_{gs} > V_{th}\)) are dependent on the geometry of the transistor. For a long-channel device the saturation current increases as \(I_d \propto (V_{gs} - V_{th})^2\), and for a short-channel device, where electron velocity saturation occurs, it increases as \(I_d \propto (V_{gs} - V_{th})\), for \(V_{ds} > V_{sat}\). The transconductance of the transistor is defined as the slope of this curve, \(g_m = \partial I_d / \partial V_{gs}\), and is an important metric of transistors.

Below threshold, the gate lifts the bands in the channel and electrons in the source region can only move to the drain region via thermal excitation over the channel barrier and diffusion along the channel. The current
therefore has an exponential relation to \( V_{gs} - V_{th} \):

\[
I_d \propto e^{q(V_{gs} - V_{th})/mkT},
\]

where \( m \) is an ideality factor, \( m \leq 1 \). One of the most important metrics for benchmarking transistors is the inverse slope in this subthreshold region; i.e. the amount \( V_{gs} \) must be varied in order to change \( I_d \) one order of magnitude. For conventional transistors, the ideal inverse slope is \( kT Q \ln 10 \approx 60 \text{ mV/decade} \) at room temperature, and is determined by the thermal smearing of the Fermi distribution. As \( V_{gs} \) becomes more negative in the subthreshold region, the drop in gate–drain potential increases. For narrow bandgap materials this can lead to band-to-band tunneling, where electrons in the valence band tunnel across the bandgap ending up in the conduction band. The result is an increase in the \( \text{off} \) current and worsening of the subthreshold characteristics. An alternative type of transistor that exploits the band-to-band tunneling to circumvent the 60mV/decade limit, is the tunneling FET. The transistor is turned on and off by enabling and disabling band-to-band tunneling in a \( p-i-n \) diode by a control gate in the \( i \) region [51]. This has also been successfully implemented in a nanowire transistor [52].

### 4.2 Device processing

In this chapter three different nanowire FETs will be presented. The starting point of these devices is a conducting InAs(111)B substrate with positioned Au particles, similar to that described in Section 2.5.1. In all three devices the substrate is used as the source contact, but different growth parameters and processing conditions are used in the different devices. Layouts for the three devices, presented in Sections 4.3, 4.4, and 4.5, are shown in Fig. 4.4(a), (b), and (c), respectively, and are hereafter called (a), (b), and (c). Note that the layouts in Fig. 4.4 are not on the same scale.

In order to obtain as clean a channel surface as possible, the gate dielectric is deposited directly after growth. In device (a), a 50 nm thick sputtered SiN\(_X\) layer is used, and in devices (b) and (c), 10 nm of HfO\(_2\) is deposited by atomic layer deposition (ALD).

In devices (a) and (c), the gate is formed by metal sputtering. The gate length is defined by spin-coating the sample with a polymer, etching it back using O\(_2\) plasma and then using it as an etch mask in wet etching of the metal. The polymer protects the metal on the substrate surface and on the
lower parts of the nanowires, and thus the resulting gate length is defined by the polymer etch-back. In devices (a) and (c), the gate metal is deposited directly on top of the gate oxide, but in device (b) a 100 nm thick SiO$_X$ insulator is inserted to further separate the gate from the substrate (source). During the evaporation of the SiO$_X$ the sample is tilted and rotated, which results in a SiO$_X$ layer on the substrate and in porous SiO$_X$ on the nanowire side facets. The material on the sides can be etched by quickly dipping it in HF, effectively cleaning the nanowire sides. In device (b), half the gate oxide is deposited before SiO$_X$ evaporation and the second half directly after the HF dip. The chromium gate metal is then formed, by vaporization, similar to the SiO$_X$, and the gate length is determined by the length of the deposition process. This is a great advantage over the process used to fabricate devices (a) and (c), and over conventional gate length fabrication using lithography. After gate formation, the gate oxide at the top of the nanowires is removed in a similar way as for the sputtered gate.

In Fig. 4.5(a) a nanowire similar to devices (a) and (c) can be seen. The nanowire was initially covered with HfO$_2$ and sputtered Cr, which was
4.3. Long-channel transistors

sequentially removed at different heights using a polymer as an etch mask. The gate lengths were determined from similar images to be $\sim 1 \, \mu\text{m}$ and $\sim 100 \, \text{nm}$ for device (a) and (c), respectively. In Fig. 4.5(b) a cross-sectional view of device type (b) can be seen directly after gate formation. Both the SiO$_X$ and the Cr are visible, and both show thinning close to the nanowire due to self-shadowing during evaporation, but no residues are visible on the nanowire sides. The HfO$_2$ gate oxide is visible as thin bright lines above and below the SiO$_X$.

![Figure 4.5](image)

**Figure 4.5:** (a) SEM image showing a nanowire covered with Cr and HfO$_2$ up to certain heights. (b) SEM image showing the SiO$_X$ spacer layer and the Cr gate. Both layers show thinning close to the nanowire due to self-shadowing during evaporation. (c) The air-bridge drain contact of device (a).

The gate–drain separation is a spin-on polymer in devices (b) and (c), whereas in (a) the polymer was removed after a thick drain contact was deposited, leaving an air-bridge. The advantage of using air is that it has a low dielectric constant, $\kappa = 1$, which minimizes stray capacitances. IBM has announced that air gaps will be used to reduce interconnect capacitances [53]. However, nanowires are fragile, and small mechanical movements of the bridge will break the nanowires. Therefore, in devices (b) and (c), the polymer was left in place to support the drain contact and the nanowires.

All measurements were made on the devices at room temperature, unless otherwise specified.

4.3 Long-channel transistors

The output characteristics of the long-channel device (device (a)) are shown in Fig. 4.6(a). The device shows good current saturation, starting at
Chapter 4. Device applications of nanowires

\(V_{ds} = 0.2\) V. For \(V_{ds} > 0.6\) V the output conductance increases, probably due to band-to-band tunneling in the gate–drain region. The current is shown both as the absolute value and normalized to the total nanowire circumference of the device. The normalized values allow the transistor to be compared to other devices.

\(\begin{align*}
\text{Figure 4.6: (a) Output characteristics of the long-channel device. (b) Transfer characteristics of the same device plotted on a logarithmic scale showing an inverse subthreshold slope of 100 mV/decade.}
\end{align*}\)

The transconductance was found to be 2 mS or 200 mS/mm. In Fig. 4.6(b), the transfer characteristics of the same device are shown, plotted on a logarithmic scale, and the inverse subthreshold slope is 100 mV/decade for \(V_{ds} = 0.4\) V. By assuming a \(I_d \propto V_{gs}^2\) relation the threshold voltage was determined to be \(-0.15\) V.

The device in this section had a gate length of \(\sim 1\) µm for nanowires with a diameter of 80 nm. Theoretical modeling suggests that a gate length/nanowire diameter ratio of around 1–1.5 is necessary for the transistor to be electrostatically well designed [54]. Making the ratio even smaller will worsen the performance and short-channel effects will become visible. The gate dielectric in this device, SiNX, deposited by chemical vapor deposition, is known to contain trapped charges related to hydrogen incorporated in the material, which can cause hysteresis and poor radio frequency performance [55]. Therefore, the gate oxide was changed in the following devices to HfO2, which ideally has a dielectric constant of 20–25 compared to SiNX, which has \(\sim 7\) and SiO2, which has \(\sim 4\).

52
4.4 Short-channel transistors

In this device the gate formation technique was changed to evaporation, where the gate length is more or less directly controlled by deposition, instead of etch-back as in the previous device. The gate length in this device was 50 nm and the nanowire diameter 50 nm. Furthermore, a spacer layer between the source (substrate) and the gate was introduced to reduce parasitic capacitances. The top part of the nanowire was also n-doped with Se during growth, see Fig. 4.4(b), as indications were observed that this reduced the resistivity by a factor $>2$.

![Output characteristics of the short-channel device.](image)

Fig. 4.7 shows the output characteristics of the device. The saturation is not as pronounced as in the previous device, which is probably due to the much smaller gate length/diameter ratio. As the gate length/diameter ratio becomes smaller the potential control in the channel becomes more difficult. As $V_{ds}$ is increased the electric field from the drain also lowers the bands in the channel, which lowers the source–channel barrier. This effect is called drain-induced barrier lowering (DIBL) and results in higher output conductance than normal [7]. At $V_{ds} > 0.6$ V the output conductance increases, which is probably due to the result of increasing band-to-band tunneling, similar to the case of the long-channel device. The absolute and the normalized current is higher than in the long-channel device, which is expected since these devices are much shorter in length and the nanowires are n-doped in the drain region.

The transfer characteristics are shown in Fig. 4.8(a) and (b), on linear...
Figure 4.8: (a) Linear transfer characteristics and (b) logarithmic transfer characteristics of the short-channel device. The threshold voltage was determined to be 0.1 V and the inverse subthreshold slope 88 mV/decade for $V_{ds} = 0.5$ V and 103 mV/decade for $V_{ds} = 1.0$ V.

and logarithmic scales respectively. The threshold voltage $V_{th}$ was determined from the region of highest transconductance and found to be 0.1 V. $V_{th}$ is shifted relative to the long-channel device without changing the doping in the channel. Both devices are nominally undoped in the channel, but other factors also affect the carrier concentration, such as nanowire diameter, gate oxide and other materials in close proximity to the nanowire. The threshold voltage is positive and the transistor therefore operates as an enhancement mode transistor. The highest transconductance value in these devices was 445 mS or 520 mS/mm for $V_{ds} = 0.5$ V, which is higher than the long-channel device. For $V_{ds} = 1.0$ V, a value of 800 mS/mm was determined, which is comparable to values found in other InAs nanowire devices [56]. These values are so far lower than today’s high electron mobility transistors, which usually show transconductance values of around 2000 mS/mm [57], but the comparison is not straightforward due to the difference in device geometry.

The subthreshold characteristics seen in Fig. 4.8(b) show an inverse slope of 88 mV/decade for $V_{ds} = 0.5$ V. By following the procedures in [58], $I_{on}$ is measured at $V_{gs} = V_{th} + 2/3 \cdot V_{ds}$ and $I_{off}$ at $V_{gs} = V_{th} - 1/3 \cdot V_{ds}$, resulting in $I_{on}/I_{off}$ ratios around $10^3$. The off current saturates at $V_{gs} = -0.5$ V and only a small increase in the current is observed for more negative
4.5 Heterostructure transistors

$V_{gs}$, thus no strong indication of ambipolar conductance was seen, as in carbon nanotubes [59].

4.5 Heterostructure transistors

As shown in Chapter 2, atomically sharp InAs/InP heterostructures can be grown without the formation of extended crystal defects, despite the 3.1% lattice mismatch. InAs/InAs$_{1-x}$P$_x$ heterostructures have also been reported, with controlled and homogeneous P content from 10% to 100% [60]. Since the current direction is the same as the growth direction, the band structure can be varied quite freely along its path. In this device a 50 nm long undoped InAs$_{0.8}$P$_{0.2}$ segment was grown in an otherwise Se-doped InAs nanowire, see Fig. 4.4(c). The InAs$_{0.8}$P$_{0.2}$ segment creates a 130 meV conduction band barrier, as determined from thermal excitation measurements. In order to ensure that the gate overlaps the entire InAs$_{0.8}$P$_{0.2}$, a slightly longer gate ($\sim$ 100 nm) was fabricated using the same technique as in the long-channel device described in Section 4.3. To be able to determine the effects of inserting a barrier in the channel, two devices were processed in parallel, one in which the shaded region in Fig. 4.4(c) was undoped InAs and one in which it was undoped InAs$_{0.8}$P$_{0.2}$.

Fig. 4.9(a) and (b) show the output characteristics of the InAs and the InAs$_{0.8}$P$_{0.2}$ devices respectively. Both devices show good saturation and similar normalized currents. The most pronounced difference is the large increase in output conductance seen for the InAs device at $V_{gs} = -0.5$ V and $V_{ds} > 0.5$ V, which was not observed in the InAs$_{0.8}$P$_{0.2}$ device. In these devices the gate length is long enough to suppress short-channel effects, and effects such as DIBL cannot explain the data shown in Fig. 4.9.

In these two devices the gate–drain region is similar, see the shaded area in Fig. 4.9(c) and (d). Thus, the same number of electron–hole pairs is created in the two devices due to band-to-band tunneling. However, band-to-band tunneling is still believed to be the cause of the difference in output characteristics. The holes created in this process drift back into the channel region and may become trapped close to the channel. These holes gate the channel, which leads to lowering of the bands and an increase in the subthreshold current for both devices, but the effect is more severe in the InAs device. The reason is that the gate voltage at which the band-to-band tunneling becomes non-negligible is far into the subthreshold region for the InAs$_{0.8}$P$_{0.2}$ device, but much closer to $V_{th}$ for the InAs device. That
Chapter 4. Device applications of nanowires

Figure 4.9: (a) Output characteristics of an InAs nanowire transistor. (b) Output characteristics of an InAs$_{0.8}$P$_{0.2}$ nanowire transistor. $V_{gs}$ was stepped from −0.5 V to 0.5 V in steps of 0.2 V for both devices. (c) Sketch of the band diagram for the InAs device and the effect of feedback gating in the channel (dashed lines). (d) Sketch of the band diagram for the InAs$_{0.8}$P$_{0.2}$ device and the effect of feedback gating in the channel (dashed lines).

is, the Fermi level is much closer to the conduction band edge in the InAs device than in the InAs$_{0.8}$P$_{0.2}$ device. Therefore, a shift in the bands will lead to a much greater increase in the electron concentration and current in the InAs device than the InAs$_{0.8}$P$_{0.2}$ device. This effect is probably most severe in the bulk of the nanowire since the negative gate lifts the bands at the surface. Thus, the increase in current in the InAs device is not the current from band-to-band tunneling but, as described above, the resulting feedback gating of the holes created in this process. A similar feedback mechanism has been seen in InAs/AlSb FETs [61]. The suggested mechanism is supported by quasi-3D modeling in ATLAS by Silvaco [62].

Furthermore, the negative gate potential could cause an inversion at the nanowire side facets, which would further screen the bulk of the nanowire from the gate potential. This effect would be worse for the InAs device due to the narrower bandgap.
Figure 4.10: (a) Transfer characteristics of the InAs device (dashed line) and InAs$_{0.8}$P$_{0.2}$ device (solid line) on a linear scale. The threshold voltages were determined to be $-0.3$ V and $0.2$ V for the InAs and InAs$_{0.8}$P$_{0.2}$ devices, respectively. (b) Transfer characteristics of the InAs device (dashed line) and InAs$_{0.8}$P$_{0.2}$ device (solid line) on a logarithmic scale. The inverse subthreshold slope is 210 mV/dec and 120 mV/dec for the InAs and InAs$_{0.8}$P$_{0.2}$ devices, respectively.

The transfer characteristics of the two devices are plotted on linear and logarithmic scales in Fig. 4.10(a) and (b). The dashed lines represent the InAs device and the solid lines the InAs$_{0.8}$P$_{0.2}$ device, where the current has been normalized in order to allow direct comparison. The threshold voltages were found to be $-0.3$ V for the InAs device and $0.2$ V for the InAs$_{0.8}$P$_{0.2}$ device. The reason for the shift is the change in Fermi level pinning and the presence of a barrier in the conduction band. For InAs the Fermi level is typically pinned in the conduction band [37, 63], whereas in InP the Fermi level is pinned below the conduction band edge [64, 63]. It is reasonable to believe that as the P content increases in InAs$_{1-x}$P$_{x}$ the Fermi level pinning gradually moves from above the conduction band to below it, and that this leads to a lower carrier concentration. Thus, $V_{th}$ can be controlled by the P content in the channel, without changing the doping.

The subthreshold region shows a remarkable difference in minimum current levels, as can be seen in Fig. 4.10(b). The current at $V_{gs} = 0.5$ V is 4 orders of magnitude lower in the InAs$_{0.8}$P$_{0.2}$ than in the InAs device. The maximum on/off current ratio for a gate swing of 0.5 V increases from 30 to 6000 when the barrier is introduced. This can also be explained by the feedback gating from holes created in the band-to-band tunneling process.
In research, one is constantly faced with questions and problems requiring answers and solutions. Answering one question usually leads to new questions, and so the field of research widens continuously. Below are some remarks on further optimizations and issues that need consideration in the future of nanowire research.

As an experimentalist, I have constructed models in an attempt to explain experimental results, and not vice versa. The difficult part in developing models is deciding what to include and what to leave out. If too many parameters are included in a model it will be possible to explain any experimental observation, but the risk of it being wrong is large. The beauty of a simple model that only takes the most relevant physical processes into account is, in my opinion, underestimated. That is one of the reasons why three different models were presented in Chapter 1. They could certainly be combined in such a way as to be helpful in future experiments.

The development of a process for the growth of short InAs and InP heterostructures, presented in Chapter 1 will be very helpful for future advanced heterostructure designs. The heterostructure control on the atomic scale in bulk semiconductor growth has, for example, enabled extreme devices such as the quantum cascade laser. Furthermore, implementing such a device in nanowires requires a high-quality, preferably epitaxial, high-bandgap shell around the nanowire [65]. The shell moves surface charges far away from the active part of the device. This is necessary since the random surface charges create irregularities in the potential, destroying the band structure design. A similar solution would also benefit studies of electron transport at low temperatures, presented in Chapter 2, and the room-temperature transport in the vertical wrap gate transistors, presented
in Chapter 4. A high-bandgap shell would result in less scattering and possibly higher electron mobility. At the same time, adding layers between the channel and the gate will move the gate further away from the channel, decreasing the gate control. This is a trade-off that must be optimized. So far, radial heterostructures have not been included in these devices, but this will be necessary in order to obtain high-performance transistors. Furthermore, other axial heterostructures than the one presented in Chapter 4 could be designed to improve transistor performance. The work on heterostructure transistors presented in this thesis is merely in the beginning of this field.

Generally speaking, I believe that nanowires have great potential for future devices in various fields. The unique property of nanowires is their extreme geometry, being almost one-dimensional. Among other things, this allows for effective strain relaxation, and thus the number of materials that can be epitaxially combined can be drastically increased. This will lead to greater freedom in band structure design. However, in many material combinations, the band structure alignment remains unknown. The effective strain relaxation is also very interesting from a commercial perspective, since it allows III–V nanowires to be grown epitaxially on Si, reducing production costs. Growth of III–V nanowires on Si has already been demonstrated [66, 67, 68], but the growth control must be improved.
Populärvetenskaplig sammanfattning


I denna avhandling presenteras ett alternativ till den konventionella transistorn med syftet att få fram strömsnålar och snabbare elektronik. Majoriteten av de transistorer som produceras idag är tillverkade i det halvledande materialet kisel (Si) och strömkanalen är placerad horisontellt. Vi har utvecklat en transistor med vertikal strömkanal av ett annat halvledande material, indiumarsenid (InAs). Strömkanalen utgörs av en tunn tråd, så kallad nanotråd. Skillnaderna kan tyckas vara petitesser, men har i själva verket en avgörande inverkan på transistoregenskaperna. Fördelen med InAs jämfört med Si är bl a att strömmen flyter mycket lättare i InAs, man säger att elektronerna har högre mobilitet. Detta gör att transistorer gjorda av InAs kan drivas med lägre spänning och de drar därmed mindre energi samtidigt som hastigheten ökar. Varför görs då inte alla transistorer av InAs? Jo, kostnaden! Att göra transistorer helt i InAs är normalt sett mycket dyrare än Si. En stor del av kostnaden ligger i startmaterialet, en så kallad wafer, som ofta måste vara i samma material.
som transistorn. Fördelen med att göra transistorer av nanotrådar är att dessa kan tillverkas, med hög kvalitet, på en Si wafer. Detta tar bort en stor del av kostnaden vilket gör tekniken intressant.

De transistorer som presenteras i denna avhandling är på forskningsstadiet och har stor förbättringspotential, men deras karaktäristik är trots detta jämförbar med dagen kommersiella transistorer som finns i datorer, som bygger på en teknik som mognat i mer än 30 år.

För att kunna tillverka en nanotrådstransistor krävs givetvis att man kan göra en nanoträd. Dessa framställs genom kristallväxt och har en hög kristallkvalitet, dvs, indium- och arsenikatomerna sitter varvade i en periodisk struktur. Helst skall nanotrådarna bara finnas i på förhand utvalda områden. Dom skall ha rätt längd och diameter, och leda ström bra. Detta kräver grundläggande studier av nanotrådsväxt och elektrontransport genom nanotrådarna. I kapitel 2 i avhandlingen presenteras de studier av nanotrådsväxt som har utförts och i kapitel 3 finns de grundläggande fysikstudierna. Tillverkning och karaktärisering av nanotrådstransistorerna presenteras i kapitel 4.


References


References


