Transmission electron microscopy characterization of quantum dot based intermediate band solar cells

Espen Undheim

Physics
Submission date: May 2014
Supervisor: Randi Holmestad, IFY
Co-supervisor: Ragnhild Sæterli, IFY

Norwegian University of Science and Technology
Department of Physics

Preface

This master thesis is the result of work done between August 2012 and June 2014, as part of a two-year master degree in Physics at the Norwegian University of Science and Technology (NTNU). This thesis was done in association with the TEM Gemini center at the Department of Physics. The TEM gemini center is a collaboration between the Department of Physics, Department of Materials Science and Engineering and SINTEF Materials and Chemistry, Trondheim.

I would like to thank my supervisor Randi Holmestad, for her encouragement and guidance during my thesis work. During our weekly meetings, her enthusiasm always made my day a bit brighter.

I would also like to thank my co-supervisor Ragnhild Sæterli, Per Erik Vullum and Antonius T. J. van Helvoort for their help, TEM-training and for always being available if I had any questions.

A special thanks to my family for their continual support and for always being there if I need someone to talk to.

Trondheim, May 29th, 2014

Espen Undheim

Abstract

In this thesis, two samples of quantum dot based intermediate band solar cells were studied and compared by transmission electron microscopy. These samples were grown on a (100) GaAs substrate, with a general structure consisting of 50 InAs quantum dot layers, where each quantum dot layer was separated by a spacer layer. The difference between the two samples was that one sample had GaAs spacer layers, while the other had GaNAs spacer layers. The N was added as a strain compensator, and it was therefore expected that this sample would exhibit less defects.

By bright field (BF) and scanning transmission electron microscopy (STEM) the crystal structure of both samples were studied. This showed that the sample containing no N had fewer defects, compared to the other sample.

Using energy dispersive X-ray spectroscopy and electron-probe micro analysis it was determined that the N content in both samples was either below the detection limit of 200 ppm or zero.

The quantum dot sizes were then found for both samples by STEM and it was seen here that the sample supposedly containing N, had on average higher quantum dot sizes.

The better crystal structure of the sample containing no N, were attributed to the lower average quantum dot sizes. As the N content of both samples was determined to be insignificant, the only likely cause for the difference in size where attributed to a difference in growth parameters.

Using electron energy loss spectroscopy and BF imaging, the quantum dot sheet density was found. The sheet density of both samples was around 10^{11} cm^{-2}, with the sample containing no N having a factor 2 higher density compared to the other sample.

Lastly, the polarity for both samples was determined by convergent electron beam diffraction, and these results were confirmed by a high resolution STEM image.

Sammendrag

I denne avhandlingen ble to prøver av kvanteprikk basert mellombånd-solceller studert og sammenlignet med transmisjonselektronmikroskopi. Disse prøvene ble grodd på en (100) type GaAs-substrat, med en generell struktur bestående av 50 InAs kvanteprikk lag, hvor hvert kvanteprikk lag var separert med et mellomlag. Forskjellen mellom de to prøvene var dette mellomlaget, der en prøve hadde GaAs mellomlag og den andre hadde GaNAs mellomlag. N ble tilsatt for å redusere belasting som bygger seg opp i materialet på grunn av en forskjell i gitterparameter, og det ble derfor forventet at denne prøven skulle inneholde mindre defekter.

Ved bruk av lysfelt, og skanning transmisjonselektronmikroskopi (STEM) ble krystallstrukturen til begge prøvenen undersøkt. Det ble vist at prøven som ikke inneholder N hadde færre defekter i forhold til den andre prøven.

Ved hjelp av røntgenspektroskopi og elektronsondemikroanalyse ble det fastslått at N -innholdet i begge prøvene var enten var under deteksjonsgrensen på 200ppm eller null.

Kvanteprikk størrelsene ble deretter funnet for begge prøvene ved STEM og det ble sett her at prøven som angivelig inneholdt N, hadde i gjennomsnitt høyere kvanteprikk størrelse.

Den bedre krystallstrukturen til prøven som ikke inneholdt noe N, ble tilskrevet den lavere gjennomsnittlige kvanteprikk størrelsen. Ettersom N innholdet i begge prøvene var ubetydelig, ble denne forskjellen i kvanteprikk størrelse tilskrevet en forskjell i vekstparametre.

Elektron energitap spektroskopi og lysfelt bilder ble brukt til å finne kvanteprikk tettheten for de to prøvene. Tettheten av begge prøvene var rundt 10^{11} cm^{-2}, men prøven som angivelig inneholdt N hadde en faktor 2 lavere tetthet.

Til slutt ble polariteten for begge prøvene bestemt ved konvergerent elektronstråle-diffraksjon, og resulatene fra dette ble bekreftet med å sammenligne med et høy oppløsning STEM bilde.

List of abbreviations

ABF Annular bright field
ADF Annular dark field
AlGaAs Aluminium Gallium Arsenide
BF Bright field
CB Conduction band
CBE Conduction band edge
CBED Convergent electron beam diffraction
DF Dark field
ED Electron diffraction
EDS Energy dispersive X-ray spectroscopy
EELS Electron energy loss spectroscopy
EPMA Electron-probe micro analysis
FOLZ First-order Laue zone
Ga(N)As Gallium (Nitrogen) Arsenide
GaN Gallium Nitride
HAADF High angle annular dark field
HRTEM High-resolution transmission electron microscopy
HOLZ Higher-order Laue zone
IB Intermediate band
IBSC Intermediate band solar cell
InAs Indium Arsenide
LAADF Low angle annular dark field
ML Monolayer
MBE Molecular beam epitaxy
PV Photovoltaics
QD Quantum dot
SAD Selected area diffraction
TEM Transmission electron microscope
SADP Selected area diffraction pattern
STEM Scanning transmission electron microscope
SK Stranski-Krastanov
SOLZ Second-order Laue zone
VB Valence band
VBE Valence band edge
VLM Visible Light Microscope
ZOLZ Zero-order Laue zone

Contents

Preface i
Abstract iii
Abstract v
List of abbreviations vii

1 Introduction **1**
1.1 Motivation 3

2 Theory **7**
2.1 Solar cells 7
2.1.1 Loss Mechanisms in solar cells 11
2.1.2 Ideal solar cell efficiencies 15
2.1.3 Intermediate band solar cells 16
2.1.4 Quantum dot intermediate band solar cells . . 18
2.2 Materials 23
2.3 Transmission electron microscopy 25
2.3.1 Resolution limit 26
2.3.2 The Instrument 27
2.3.2.1 Electron sources 27
2.3.2.2 Electron optics 29
2.3.3 Electron-matter interactions 32
2.3.4 Diffraction 34
2.3.4.1 Bragg's law 35
2.3.4.2 Laue equations and the Ewalds sphere 36
2.3.4.3 Atomic form factor and structure factor 39
2.3.5 Contrast 40
2.4 Imaging & diffraction techniques 42
2.4.1 Bright field and dark field 43

2.4.2 Scanning transmission electron microscopy . . . 45
2.4.2.1 Contrast in STEM mode . . . 47
2.4.2.2 Spherical aberration . . . 48
2.4.3 Electron diffraction . . . 49
2.5 Spectroscopy techniques . . . 55
2.5.1 Energy dispersive X-ray spectroscopy . . . 55
2.5.2 Electron-probe micro analysis . . . 56
2.5.3 Electron energy loss spectroscopy . . . 56
2.6 Molecular beam epitaxy . . . 59
2.7 Notation . . . 61

3 Method and experiment **62**
3.1 Samples . . . 62
3.2 Sample preparation . . . 64
3.3 Experimental equipment . . . 67

4 Results and discussion **69**
4.1 Structure . . . 69
4.1.1 Defect structures and QD stacking . . . 75
4.2 The hunt for the missing Nitrogen . . . 85
4.3 Quantum dot size and density . . . 93
4.4 HRTEM imaging of QDs . . . 103
4.5 Polarity determination . . . 118

5 Conclusion **123**

6 Further work **125**

Chapter 1

Introduction

As the living standards and the world population increases, so does the energy consumption. Today about 80% of the energy production is through the use of fossil fuels, i.e. coal, oil and gas [1]. These contribute to the release of harmful gasses, which includes CO_2, into the atmosphere. With evidence piling up that climate change is a result of human influences, and the fact that the amounts of fossil fuels are finite, the distribution of energy sources must change. The energy distribution as of 2010 can be seen in figure 1.1.

The energy production must shift to nuclear and renewable energies. The potential of nuclear energy (not counting fusion) is rather limited, as this is also a finite energy source. Combining this with the fact

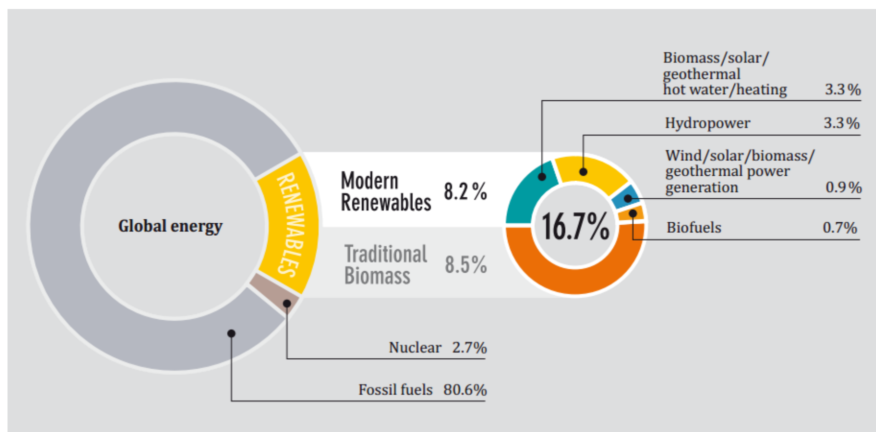


Figure 1.1: The global energy distribution as of 2010 [1]

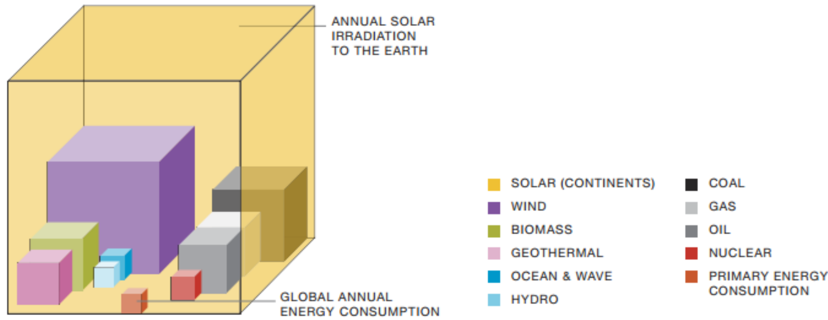
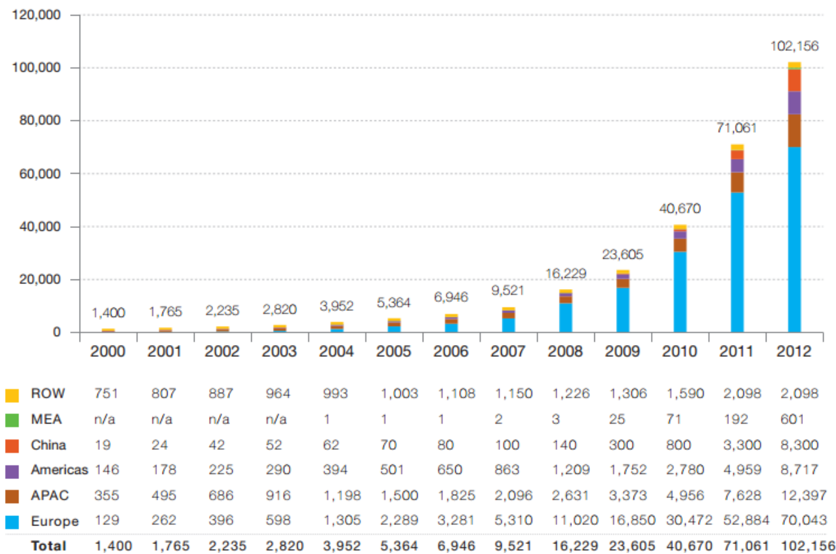


Figure 1.2: The energy potential of various energy sources. For the fossil fuels and nuclear energy, the energy potential is based on their total reserves, while for renewable energies the size indicates their yearly potential [2]

that the by-products are highly dangerous, due to radiation, indicates that the shift should be mainly focused towards renewable energies, e.g. wind, hydro, or geothermal power.

Another source of clean renewable energy is solar energy. As of 2010 solar power only accounted for 3.3% of the global energy production, but it's potential is much greater. This is represented visually in figure 1.2, and from this it can be seen that the potential for solar energy dwarfs all the others. The total amount of solar radiation impinging on earth could, with current technology meet the global energy needs 10,000 times over [2], given that every continent were covered by solar cells or solar thermal heaters. These two ways are the two primary methods for extracting solar power. Solar cells utilize the incoming photons to generate electricity, while solar thermal heaters can be used in different ways. These include passive heating of houses or by using a solar collector to generate hot water that can be used for heating houses or in hot water systems, e.g. shower water. Another application is to generate electricity by heating water beyond the boiling point, and then using the steam to generate electricity.

Of the two, the one dominating the market is photovoltaics (PV). PV has experienced a massive growth the latest year, where at the end of 2009 the world's cumulative installed PV capacity approached 24 Giga Watt (GW) and just three years later this number had reached over a 100 GW [3]. This growth is shown in figure 1.3 for the period 2000-2012. Still, the cost of electricity from solar power isn't able to compete with the price of electricity from fossil fuels. Though it is



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
ROW	751	807	887	964	993	1,003	1,108	1,150	1,226	1,306	1,590	2,098	2,098
MEA	n/a	n/a	n/a	n/a	1	1	1	2	3	25	71	192	601
China	19	24	42	52	62	70	80	100	140	300	800	3,300	8,300
Americas	146	178	225	290	394	501	650	863	1,209	1,752	2,780	4,959	8,717
APAC	355	495	686	916	1,198	1,500	1,825	2,096	2,631	3,373	4,956	7,628	12,397
Europe	129	262	396	598	1,305	2,289	3,281	5,310	11,020	16,850	30,472	52,884	70,043
Total	**1,400**	**1,765**	**2,235**	**2,820**	**3,952**	**5,364**	**6,946**	**9,521**	**16,229**	**23,605**	**40,670**	**71,061**	**102,156**

Figure 1.3: Evolution of global PV cumulative installed capacity 2000-2012. The y-scale is in units of mega watts (MW) [3]

close in countries where the price of electricity is high and the amount of solar radiance is high. In order to reduce the cost per watt for solar cells, new technologies are needed.

1.1 Motivation

Most solar cells on the market currently are Silicon (Si) based, either multi-crystalline or single-crystalline. These cells belong to a category called first generation solar cells. The maximum possible efficiency of these cells are about 30 - 40 % [4], depending on the concentration of light, but currently the majority of cells on the market have an efficiency around 20 %. Here the efficiency is defined as

$$\eta = \frac{E_{electricity}}{E_{incident}} \tag{1.1}$$

where $E_{electricity}$ is the energy extracted from the solar cell and $E_{incident}$ is the energy of the incident solar radiation.

As the technology of these Si based cells have matured, the majority of the cost is being associated with the starting materials. These are

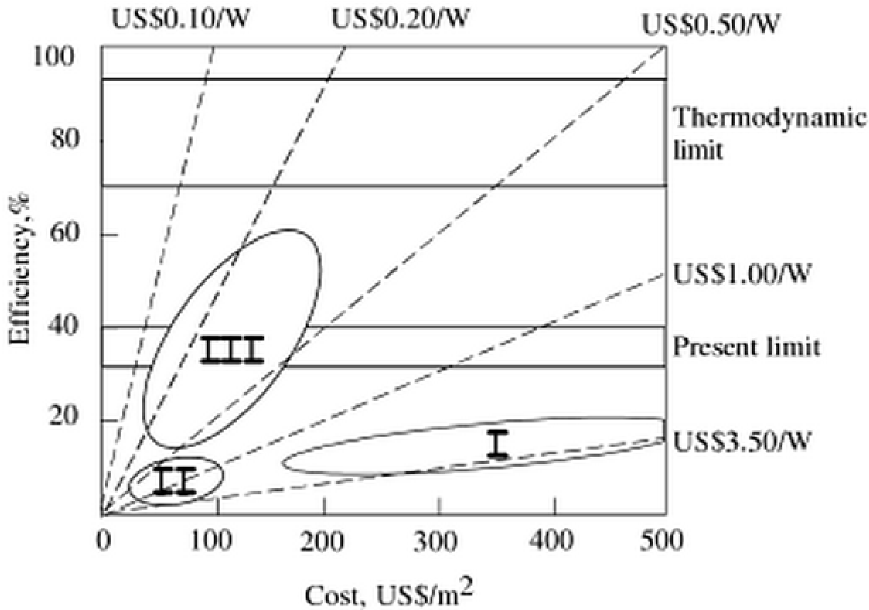


Figure 1.4: The projected cost of the different generations of solar cells, with the areas of efficiency and cost defining each generation indicated by I, II and III [5]. This was published in 2003, and so the dollar values does not take inflation into account

already mass produced and there is therefore not much room for improvement, and the costs will eventually reach a certain limit. Thus to reduce the cost, either the materials needed to produce a cell must be reduced, or the efficiency must be increased. Second generation cells are defined as those that reduce cost by the first principle, and third generation cells are a combination of the two. The projected costs for the three different generations of solar cells are shown in figure 1.4.

The second generation cells are also referred to as thin-film solar cells and because they use less material, the cost per m^2 will be lower. These cells are also limited by the same efficiency as first generation cells, and so the price per watt can only go so low.

The efficiency of a solar cell will depend on its surface area, as the energy possible to harvest will increase with increasing surface area. Therefore, if the efficiency can be doubled, the cell will produce the same amount of electricity, with only half of the surface area. This

reduces the amount of material needed, reducing the cost per m^2

Third generation cells utilizes a higher efficiency, coupled with thin-film technology and the reduction of surface area, in order to achieve low costs. This has yet to be realized, and most third generation solar cells are still at the laboratory stage. Currently the most realized third generation cells are the multi-junction cells, but these are still very expensive, even though the efficiency have reached up to 44,7% [6] for the record holding cell.

Another third generation concept that is promising are the intermediate band solar cells, which are the focus of this thesis. These cells are able to increase the efficiency by utilizing more of the sunlight. There are several ways of creating this type of solar cells, but currently the most studied method is through the use of quantum dots.

Quantum dot based intermediate band cells require a high density of quantum dots in order to have sufficient absorption of light. This necessitates the need for growing several layers. In this thesis the cells made were based on the Indium Arsenide (InAs) / Gallium Arsenide (GaAs) material system, which has a large lattice mismatch. Structures of these materials will therefore be strained. The situation is only worsened as the number of layers increases. Due to an increase in the strain energy, defects will start to form as the number of layers increase. These defects reduces the strain energy, but they also reduce the efficiency of the cell.

As a way of compensating for the build up of strain, Nitrogen has been added to one of the samples. The aim of this thesis is therefore to look at structural differences and parameters such as size and density differences in the quantum dots, in the two samples. These parameters will be related to composition and growth parameters.

These materials will be investigated by a (scanning) transmission electron microscope. This instrument is ideal for studying nanoscale structures and material interfaces at high spatial resolution. It is also possible to perform compositional analysis of a material in the same instrument, making this instrument an excellent tool for investigating and understanding these materials.

The information gathered will help growers expand their knowledge about these materials and the growth process, hopefully leading to better quality materials with higher efficiencies.

The thesis is divided into three main parts, chapter 2 on theory, chapter 3 on the experimental details, and lastly chapter 4 is the

results and discussion.

In chapter 2 the underlying theory of the concepts used and studied in this thesis will be explained. These include the working principles of solar cells, loss mechanisms and efficiency limits in solar cells, and the working principle behind the intermediate band solar cells, especially the quantum dot based intermediate band solar cells. The techniques and instruments used to study these materials will also be explained. Finally the growth mechanism for these samples and the material parameters and properties will be described.

In chapter 3 the experimental details are given. Here information about how the samples were grown, and growth parameters, are given. Also included here, are the sample preparation procedure and the experimental equipment used.

In chapter 4 the results obtained are presented and discussed. The results presented first are just a general structure for both samples, showing how the grown materials fit with the expected structure. Next the defect structures present in these materials are showed and discussed in detail. The chemical composition of both samples are then determined and discussed. After this the results relating to the quantum dot sizes are presented. Then a more in depth results and discussion part on dislocations in quantum dots follows. Finally, the results from the polarity determination are shown.

Chapter 2

Theory

This section will present some of the theoretical background for the project. This includes the working principles of solar cells, intermediate band solar cells, and transmission electron microscopy. The last one in particular will be described very thoroughly and this description will also include the various techniques that has been used on the TEM during this master thesis. Most of the information regarding TEM is from Williams and Carter's book on Transmission electron Microscopy [7]. The working principles of both molecular beam epitaxy (MBE), electron-probe micro analysis (EPMA), and photoluminescence (PL) will also be described briefly. For further reading on the working principles of solar cells, beyond what is described in this section, see Jenny Nelson's book, *The Physics of solar cells* [8].

2.1 Solar cells

The basic property of a solar cell is the ability to convert photons to an electric current. Most solar cells utilize the inherent properties of semiconductors to accomplish this. At zero kelvin semiconductors have a completely filled valence band (VB) and an empty conduction band (CB). Separating these two bands is a "forbidden" energy band, called the band gap. The size of the band gap is defined as the difference in energy between the conduction band edge (CBE), the lowest energy state of the CB, and the valence band edge (VBE), the highest energy state in the VB. Depending on how these align, the material may be a direct or an indirect band gap semiconductor. This is shown in figure 2.1a and 2.1b.

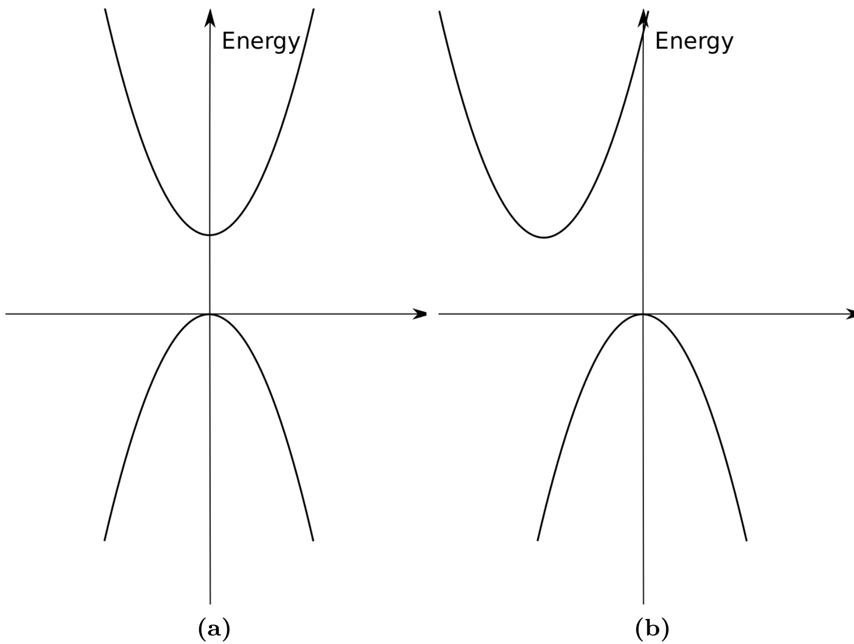


Figure 2.1: Simple model of a band structure for a **(a)** direct band gap material and an **(b)** indirect one. The real band structure will be much more complicated, this is just to show the alignments of the top and bottom of the two bands. The horizontal-component in these graphs is called the wavevector. The wavevector is proportional to the momentum of the electron.

The energy level in the middle of the band gap, for a pure semiconductor at 0° Kelvin, is called the Fermi level. The Fermi level is defined as the energy level where the probability of finding an electron is 50 %. This level will change depending on various parameters, e.g. temperature and dopant concentration. As the Fermi level in a material has to be constant under equilibrium conditions, the band diagram can be tailor-made by doping different regions of the material differently. The most standard of this application is the creation of p-n junctions in conventional solar cells.

The process by which current is created in a solar cell material is called the photovoltaic effect. The principle behind this effect is the excitation of an electron through the absorption of a photon. The absorption of a photon excites an electron in the VB into the CB. This electron leaves behind a hole in the VB, which is an empty energy state. This hole will allow electrons in the VB to move by "jumping" to this

empty state. When a electron "jumps" it leaves behind a new empty state. By this mechanism the electrons in the VB can move and there will be a current in the material. Instead of regarding this current as electrons "jumping" from state to state, it can be thought of as a single particle, a hole, moving in the VB. The electron in the CB is free to move, as most of the states here are unoccupied. Both the electron and the hole will therefore act as charge carriers in the solar cell. Through the photovoltaic effect the absorbed photon has therefore created two charge carriers, often referred to as an electron-hole pair (e-h pair).

To extract current from the solar cell the e-h pairs need to be separated. If they are not separated the e-h pair will quickly recombine. The various types of recombination processes are explained in detail in section 2.1.1. The separation of the e-h pairs is achieved by creating an internal electric field. This is done by creating a junction in the material between differently doped regions. The two types of doping are called n-type and p-type doping and the dopant atoms are referred to respectively as donor and acceptor atoms. This junction is therefore called a p-n junction. The electric field will be situated at the junction and will extend into the p and n side. How far it extends into either side is dependant on the dopant concentration, and it will not necessarily be symmetrical.

The n-type side is doped with atoms that has one extra electron (or more) compared to the host material, i.e. it "donates" electrons to the host material. The donor atom introduces states in the band gap, and how close these states are to the CB are determined by the type of atom used. It is preferential to choose donor atoms that have states close to the CB, as this will allow all the electrons to be excited into the conduction band at room temperature by thermal excitations. This will also reduce so called non-radiative recombination, and the reason for this will be explained in section 2.1.1. This type of doping raises the Fermi level towards the CB, see figure 2.2.

P-type doping is when the dopant atom has one electron (or more) less than the host material. The donor atoms "accepts" electrons from the VB by thermal excitations, leaving behind holes in the VB that acts as charge carriers. These dopant atoms creates states in the band gap, and for the same reasons as in the n-type case these states should be close to the VB. In this case the Fermi energy level is shifted towards the VB, see figure 2.2.

The p-n junction can be divided into three regions. A region where

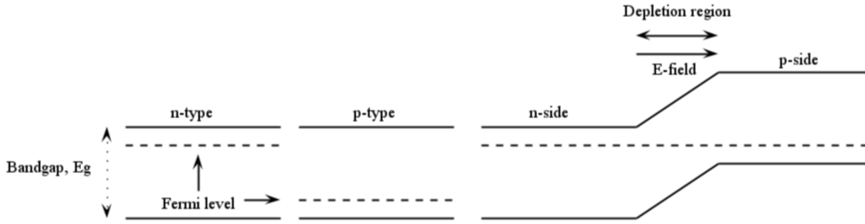


Figure 2.2: To the left is the band diagram for physically separated p- and n-type semiconductors. To the right is the energy band for a p-n junction. This is the case when the p and n side are connected. Any e-h pairs created in, or near, the depletion region will be separated by the electric field. The direction of the electric field is from the n-side to the p-side. Electrons, being negative particles, moves against the field and will therefore move towards the n-side. Whereas holes can be seen as positive particles and will therefore move with the field, towards the p-side.

electrons are the majority charge carriers, the n-side, a region where the holes are the majority charge carries, the p-side, and lastly the depletion region. The depletion region will contain no charge carriers in equilibrium and hence the name. These regions can be seen in figure 2.2. The depletion region is created due to the diffusion of charge carriers towards regions of lower hole and electron concentrations, i.e. holes towards the n-side and electrons towards the p-side. The diffusion of the charge carriers leaves behind positively charged impurity ions on the n-side and negatively charged impurity ions on the p-side. This creates a charge inequality and a electric field is set up, going from the n-side to the p-side. The size of this region is mainly determined by the dopant concentrations. This is illustrated in 2.3. The e-h pairs created in this region will be separated and can therefore contribute to the current.

When the solar cell is illuminated by light, e-h pairs will be generated and those generated in the depletion region will be separated by the electric field. The electrons, moving against the field, will move towards the n side and the holes will move towards the p side. This is called the drift current. There will also be a current of holes towards the n side and electrons towards the p side because of concentration differences. This diffusion current will be small, but the charge carriers that move across the depletion region will recombine and thus lowering the concentration even further. Due to these currents there will be a net increase in the population of electrons and holes on the n side and

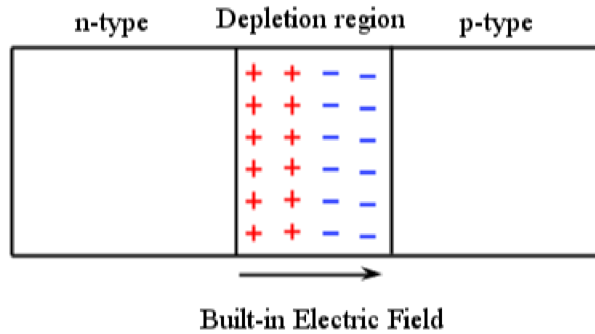


Figure 2.3: A typical p-n junction. Because of the diffusion of charge carriers, there will be a region containing no charge carriers called the depletion region. This diffusion leaves behind ions of the impurity atoms, and as a consequence of this there will be a charge difference and a electric field. This electric field is useful for separating the e-h pairs that are created in this region.

p side respectively. Because the relaxation time for the electrons in the CB, on the n-side, are much lower compared to the relaxation time across the depletion region, they can be considered to be in thermal equilibrium in the CB. The same is true for holes in the VB on the p-side. These two sides will now have different Fermi-levels, and these levels are referred to as quasi-Fermi levels with the notation E_{F_n} for the n side and E_{F_p} for the p side. The difference in the quasi-Fermi levels will determine the maximum output voltage of the solar cell. The energy band diagram for non-equilibrium conditions is shown in figure 2.4.

2.1.1 Loss Mechanisms in solar cells

The efficiency of a solar cell is determined by how much of the incoming energy it can transform into usable electric current. This efficiency is limited by several factors and different solar cell design have been made that try to reduce losses from different effects. The different losses can be explained by following the path a photon takes before it can be extracted as current.

First an incoming photon will have a probability of being reflected. By reflection, the amount of photons that are available are reduced and this will therefore constitute a loss in energy and efficiency. This situation can be improved by etching the surface, usually either in

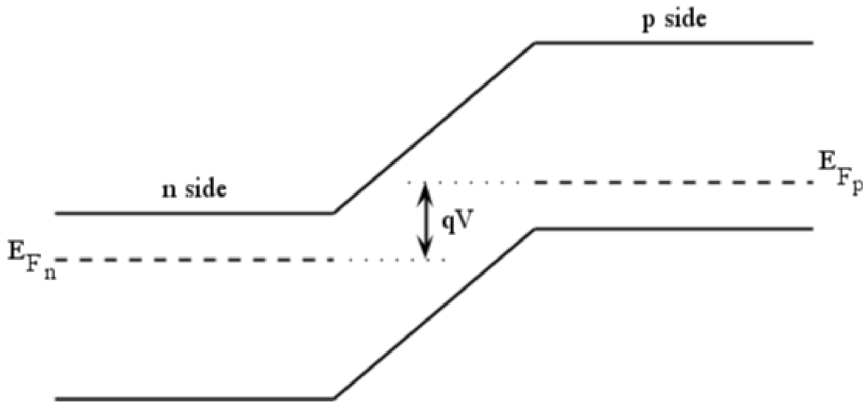


Figure 2.4: The energy band diagram for non-equilibrium conditions in a p-n junction. This is the case when light is shining on the cell, causing the electron and hole populations on the n and p side to change. Both of these charge carriers will reach a thermal equilibrium on their respective side, establishing different quasi-Fermi levels. The output voltage of the cell, V, is also shown schematically as the difference in these two levels.

pyramid shapes or inverted pyramids. This "traps" the light which reduces the amount that is reflected. Another method that is often used in conjunction with this is to coat the top of the solar cell with an anti-reflection coating(ARC), usually an oxide layer. This ARC will have a refractive index between that of air and the solar cell material. The reason for this is that the reflectance, the amount of the incoming light that is reflected, is dependant on the difference in refractive index. When this difference is small, so will the reflectance be. This can be regarded as a way of "easing" the passage of the light through the material.

After the photon has passed the ARC and entered the material it can be absorbed. Assuming that the material has been made thick enough most, if not all, of the incoming photons will be absorbed. One loss factor at this stage is that photons with energy below the band gap will not be absorbed, and will therefore not contribute to the current. This effectively removes part of the solar spectrum, reducing the efficiency. Even photons that are absorbed represent a loss in efficiency. The losses here will depend on how much greater the energy of the photon is compared to the band gap energy. How far the electron is excited into the CB, and holes into the VB, will depend on the energy of the photon. Since most states in the CB are unoccupied, the

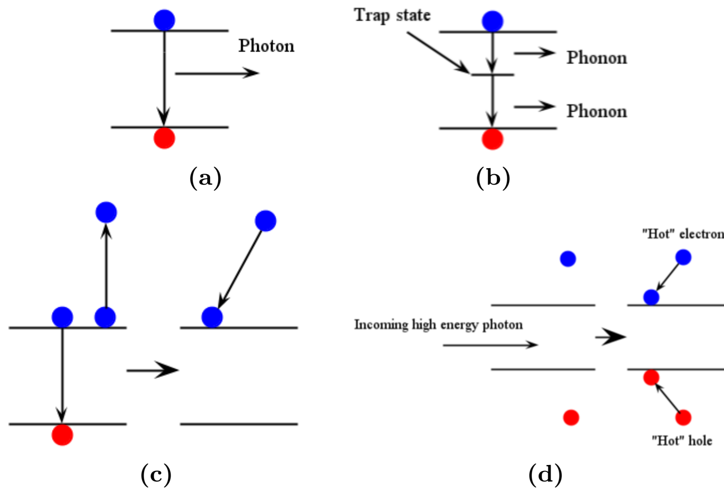


Figure 2.5: **(a)** Radiative recombination. The recombination of an e-h pair resulting in the creation of a photon. This photon will have the energy equal to the bandgap energy **(b)** Non-radiative recombination. The recombination of an e-h pair through a trap state in the band gap. At each step a phonon is created. **(c)** Auger recombination. This is similar to the case of radiative recombination, only instead of creating a photon the energy from the recombination is given to another electron in the CB. **(d)** Hot carriers. A high energy photon hits a electron in the VB causing it to be excited high into the CB. The electron (and hole) then relaxes down to the conduction band edge (valence band edge) by thermalisation.

electrons will quickly relax into these lower energy states (or holes into higher energy states in the VB). Thus some of the energy from these photons will be lost through the relaxation of these charge carriers. The energy lost in this relaxation process is through thermalisation. In this process the electrons or holes loose energy through the creation of phonons. Electrons (holes) that are excited high into the CB (VB) are called hot carriers and represent a large loss in efficiency. This process is very quick ($\backsim 10^{-15}$ s) and it is therefore very hard to utilize these "hot" carriers. See figure 2.5d for a simple picture of this process.

When the photon has been absorbed and a e-h pair has been created there are several processes in which the electron and hole can recombine. This means less charge carriers and therefore less current, which translates into a loss in efficiency. These types of recombination can be summarized as follows:

- Radiative recombination. This is the case when an electron and a hole recombine, creating a photon in the process. The created photon will have an energy equal to the band gap energy. This is shown in figure 2.5a. This process will be significant in direct band gap semiconductors, but in indirect band gap materials it is insignificant and usually neglected.

- Non-radiative recombination. In this process the e-h pair will recombine through an intermediate energy state, called a trap state, in the band gap. This process is shown in figure 2.5b. These states are introduced either because of dislocations in the material or from impurity atoms. The rate at which the carriers enters the trap states are dependant on the distance from the band edges. A trap state close to the conduction band edge will quickly be filled with electrons from the CB, but the rate of recombinations will not be large. The reason for this is that the difference between the trap state and the valence band edge is large, so the rate at which the electron goes from trap state to valence band edge is low. This is why trap states near the middle of the band gap have the highest recombination rate. As the electron falls into the trap state, and from there to the VB, energy is released through the creation of phonons. This loss can be reduced by reducing the amount of defects and impurities in the material through better fabrication methods.

- Auger recombination. This process is very similar to the case of radiative recombination. The difference here is that instead of creating a photon, the extra energy is given to another electron in the CB. This causes it to be excited further above the bandgap, the extra energy lost here is through the same process as for hot carriers (figure 2.5c.

Assuming the e-h pairs haven't recombined, they will be separated by the electric field and move towards the n and p side respectively. Due to the concentration difference there will be a diffusion current of holes towards the n side and electrons towards the p side. These charge carriers are called minority charge carriers, as they are a minority on these sides. These minority charge carriers will cause recombinations. This effect can be reduced by including a highly doped region between the contact and the n, or p, side called a back surface field (BSF). This region will represent a barrier to the minority charge carriers, but not

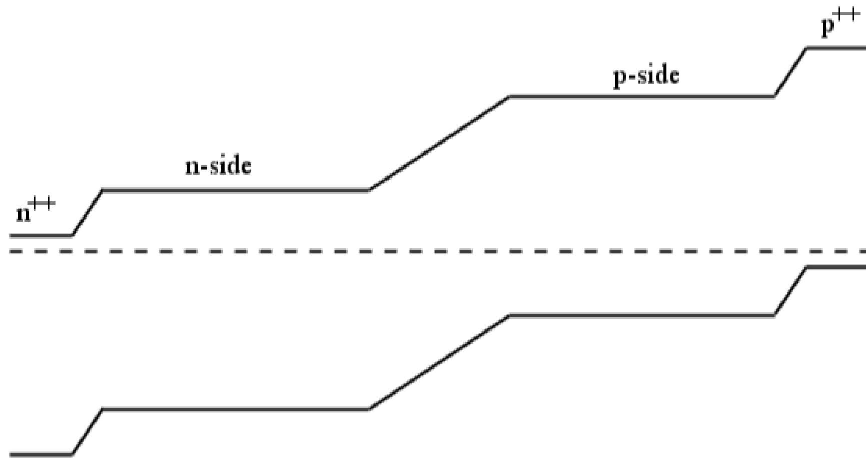


Figure 2.6: The band diagram for a p-n junction with a BSF on both sides. These highly doped regions, with notations p++ and n^{++}, represent barriers for the minority charge carriers and there will therefore be less recombinations.

to the majority charge carriers. This reduced the amount of minority charge carriers here and therefore there will also be less recombination. This is shown in figure 2.6.

In addition there will also be some losses at the contacts, and from the contacts placed on the front of the solar cell. The last one is due to shading, which effectively reduces the area of the cell, and will depend on how the contacts are integrated into the cell.

2.1.2 Ideal solar cell efficiencies

The first calculations on the theoretical maximum efficiency for a ideal single band solar cell were first done by Shockley and Quiesser in 1961 [4]. For these calculations they assumed the following: All photons with energy greater than the band gap are absorbed, no non-radiative recombination, the only loss mechanism is radiative recombination, and all generated carriers that do not recombine are collected. From these assumption the maximum efficiency for a single bandgap solar cell was found to by 33.7% in the case when the band gap is equal to 1.34 eV. This is the case when the illumination is equal to a spectrum of 1 sun passing through the atmosphere. By using concentrator systems, lenses and mirrors, it is possible to increase this

efficiency to 40.7% [9]. The illumination in this case is the equivalent of 46050 suns. These limits are referred to as the Shockley-Queisser limit.

As these calculations does not include losses from recombinations, other than from radiative recombination, the low efficiency cannot be attributed to recombination processes. The only other loss mechanisms that could cause this large a loss in efficiency is the below band gap- and high energy photons. The types of solar cell that take advantage of these photons are the third generation solar cells. These include multi-junction cells, "hot carrier" cells, and intermediate-band solar cells. The last one will be described in detail in section 2.1.3 as this is the focus of the thesis. Currently the most successful of these are the multi-junction cells. The current world record holder (as of march 2014) is a four-junction cell with an efficiency of 44.7% at illuminations equal to 297 suns [6].

The multi-junction cells work by including many band gaps within one cell. This can either be achieved by having each junction separated physically and then connecting each junction in series with multiple contacts. The other method is to integrate the cells into one structure and using tunnel junctions to connect them in series. In both of these cases the structure is arranged in such a way that the highest band gap material is on top and the lowest band gap at the bottom. This ensures that there is minimal thermalisation losses, as the highest energy photons are used to excite electrons in the highest band gap material and so on. The first method requires very fine circuitry and is therefore very expensive. The second method needs to use materials which have the required band gaps and a low lattice mismatch. If the lattice parameters are very different, the material will be strained and dislocations in the material will occur. These dislocations will lower the performance of the cells. The challenge with this method is therefore related to finding the proper materials to improve the crystal quality. The latter method is very promising, but as mentioned there are some problems that needs to be solved before these types of cell will be cost effective compared to silicon solar cells, which currently holds the majority of the market share.

2.1.3 Intermediate band solar cells

An intermediate band solar cell (IBSC) is similar to a multi-junction cell in one sense. That is the use of multiple band gaps to increase

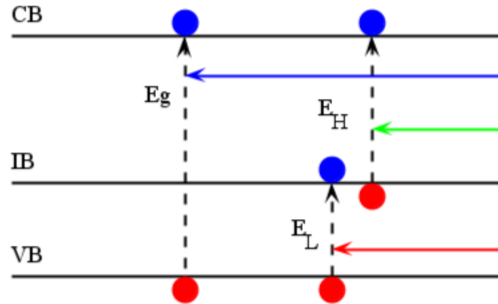


Figure 2.7: An intermediate energy band is introduced into the bandgap of a material. This creates three different bandgaps, making it possible to utilize the photons with energies lower than the original bandgap. The notations for the bandgaps are E_L for the lowest bandgap, E_H for the middle bandgap, and E_G for the original band gap of the material. The three different absorption possibilities are shown here, with the incoming photons marked with blue, green, and red indicating the energy of the photon in decending order.

efficiency. The difference is that instead of having different band gap materials stacked on top of each other, an intermediate energy band (IB) is introduced within the band gap of a single material. The IB effectively creates three different band gaps within a single material. A schematic of this is shown in figure 2.7.

This allows for three different absorption possibilities within the material, VB $\rightarrow$ IB, IB $\rightarrow$ CB, and VB $\rightarrow$ CB. As a consequence of this, below band gap photons can now be absorbed and generate charge carriers. For a material with a single IB with the optimum band gaps, $E_L = 0.71$ eV, $E_H = 1.24$ eV and $E_G = 1.97$ eV, the theoretical maximum efficiency is 63.2% [9]. To achieve the optimum operating conditions, the IB band needs to be half filled [10]. This ensures that there are always electrons in the IB available for excitations into the CB and that there are available states for the electrons that are excited into the IB. A half filled IB is not necessarily needed if the cells are used in conjunction with a concentrator system [11].

The IB is essentially a collection of energy levels within the band gap, and as explained in section 2.1.1 such levels usually increases the amount of non-radiative recombination. The reason this doesn't happen in an IBSC is that the electrons are in a delocalised state in the IB. This is also the case for both the VB and CB. When an electron is in a delocalised state it is not associated with one specific

state, but extends across all states. During a recombination event the electron changes from one delocalized state to another. This doesn't perturb the lattice in any significant way, and no phonons will be created during this recombination. For a more thorough explanation of this see [12, p. 2]. This delocalised nature of the IB is why it is referred to as a band instead of a collection of energy levels.

The energy lost during the recombination is therefore lost through either radiative recombination or auger recombination. The latter is not as detrimental to the efficiency as in the case of conventional solar cells [13]. This is because the energy produced during the recombination can be given to an electron in the IB, exciting it to the CB.

When manufacturing a IBSC the IB material needs to be separated from the contacts, as this would extract a current from the IB. Isolating the IB from the contacts also makes it possible for the quasi-Fermi levels to split when the cell is illuminated. This splitting gives a high output voltage, without changing the current, and it is because of this that IBSCs have such a high maximum efficiency. The band diagram for both equilibrium and the non-equilibrium case is shown in figure 2.8a and 2.8b, respectively. The splitting occurs because the carrier relaxation within the bands is a much faster process compared to the recombination between the bands. This means that the carrier concentration in each band can be described by its own quasi-Fermi level, E_{FV}, E_{FC}, and E_{FI}.

To separate the IB from the contacts, the IB material is sandwiched between a single band gap material, where one side is p-doped and the other is n-doped. These regions are called emitters. Ideally the emitters should be made from a material with higher band gap than the IB material, as this would give a better contact selectivity for electrons and holes. This means that only holes would be extracted from the p-side and only electrons from the n-side. It will also reduce the amount of recombinations at the contacts. These emitters can be seen in the band diagrams in figure 2.8a and 2.8b.

2.1.4 Quantum dot intermediate band solar cells

There are different ways of creating an IB material, some examples are impurity doping or quantum dots (QD). This section will detail the latter method, as this is the basis of the materials studied in this thesis.

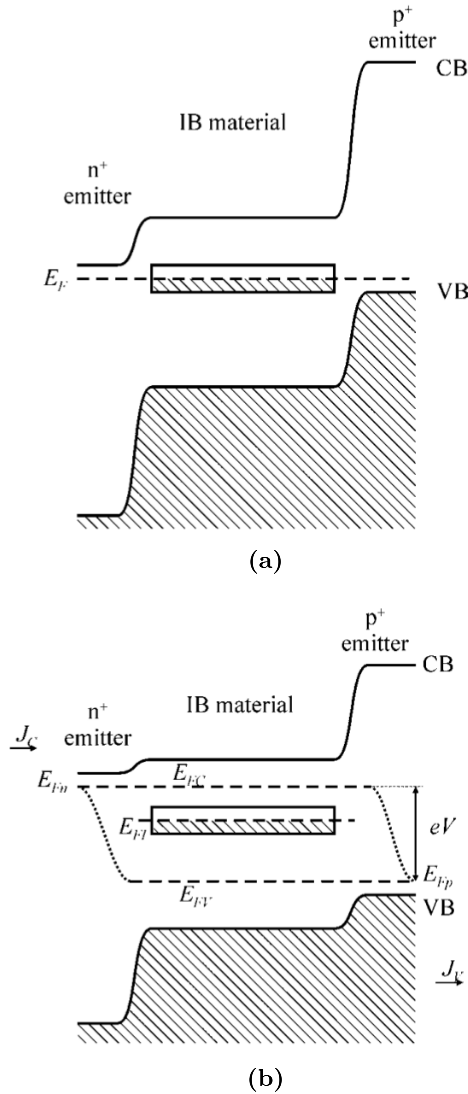


Figure 2.8: Energy band diagram of a IBSC under different conditions: **(a)** Equilibrium case **(b)** Under illumination. In this case there is a splitting of the Fermi levels into three different quasi-Fermi levels. The difference between E_{FV} and E_{FC} determines the output voltage of the cell. Image from reference [14].

QDs are pieces of crystalline semiconductors, small
enough to exhibit quantum confinement effects, embedded in a matrix

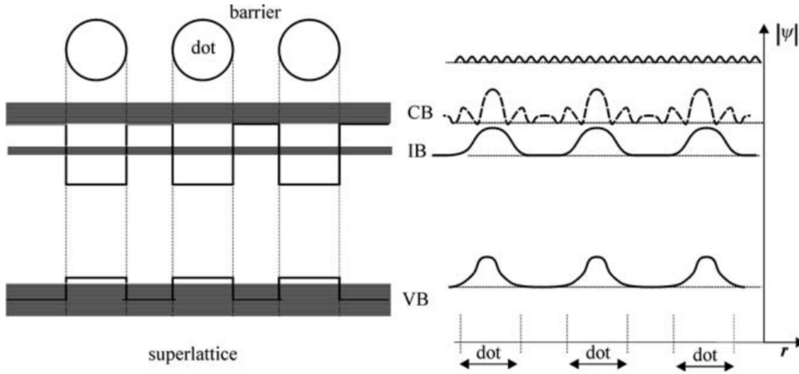


Figure 2.9: The left hand side shows how the bands form in the QD based IB material. Due to the regular superlattice of QD and barrier material, the wave functions will start to overlap creating a new VB and an IB. The wave functions are shown schematically on the right hand side. . Image from reference [14].

material, usually with a higher band gap. The higher band gap of the matrix material will be a barrier for the electron in the QD and the matrix material will therefore be referred to as the barrier material. Due to the quantum confinement effect the QDs will have discrete energy states, in contrast to bulk materials where the energy states are continuous. This makes it possible to tailor the band gap, in order to optimize the efficiency, by changing the size of the QD.

The wave function of an electron in the CB of the QD will extend slightly into the barrier material. By placing the QDs in a tightly spaced pattern the wave functions will overlap. This overlap will create a IB in the material and the electrons here will be in a delocalised state. A schematic of this is shown in figure 2.9. This will increase the probability of absorption [12] and reduce non-radiative recombination, as explained in the previous section.

The width the IB will depend on how far the QDs are placed from each other [15]. Figure 2.10 shows an example of this for calculations done on spherical QDs arranged in a cubic pattern.

When choosing the QD material, it is also important to keep in mind what the VB offset of the QD material is. Ideally it should align with the VB of the matrix material, but this is often not achievable. The VB offset will effectively reduce the band gap in the IB material,

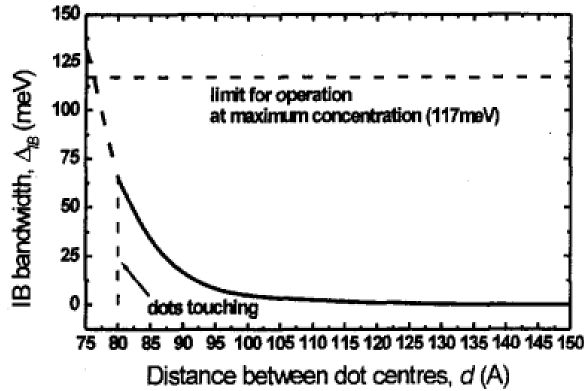


Figure 2.10: The width of the IB as a function of the distance between each QD. The calculation is based on the tight binding approximation for spherical QDs with a radius of 40 Å distributed in a cubic array. [15]

making the band gap less than optimal. This can be seen in figure 2.9. This situation can be remedied by starting with a barrier material with a higher than optimal band gap.

The first QD-IBSCs were demonstrated by Luque et al. in 2004 [13]. These cells were grown by MBE using the Stranski-Krastanov growth mode. Both MBE and the Stranski-Krastanov growth mode will be explained in more detail in section 2.6. These cells where made by growing QDs in layers with a 10 nm spacer in between each layer. A total of 10 layers were grown, making the total thickness of the IB material 100 nm. Here the QDs were made of InAs and the spacer layer of GaAs. The InAs/GaAs material is excellent for testing the theory of IBSCs, due to it being a well understood system, but in terms of efficiency it is not optimal.

A reference cell of GaAs were also made. Subsequent analysis revealed that the quasi-Fermi levels of the CB and IB were separated. This is a important result as it shows that the underlying theory of IBSCs is well founded. These cells exhibited a response in the quantum efficiency for sub-band gap photons that were absent in the control cell. The current from this region was quite low and this was attributed to the small number of QD layers [16], since each QD is very small and doesn't absorb much light. The structure of the cell is shown in figure 2.11a and the quantum efficiency, for a model system, the GaAs control cell and the IBCS cell, is shown in figure 2.11b.

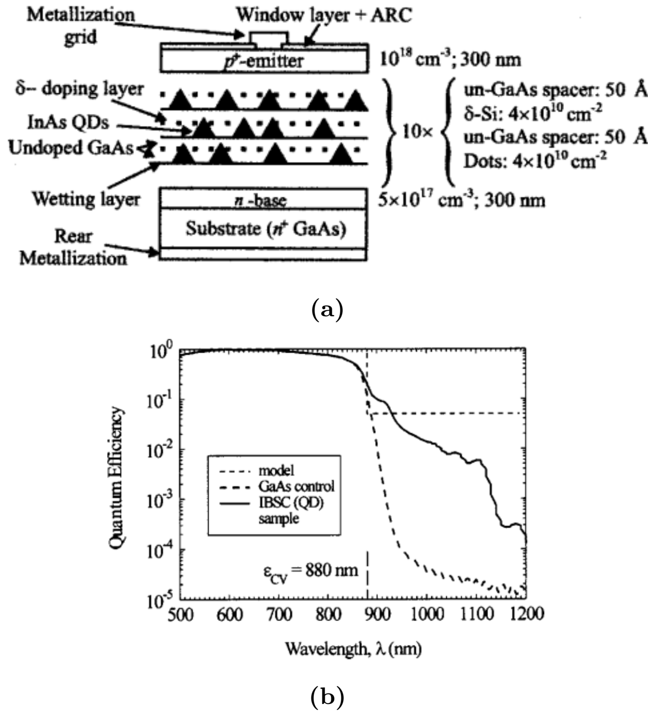


Figure 2.11: (a) Schematic of a QD-IBSC with 10 QD layers separated by a 10nm spacer. The IB material has been doped with Si to provide a half filled IB. On the surface a ARC has been applied to reduce losses from reflection. **(b)** External quantum efficiency of the cell in (a). The IBSC cell shows a sub-band gap response, indicating that the cell has an IB. The response is still quite low due to a small number of QD layers. Both images from [13]

To achieve a higher photocurrent from the photons in the sub-band gap region, the number of QD layers would need to be higher [17], but this introduces a new problem. Because the lattice mismatch between InAs and GaAs is quite high (7%), the system will be strained. As the number of layers increases, so does the strain. The strain damages the crystal structure of the cell by introducing dislocations and stacking faults. These increases the amount of non-radiative recombinations and so the efficiency will decrease.

Various methods have been implemented to try to compensate for this strain. Hubbard et al. showed that the inclusion of GaP into the structure compensated for strain [18]. Zhang et al. showed the

possibility of strain compensation by the use of GaNAs. By embedding InAs QDs in a layer of GaNAs, the optical properties were improved as a result of strain reduction in the system [19], [20]. Oshima et al. made cells with 20 layers of InAs QDs by using GaNAs to compensate for strain [21]. These cells contained few defects and had a four times higher short-circuit current compared to strained cells with identical structure. Other papers published on the use of GaNAs as a strain compensator can be found in references [22–27].

However even these cells show a decrease in the efficiency, due to a decreased open circuit voltage, compared to a reference GaAs cell. This decrease in voltage can be attributed to the VB offset and a the existence of a wetting layer, introduced during the growth. The wetting layer is a 2-3 monolayer (ML) thick layer, which introduces energy states, below the CBE of the material, that form a continuum with the CB of the material. This lowers the effective CBE and combined with the VB offset, which increases the VBE, the new band gap will be lower than the GaAs reference cell. As the maximum output voltage is proportional with the band gap, the IBSCs cell will have a lower output voltage compared to the reference cell. It is expected that this is possible to remedy by using concentrated light [28].

2.2 Materials

The solar cells studied in this thesis were mainly composed of four different compound semiconductors; Gallium Arsenide (GaAs), Indium Arsenide (InAs), Gallium Nitrogen Arsenide (GaNAs), and Aluminium Gallium Arsenide (AlGaAs). A compound semiconductor is a semiconductor composed of two or more elements from different groups in the periodic table. All of the materials used are III-V semiconductors, where the group III elements are Al, Ga, and In and the group V elements are N and As. The number of the group is determined by how many electrons the elements have in their outer shells. The group III elements will therefore easily bond with the group V elements, as this fill their outer shells with electrons, making this arrangement energetically favourable.

In AlGaAs, a fraction of the Ga atoms has been replaced by Al, and in GaNAs, a fraction of the As atoms has been replaced by N. This is written as $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ and $\mathrm{GaN}_x\mathrm{As}_1 - x$, were x is a number between 0 and 1 and indicates the percentage of the specific atoms in

Table 2.1: The band gaps and lattice parameters at 300 ° K for the relevant materials. For AlGaAs and GaNAs these parameters will change depending on the fraction of elements in the compound, and therefore only the limiting cases are listed, i.e. $x = 0$ (GaAs) and $x = 1$ (GaN, AlAs)

Material	Band gap [eV]	Lattice parameter [Å]
GaAs	1.42 [29]	5.6533 [30]
InAs	0.356 [29]	6.0583 [31]
GaN	3.51 [32]	(3.1890, 5.1850) [32]
AlAs	2.16 [29]	5.6611 [30]

the material. For example in $Al_xGa_{1-x}As$ x amount of Ga atoms have been replace by Al, but the As content is the same.

Properties such as band gap and lattice parameter will depend on how large the fraction is. For example $Al_xGa_{1-x}As$ is a direct band gap material for $x < 0.4$, and for larger x values it changes to a indirect band gap material. In this thesis the Al content in AlGaAs was 25%, while the N content in GaNAs was more uncertain.

The lattice parameters and band gaps for the various materials studied in this thesis are summarized in table 2.1.

The crystal structure of GaAs, InAs and $Al_xGa_{1-x}As$, is called Zincblende. This is the case for all x values in the $Al_xGa_{1-x}As$.
In GaN_xAs_{1-x}, the crystal structure will be Zincblende up to a certain concentration of N. At this point the structure will change to the Wurtzite structure. In this thesis the N content is not high enough to change the structure to Wurtzite, so the only phase existing in these solar cells were Zincblende.

The Zincblende structure consists of two interlocking face centered cubic (FCC) lattices, consisting of different atoms, where one FCC lattice is shifted by $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ compared to all the atoms in the other. The unit cell of a FCC structure will be a cube with atoms at all faces and corners, and can be described using only one lattice parameter, usually denoted as a. This is the length of the sides in the cube.

Wurtzite is a hexagonal crystal structure, composed of two different atoms. The unit cell of the Wurtzite structure is a cube that is elongated along one axis. As a consequence of this, two lattice parameters are needed to describe this structure. These are usually denoted as a and c, where a is length of the shortest side and c the length of the elongated side.

These structures can be seen in figure 2.12a and 2.12b respectively.

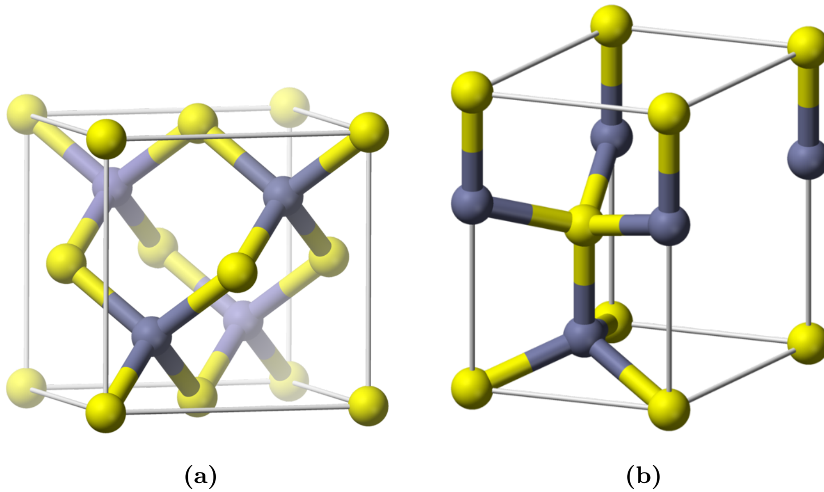


Figure 2.12: **(a)** Zincblende unit cell [33]. The grey and yellow spheres represent different types of atoms. The grey atoms could for example be Ga and the yellow As. **(b)** Wurtzite unit cell. Again the differently coloured spheres represents different atoms. [34]

2.3 Transmission electron microscopy

Transmission electron microscopy (TEM) is a technique that is used to study materials through a range of length scales, from a nano scale, <100nm, to micrometer lengths. The TEM is a versatile tool, with a multitude of different characterization techniques. It can be used for strain quantification, studying material composition, characterizing the crystal and energy band structure, just to name a few.

The newest TEM's are capable of resolving points with sub-ångstrom (Å) ($<0.1nm$) distance between them and this offers valuable and very precise information on crystal structure, atom placement, strain, and chemical composition. The TEM is used to study various types of materials, from semiconductors and metals to biological materials, such as cells and viruses.

This chapter will go through the working principle of the TEM, the various imaging modes, and the different mechanisms for contrast in these imaging modes. The optics and the lens system will also be briefly explained. Unless otherwise specified, the material presented here are from Williams and Carter's series of books on *Transmission Electron Microscopy* [7].

2.3.1 Resolution limit

The resolution in an image is defined as the minimum distance between two point that can be resolved. For a visible light microscope (VLM) this distance can be calculated by the following equation

$$\delta = \frac{0.61\lambda}{\mu sin(\beta)} \tag{2.1}$$

called the Rayleigh criteria. Here δ is the smallest distance that can be resolved, λ is the wavelength of the light that is used, μ is the refractive index of the medium between the lens and the material, and β is the semi-collection angle of the magnifying lens. The expression in the case of TEM is very similar, and can be expressed as

$$\delta = 1.22\frac{\lambda}{\beta} \tag{2.2}$$

The μ factor is disregarded here because the electron beam is travelling through vacuum. The other factors are the same as for equation 2.1.

The minimum distance the human eye can resolve can be found by using equation 2.1. This distance will depend on how good the eye is, but for an average eye the distance is $0.1 - 0.2$ mm. Anything that can reveal details below this distance is referred to as a microscope.

In VLM the light used is in the visible region, and this means that the wavelength is between $400 - 700$ nm. Assuming that the factors $\mu sin(\beta)$ to be unity, the minimum distance that can be resolved will be in the region $24.4 - 42.7\mu$m. By exploiting the wave characteristics of the electrons the resolution can be improved drastically.

In 1914 Louis de Broglie proposed that all particles exhibit a wave-like characteristic and would therefore have a wavelength. After electron-diffraction experiments were carried out, which demonstrated the wave-like nature of the electron, it wasn't long until the idea of an electron microscope was born. The de-Broglie expression for the wavelength is given as

$$\lambda = \frac{h}{p} \tag{2.3}$$

where h is Planck's constant and p is the momentum of the particle. To properly describe the wavelength of an electron in the TEM, the relativistic expression for the momentum needs to be used in equation 2.3. The reason for this is that the high voltage in the TEM accelerates

electrons to such high speeds that relativistic effects become important. The following equation gives the wavelength of an electron, accelerated by a voltage V, travelling at relativistic speeds

$$\lambda = \frac{h}{\sqrt{2meV(1 + \frac{eV}{2mc^2})}} \tag{2.4}$$

Here m is the mass of the electron, e is the elementary charge, and c is the speed of light. For a 200 kV accelerating potential, the wavelength is 0.00251 nm. By the TEM equivalent of the Rayleigh criteria, assuming $\beta = 1$ to make the calculation easier, the resolution will be 0.0305 Å. This value is on the same scale as the size of atoms, making electrons an ideal source for imaging atomic structures. The resolution in a real microscope will be less than this, due to a different β value and imperfect lenses. The last point here will be explained in section 2.3.2.

2.3.2 The Instrument

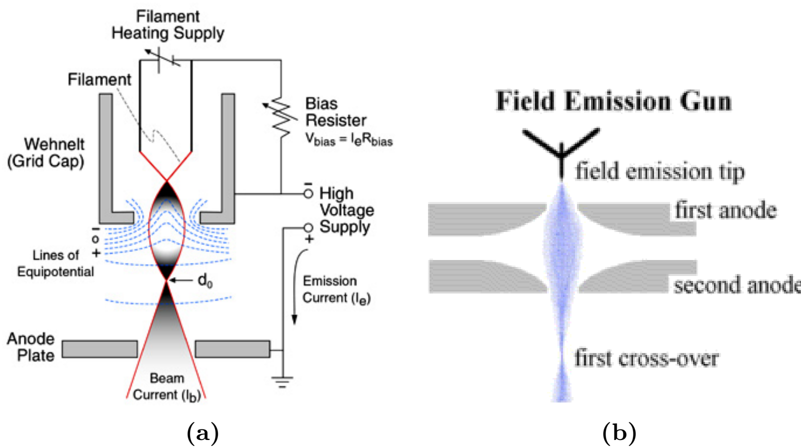
A transmission electron microscope consists of an electron source, an optics system for transmitting the electron beam through the instrument, and various detectors, ranging from CCD cameras for regular imaging to EDX detectors for chemical composition. The following paragraphs will quickly explain the working principles behind some of the components in a TEM. Lastly it will be explained how these come together to form a working electron microscope.

2.3.2.1 Electron sources

The electron sources can be divided into two categories:
Thermionic sources and field emission sources. When these are combined with a accelerating voltage they are referred to as electron guns.

In a thermionic source a filament of Tungsten (W) or a crystal of lanthanum hexaboride (LaB_6) is heated. This raises the energy of the electrons beyond the work function of the material, making it possible for the electrons to escape the material.

In a field emission source (FE) the electrons are drawn out of a very fine tip by a high voltage. This high voltage changes the energy potential of the material at the tip, increasing the possibility that the electron can tunnel out of the material. The field emission



(a) (b)

Figure 2.13: (a) A thermionic electron gun. Here a thin filament or crystal is heated, generating a stream of electrons. These are then accelerated by an applied voltage. Image from [35]. **(b)** Field emission electron gun. In this case a high voltage is used to draw electrons out from a fine tip. These are then accelerated by an applied voltage, similar to the thermionic case. Image from [36].

guns (FEG's) can be divided into two different types: Schottky FEG and cold FEG's (CFEG). The difference between the two is that the Schottky FEG is continuously heated, while the CFEG is kept at room temperature. Contamination on the CFEG will decrease the performance, and therefore will need to be kept at a high vacuum. There will still be build up of contaminations and this is removed by "flashing" the tip. The "flashing" procedure works by either reversing the potential applied to the tip, ejecting the contaminant atoms from the tip, or heating it to $\backsim$ 5000 K quickly to evaporate the contaminant atoms. In Schottky FEGs contaminations are not a problem, as the tip is kept at a high temperature at all times. This means that the needed vacuum can be lower here, compared to a CFEG.

The FEGs have a higher current density, lower energy spread, and a higher operating lifetime compared to the LaB_6 crystal. Therefore the FEG's provide a much better resolution, in both real space and energy space, and image quality compared to thermionic sources. Of the two FEGs the CFEG is the better of the two, but requires a much higher vacuum. The only case where the LaB_6 thermionic source is better than the FEG's is during imaging at a relatively low magnification (50 - 100,000x)). This is because the size of the FEG source is too small,

and so when imaging at low magnification, the beam must spread beyond the size of the tip, causing a subsequent decrease in intensity.

2.3.2.2 Electron optics

The lenses in a TEM are analogous to convex glass lenses in a VLM. In both cases the lens needs do to two things: Gather all the rays coming from a point in an object into a point in the image plane, and to focus all incoming parallel rays to a point in the back focal plane (BFP) (figure 2.14). In a VLM this is done by having a translucent material with a different refractive index. This wouldn't work for a beam of electrons, so instead electric and magnetic fields are used. By varying the current used to set up these fields, the strength of the lens can be changed. This makes it possible to control the magnification, illumination, and other parameters, just by changing the current.

As these types of lenses are far from perfect, a more correct analogy would be to compare TEM lenses to looking through the bottom of a dirty glass bottle. The imperfection in the lenses introduces various aberrations in the beam, which reduces the possible resolution in the microscope.

The two most important limitations are spherical and chromatic

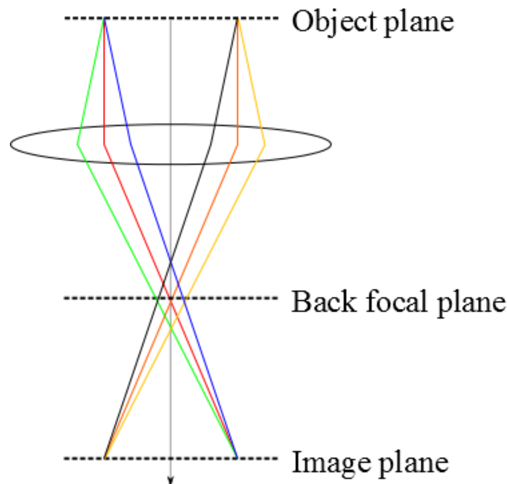


Figure 2.14: The ray diagram for a convex lens. The parallel rays intersect in the back focal plane, while all the rays originating from one point are focused to one point in the image plane. Figure adapted from figure 6.3 in [7].

aberration, but recently microscopes have been fitted with complicated systems which compensate for the spherical aberrations, which is the most detrimental to the spatital resolution of the two. These types of microscopes are very expensive, and are therefore still very rare. In other microscopes these effects can be reduced by inserting a diaphragm into the beam path.

The diaphragm is a disk with a hole in it, called an aperture, which blocks out parts of the beam and limits the collection angle of the lens. Electrons passing through the center of the lens will go through unaffected and thus the aberrations will be worse the further out from the center of the lens, or optical axis, the electrons pass. Therefore by selecting electrons that pass close the the optical axis, the effect of these aberrations are reduced. This is done by placing the diaphragm around the optical axis, excluding electrons that pass outside a diameter determined by the aperture. This decreases the brightness of the beam, but improves the resolution, and so the aperture used will be a compromise between needed resolution and brightness.

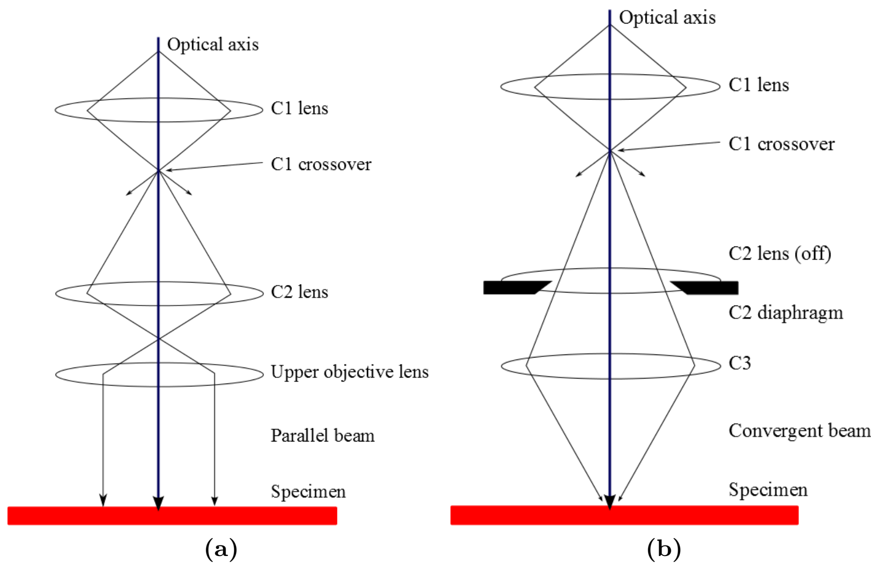


Figure 2.15: Ray diagrams for the two operational modes of the illumination system. By adjusting the strength of the condenser lenses, the beam path can be changed. The ray diagram for parallel illumination is shown in **(a)** and convergent illumination is shown in **(b)**. Figures adapted from figure 9.1 and 9.3 in reference [7], respectively.

For convenience the components of a TEM can be summarized in three different categories: the illumination system, the objective lens, and the imaging system.

The illumination system is responsible for producing a beam of electrons and bringing these to the sample. It consist of an electron gun, which produces the electrons, and the condenser lenses, which transfers the electrons to the sample. There are also other lenses which control the spread of the beam and the spot size. The illumination system can be operated in primarily two modes: parallel beam (figure 2.15a) and convergent beam mode (figure 2.15b). The first is mainly used for TEM imaging, bright field (BF) and dark field (DF) imaging, and diffraction patterns (DP), whereas the latter is used for scanning transmission electron microscopy (STEM) and convergent beam electron diffraction (CBED).

The next stage of the microscope is where the objective lens and specimen is situated. This is the part of the microscope where all the interactions between the beam and the sample occurs. Due to the effect of lens aberrations, it is important that the specimen is placed near the centre of the objective lens. As a consequence of this, the movement and rotation of the specimen will be limited.

The final stage of the microscope is the imaging system. Here various lenses are used in order to magnify the diffraction pattern (DP) or image down to the viewing screen, camera or photographic film. The two first lenses magnify the image (or DP) and are called the intermediate and diffraction lenses. The final lens projects the image onto the viewing screen or camera and is therefore called the projection lens.

These systems comprise what is called the column of the microscope and is kept under vacuum by various pumping systems. This ensures that the electrons will transmit through the column without scattering and also keeps contaminations from accumulating on the sample surface, which is detrimental to the quality of the image. Most TEMs are built such that the column is vertical, with the electron gun at the top, and so the electrons are often referred to as travelling down the column. Figure 2.16 shows a schematic of a TEM and illustrates the positions of the electron source, apertures, and the imaging system.

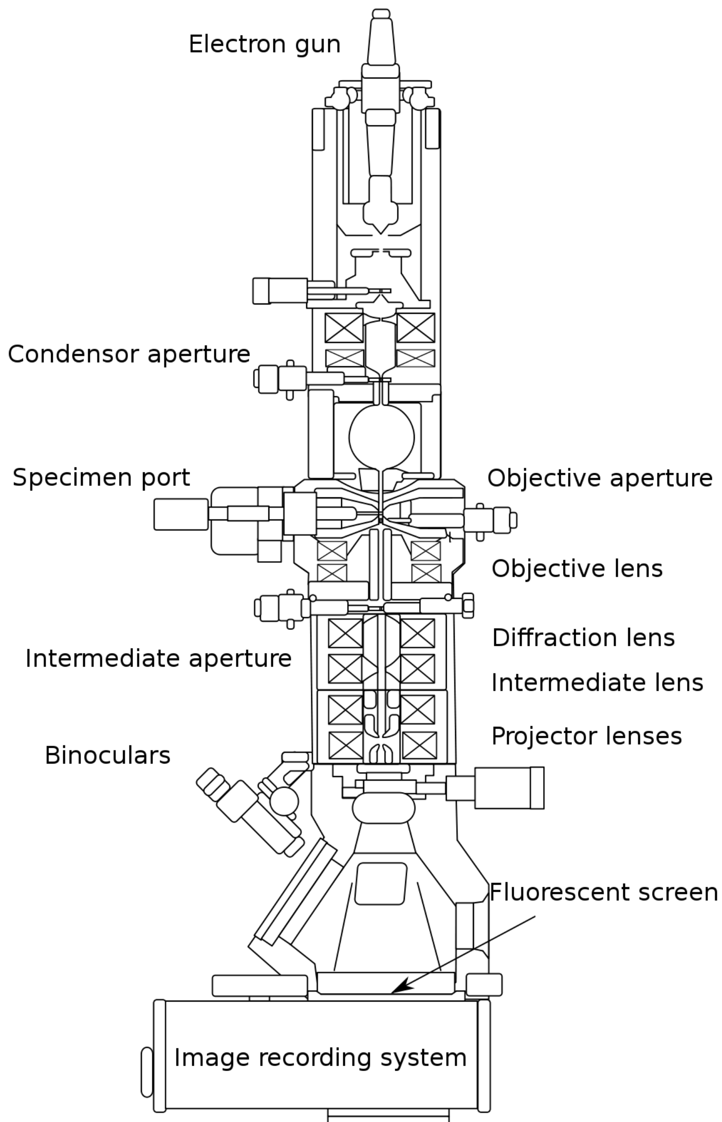


Figure 2.16: A schematic of a TEM, where all the lenses and apertures are indicated. The condensor lens is not labeled on the figure, but it is situated right above the condensor aperture [37].

2.3.3 Electron-matter interactions

An incoming electron interacts with a material through scattering events. These scattering processes can either be elastic, where no

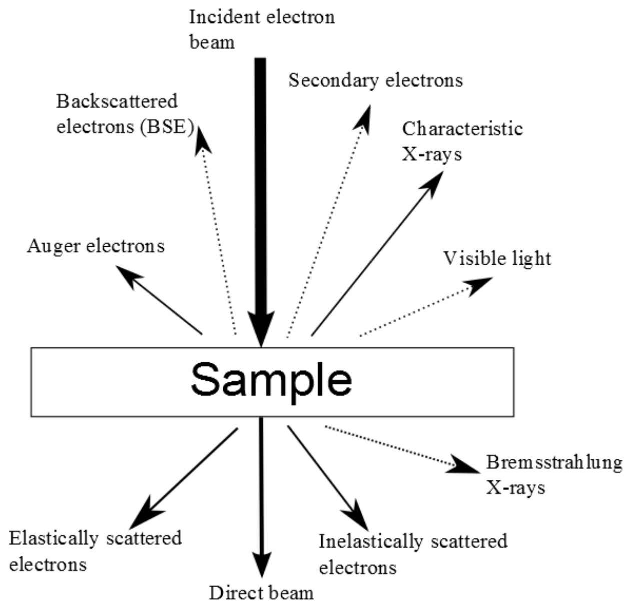


Figure 2.17: The various signals generated when the electron beam interacts with a material. The solid lines are the signals that have been utilized in this thesis. The direction of the lines gives a rough estimate of where the signals are the strongest. Figure adapted from figure 1.3 in reference [7].

energy is lost during the interaction, or inelastic, where energy is lost. The inelastic interactions produce a range of signals which can be used in various techniques related to composition or topography. Signals from the elastically scattered electrons are mostly used for crystal characterization, e.g. crystallographic orientation determination. The various signals produced when electrons interact with matter is shown in figure 2.17.

Electrons can be scattered either coherently or incoherently. Coherently scattered electrons exhibit a wave-like characteristic compared to incoherently scattered electrons which should be thought of more as particles. Scattering can also be either forward or back scattering. Elastic scattering is strongly peaked in the forward direction and for thin samples is usually coherent. At higher angles ($\backsim 10°$) elastic scattering becomes incoherent. Inelastically scattered electrons are scattered to very low angles ($< 1°$) and is mostly incoherent. Only the forward scattered electrons are utilized in a standard TEM, with the exception of the X-rays generated during the inelastic scattering

processes for composition determination.

2.3.4 Diffraction

Diffraction in a TEM can be compared to the diffraction of light through a diffracting grate or slit. In this case the periodic atomic structure of the crystals acts as a diffracting grate. Diffraction will occur as long as the wavelength of the electrons are smaller than the lattice spacings in the crystal. Because of the high voltage used to accelerate the electrons, this condition is fulfilled for all cases in a modern TEM. The electrons that contribute to the diffraction pattern (DP) are the ones that are coherent and scattered elastically to low angles. The diffraction pattern arises because of the wave-like characteristics of these electrons, resulting in constructive and destructive interference. This results in a pattern of spots, called diffraction spots or reflections (figure 2.18), each corresponding to a diffracting plane in real space. These planes are indexed by the Miller indices and the diffraction spots are indexed by the same indices.

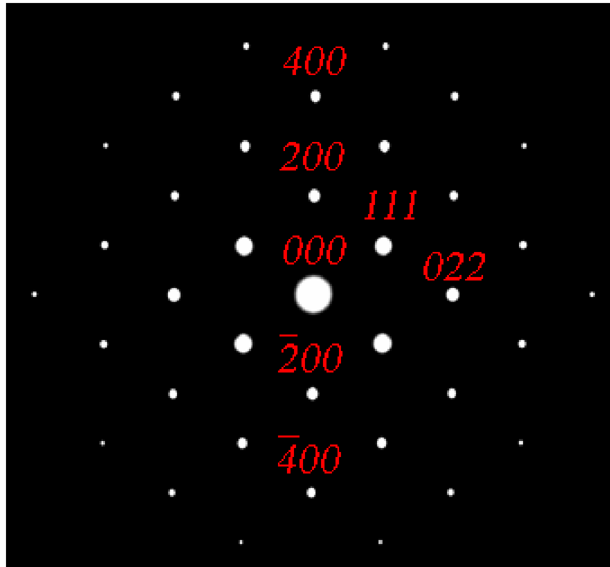


Figure 2.18: Diffraction pattern from the [110] zone axis of an FCC crystal, in this case GaAs. A few of the indices of the reflections are also shown.

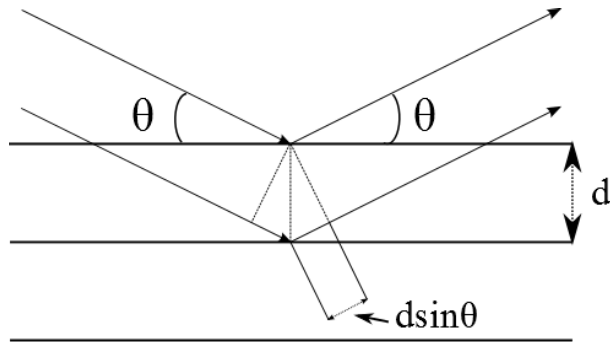


Figure 2.19: A schematic of the derivation of Bragg's law. The incoming beam is reflected of the parallel lattice planes, separated by a distance d, resulting in a path difference of 2dsinθ. The incoming angle is defined as θ. Adapted from figure 2 in reference [38].

2.3.4.1 Bragg's law

W. L. Bragg assumed that it was constructive interference from beams reflected from parallel atomic planes that caused the diffraction pattern. Although this explanation is physically wrong, Bragg's law presents a simple and useful method for calculating the distance between atomic planes. The assumptions made were that each atomic plane would reflect only a small part of the beam and that the reflection would be specular. Under these assumptions, in order to have constructive interference, the path difference between two reflected beams must be an integer value of the wavelength. This gives the following relation, based on geometrical considerations (figure 2.19).

$$2dsin\theta = n\lambda \tag{2.5}$$

here d is the distance between parallel lattice planes, θ is the scattering angle, also called the Bragg angle, n is an integer, and λ is the wavelength of the electron.

By this relation, in a given crystal, only certain values of θ will give diffraction spots. This is entirely dependent on the crystal structure, but the intensity in each spot will depend on the composition of the unit cell, which will be explained in section 2.3.4.3.

2.3.4.2 Laue equations and the Ewalds sphere

The reciprocal lattice vectors are defined similarly as the lattice vectors in real space. In real space, the lattice vectors are defined in such a way that it is possible to create a large crystal, just by translating the unit cell along these vectors. The reciprocal lattice vectors do the same thing, only with the diffraction pattern. The possible reflections in a DP is determined by the set of reciprocal lattice vectors **G**, when the condition

$$\mathbf{G} = \mathbf{\Delta k} \tag{2.6}$$

is fulfilled [38]. Here $\Delta\mathbf{k}$ is called the scattering vector and is defined as $\Delta\mathbf{k} = \mathbf{k_{out}} - \mathbf{k_{in}}$, where $\mathbf{k_{out}}$ and $\mathbf{k_{in}}$ is the wavevector of the reflected and incoming beam respectively.

This is derived by looking at an incoming electron wave hitting a crystalline specimen. The incoming electron wave is scattered from two points (volume elements), separated by a distance **r**. This introduces a phase difference between the two diffracted waves, $(\mathbf{k_{in}} - \mathbf{k_{out}}) \cdot \mathbf{r}$. Now **r** will correspond to a vector in real space in the crystal, defined as

$$\mathbf{r} = u_1\mathbf{a_1} + u_2\mathbf{a_2} + u_3\mathbf{a_3} \tag{2.7}$$

where u_1, u_2, and u_3 are integers and $\mathbf{a_1}, \mathbf{a_2}$, and $\mathbf{a_3}$ are the real space lattice vectors.

As the phase difference between the two scattered waves must be zero (or an integer value of 2π) in order for there to be constructive interference, the condition

$$(\mathbf{k_{in}} - \mathbf{k_{out}}) \cdot (u_1\mathbf{a_1} + u_2\mathbf{a_2} + u_3\mathbf{a_3}) = 2n\pi \tag{2.8}$$

must be fulfilled.

Lastly a set of reciprocal lattice vectors **G** can be defined as

$$\mathbf{G} = v_1\mathbf{b_1} + v_2\mathbf{b_2} + v_3\mathbf{b_3} \tag{2.9}$$

where v_1, v_2, and v_3 are integers and $\mathbf{b_1}, \mathbf{b_2}$, and $\mathbf{b_3}$ are the reciprocal lattice vectors. $\mathbf{b_i}$ are defined in such a way that $\mathbf{b_i} \cdot \mathbf{a_j} = 2\pi\delta_{ij}$ must be fulfilled, where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ if $i \neq j$. In a diffraction pattern, the notation used is **G** for the reciprocal lattice vector, and G for the diffracting point.

It can be shown that $\mathbf{G} \perp (hkl)$ and $|\mathbf{G}| = \frac{1}{d_{hkl}}$. Here h, k, and l are the Miller indices for a plane in real space. (hkl) denotes a plane described by the three Miller indices.

From equation 2.9 and the condition $\mathbf{b_i} \cdot \mathbf{a_j} = 2\pi\delta_{ij}$ it is clear that in order for equation 2.8 to be fulfilled $\mathbf{k_{in}} - \mathbf{k_{out}}$ must equal a reciprocal lattice vector, i.e. $\mathbf{k_{in}} - \mathbf{k_{out}} = \Delta\mathbf{k} = \mathbf{G}$ (equation 2.6).

By taking the scalar product of both sides of equation 2.6 with $\mathbf{a_1}, \mathbf{a_2}$, and $\mathbf{a_3}$ successively, the Laue conditions are found [38].

$$\mathbf{a_1} \cdot \Delta\mathbf{k} = 2\pi v_1,\ \mathbf{a_2} \cdot \Delta\mathbf{k} = 2\pi v_2,\ \mathbf{a_3} \cdot \Delta\mathbf{k} = 2\pi v_3 \tag{2.10}$$

According to these equations, in order for a diffracted beam to form, $\Delta\mathbf{k}$ must lie on the intersection of three cones, where each cone is described by the equation 2.10. It can be shown that the Laue equations are mathematically equivalent to Bragg's law.

An interesting way of showing the Laue equations graphically is by the Ewald construction, or the Ewald sphere (see figure 2.20). Here the incoming wavevector $\mathbf{k}$ is drawn originating at the center of a sphere and ending at a reciprocal lattice point. The sphere is drawn

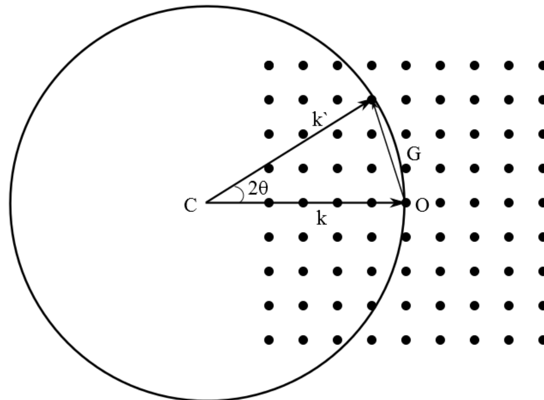


Figure 2.20: The Ewald construction for a cubic reciprocal lattice. The incoming electron wave has wavevector $\mathbf{k}$, and the scattered wave has wavevector $\mathbf{k}$‘. The difference between the two is labeled as $\mathbf{G}$, representing a reciprocal lattice vector. The O and C is shortform for the origin and center respectively. Adapted from figure 8 in reference [38].

around the start of this wavevector, with a radius equal to its length, i.e. $r = |\mathbf{k}| = \frac{2\pi}{\lambda}$. The center (C) of the sphere is determined by the length of $\mathbf{k}$ and the diffraction spot where $\mathbf{k}$ terminates is called the origin (O). The wavevector of any elastically scattered beam will have the same length as the incoming wavevector, as no energy is lost in the process, meaning that this vector will also terminate at the boundary of the sphere. Any point the sphere intersects will fulfil the Laue equations. This is shown in a 2D representation in figure 2.20. The vector between two reciprocal lattice points will be a reciprocal lattice vector, and so when $\Delta\mathbf{k}$ is equal to this, i.e when the sphere intersects a point, equation 2.6 is fulfilled.

Another point to note is that the reciprocal lattice points are not one dimensional. This would have been the case if the crystal was infinite, but obviously this is not the case. Because of the finite nature of the crystal, each point will be more like a rod. This is the reason why there will still be some intensity in the diffracted spot, even though

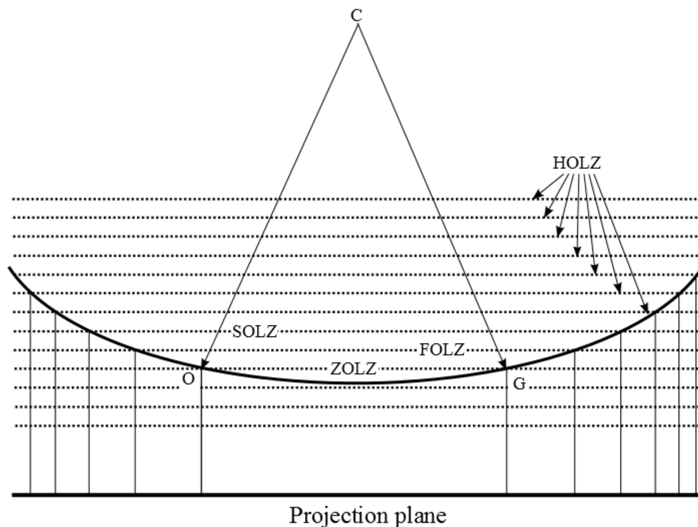


Figure 2.21: A cross section of the Ewald sphere, in the case of short wavelength. The reciprocal lattice plane which contains the origin is labeled ZOLZ, followed by FOLZ and SOLZ. The rest are just called HOLZ. The sphere will be very flat, as the wavelength is small, and it will therefore intersect more than one point in each plane at once. The projection plane is what is viewed on the imaging device (CCD camera, photographic film, viewing screen) in a TEM. Adapted from figure 12.5 in reference [7].

the Ewald sphere doesn't intersect the rod exactly in the middle.

At a zone axis the DP shouldn't contain any spots according to the Laue conditions, but this clearly isn't the case (figure 2.18). The diffraction spots in a DP on a zone axis, all come from planes that are parallel to the zone. This means that these planes will not be far from Bragg, as the Bragg angles are usually measured in milliradians(mrad). Therefore, due to the fact that the diffraction spots really are more like rods, the Ewald sphere will still intersect these spots, causing many planes to appear to be in Bragg.

When the wavelength decreases, the radius of the sphere will increase, as it is proportional to λ^{-1}. This causes the surface of the sphere to flatten and it will therefore intersect more reciprocal lattice points. This is why DPs in a TEM contains so many diffraction spots, as the wavelength here is on the order of *p*m. In contrast to X-ray diffraction, where the wavelength is on the order of Å, making it necessary to use various methods to enhance the number of diffraction spots, e.g. using white radiation or rotating the sample.

As the reciprocal lattice is three dimensional, the Ewald sphere will intersect points lying on reciprocal lattice planes that are parallel to the plane that contains the origin of the sphere. These planes are called Laue zones, the plane of points which contains the origin is called the zero-order Laue zone (ZOLZ), the next is called the first-order Laue zone (FOLZ), and the one after that is called the second-order Laue zone(SOLZ). Every plane after the ZOLS can also be referred to as a higher-order Laue zone (HOLZ). This is shown in figure 2.21. The distance between the diffraction spots will be significant, meaning that when looking at a DP centred around spots from the ZOLZ, the spots from the FOLZ won't necessarily be visible.

2.3.4.3 Atomic form factor and structure factor

The intensity in a diffraction spot will vary according the the elemental composition of the unit cell. This is described by the structure factor, which is dependant on the atomic form factor.

The atomic form factor is a measure of the scattering power of a single atom. It is mainly dependant on three factors, the electron distribution of the atom, the atomic number, and the wavelength. A larger atomic form factor means that the atom will scatter more of the incoming radiation.

The structure factor combines the atomic form factor with the

unit cell of a crystal, in order to describe how a crystal will scatter incoming radiation. It shows which hkl values that is expected to show in a DP and how the intensity in each spot will be. Which hkl values that will give non zero intensity is summarized in the selection rules. The structure factor is given as [38]

$$F_{hkl} = \sum_{i=1} f_i \exp^{2\pi i(hx_i+ky_i+lz_i)} \tag{2.11}$$

Here the sum is over all the atoms in the unit cell, f_i is the atomic form factor for atom i, (hkl) are the Miller indices, and x_i, y_i and z_i are the coordinates for atom i. $\exp^{2\pi i(hx_i+ky_i+lz_i)}$ is a phase factor, and determines how the waves scattered of (hkl) planes will interfere. Using this equation the selection rules for a crystal can be found. Below is an example of this for an FCC crystal.

$$F_{hkl} = \begin{cases} 4f, & \text{h, k, l all odd or all even} \\ 0, & \text{h, k, l mixed.} \end{cases} \tag{2.12}$$

The relation between the intensity in each (hkl) spot and equation 2.11 is given as

$$I_{hkl} \propto |F_{hkl}F^*_{hkl}| \tag{2.13}$$

Using this in conjunction with the selection rules gives the intensity in each spot, and which spots that can be present in the DP.

At a zone axis the intensity in each diffraction spot will not necessarily follow the selection rules, due to double diffraction. At a zone axis many atomic planes will fulfil the Bragg condition at once. This means that a electron that has been reflected once, can easily be reflected several more times, as its angle is equal to the Bragg angle of these other planes. As a consequence of this, an electron might scatter into a diffraction spot that should have been weak, according to the selection rules, by two or more transitions. This is called dynamical scattering and the electrons here scatter two or more times as it is transmitted through the sample. The other case is called kinematical scattering, where the electron only scatters once.

2.3.5 Contrast

The contrast in a TEM image can be divided into three categories: (I) Mass-thickness contrast; (II) Diffraction contrast; (III) Phase con-

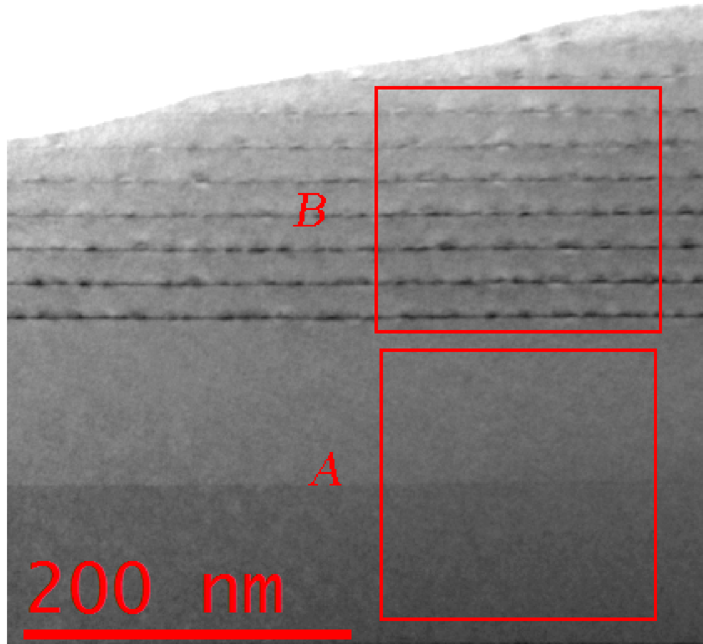


Figure 2.22: A bright field TEM image taken from one of the samples studied in this thesis. (A) shows a contrast due to a difference in mass of the two layers. The bottom layer, which appears darker is GaAs, and the one on top is AlGaAs, where some of the Ga atoms have been substituted with Al, making it a lower Z composite. (B) Shows alternating layers of dark InAs QD's/wetting layer with a GaAs spacer layer. The difference in lattice parameter of the two materials are 7%, resulting in strain surounding the InAs material. The contrast present here is therefore a result of a combination of diffraction contrast and mass contrast, since In is heavier than Ga.

trast. Which one that will dominate will depend on the magnification, imaging mode, and the material being studied.

- Mass-thickness contrast (figure 2.22 A.) is due to a variation in composition, i.e. variation in Z, and/or thickness. The contrast from a variation in thicknesses can easily be understood as a higher sample thickness would scatter more of the incoming electrons. The end result of this is that in a bright field image the thinner areas will appear lighter compared to the thicker areas which will contain less intensity. Mass contrast is because heavier atoms will scatter the electrons to higher angles. This is

a consequence of the fact that a higher Z atom will have a larger atomic form factor, which essentially is the scattering power of that atom.

- Diffraction contrast (figure 2.22 B) is due to a difference in intensity in areas which fulfil the Bragg condition compared to those that don't. In areas where the Bragg condition is fulfilled, referred to as being "in Bragg", more of the electrons will be scattered into the diffraction pattern, and thus less will be scattered to higher angles which doesn't contribute to the image. This is not the case for scanning transmission electron microscopy, but this will be explained more in detail in section 2.4.2. Diffraction contrast can be present in the sample in different ways, e.g if the sample contains differently oriented crystal grains, or if the sample is strained or bent. In the case of strain and bending, the crystal lattice will be either compressed or stretched, resulting in a slightly different lattice spacing. Bending is most often introduced during sample preparation, whereas strain is usually a inherent material property.
- Phase contrast is mostly only present in high resolution TEM images (HRTEM) and is a consequence of the electron beam interfering with itself. The distance travelled by each of the diffracted beams will differ, depending on how far off the optical axis they are scattered. This introduces a phase shift in each of these beams. To make matters worse the objective lens introduces aberrations, which complicate the phase shift even more. HRTEM images are therefore not easily interpreted and to fully analyse them, they must be accompanied by numerical simulations.

2.4 Imaging & diffraction techniques

This section will explain some of the imaging techniques used in this thesis. The signals used to form these images are the elastically scattered electrons, both coherent and incoherent. Different diffraction modes will also be explained.

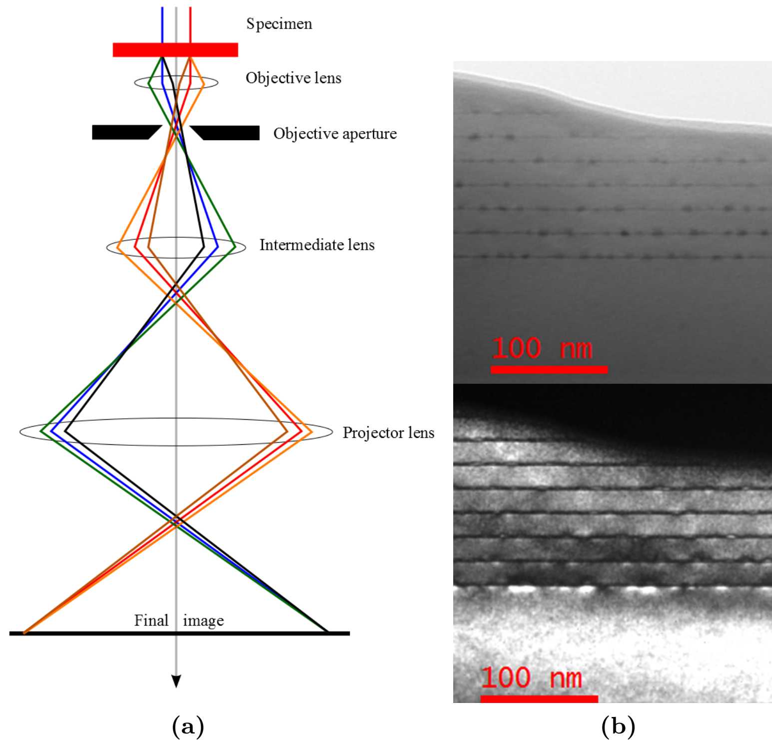


Figure 2.23: **(a)** Ray diagram for the lenses after the specimen. In the final image, beams originating from one point in the specimen are focused to a point. The objective aperture sits in the BFP, and controls which of the scattered beams form the final image. Adapted from figure 9.12 B in reference [7] **(b)** The difference between a BF (Top) and a DF image (bottom).

2.4.1 Bright field and dark field

In both bright field and dark field mode only the electrons that are scattered coherently, and therefore contributes to the DP, are used. What differs between these two, is which beams that are selected to form the image. In both cases an aperture is inserted into the back focal plane of the objective lens. This aperture is shown in figure 2.23a and is called the objective aperture. This same figure also shows the ray diagram of the image mode in a TEM. The aperture is placed around some diffraction spot, and thus limiting which of the scattered beams that will contribute to the final image. The contrast present in

the image will increase for smaller apertures, but the resolution will decrease.

In BF imaging the aperture is placed around the direct beam, reducing the amount of diffracted beams that contribute. This increases the contrast between areas that strongly diffract the beam, and those that don't. When forming a DF image, the aperture is placed around a diffracted beam. The image formed from this will then only correspond to atomic planes which are tilted in such a way that they scatter to this diffraction spot. This is useful when imaging samples consisting of different grains, materials of different crystal structure, or materials with the same crystal structure, but with different selection rules.

The name for both of these techniques relate to what the image looks like when viewing a hole in the sample. In bright field the hole will be bright, as all electrons are still in the direct beam, and in dark field all these electrons have been removed, making the hole appear dark. A BF and DF image are shown in figure 2.23b.

High resolution transmission electron microscopy (HRTEM) is a special case of bright field imaging at high resolution. Here the phase contrast becomes the main contrast mechanism in the image. Therefore, since the detector can only measure amplitude, HRTEM images cannot be easily interpreted without the use of simulations.

The periodicity of the structure can be extracted by Fourier transforming the image. This leads to a pattern which is similar to a diffraction pattern. A mask can then be applied to this image, and it is therefore possible to extract the periodic elements, which correspond to the spot selected by the mask, in the image. These periodic elements will correspond to atomic planes, and this makes it very easy to get information about dislocations, along the various planes. An example of this is shown in figure 2.24a, where a HRTEM image of a QD is (fast) Fourier transformed (FFT) (figure 2.24b). A mask is then applied to the FT image. Lastly an inverse (fast) fourier transformation (IFFT) is used on the mask, resulting in figure 2.24c. In the red rectangles in this image, dislocations in the atomic planes can be seen.

Much more can be said about HRTEM images, but this is the extent to which they were used in this thesis so they won't be explained in any more detail. For more information see volume 3 of Williams & Carter's book on transmission electron microscopy [7].

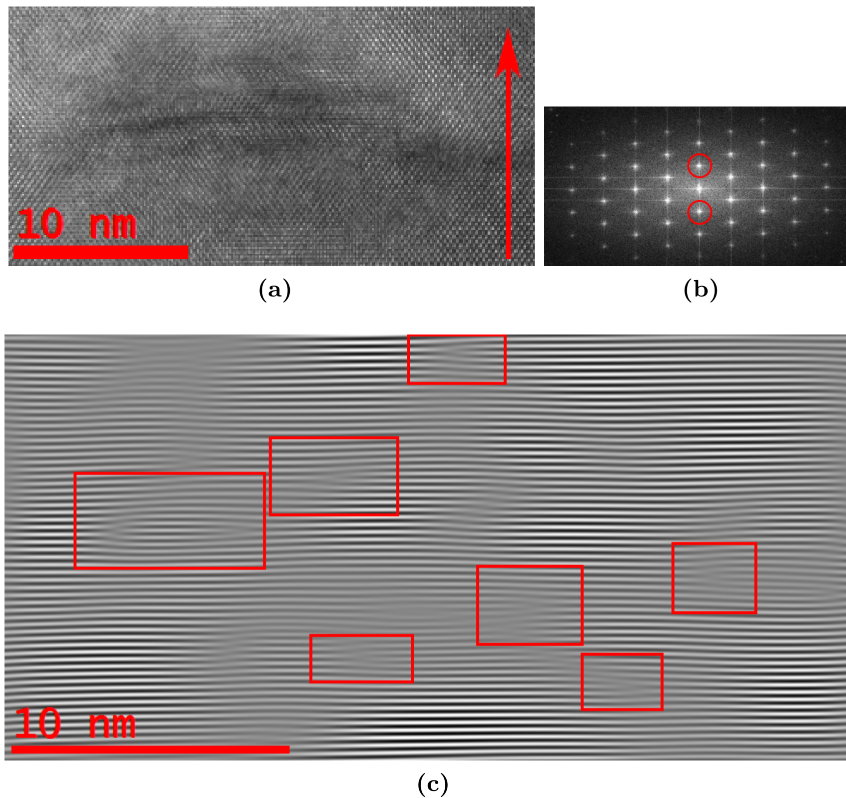
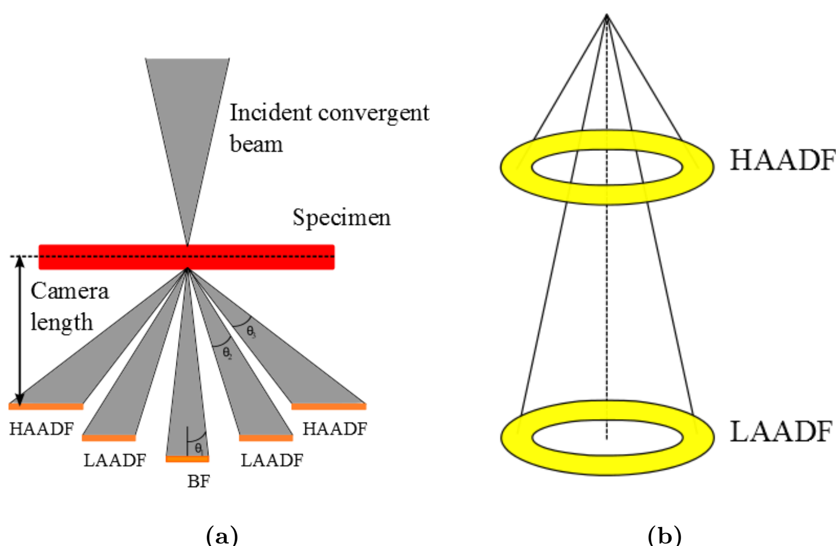


Figure 2.24: (a) A HRTEM image of a QD. The QD shows signs of strain, in the form of the varying dark and bright areas, and the outline of the QD is not easily interpreted from the image. It is also extremely hard to see the dislocations from this image. The red arrow indicates the growth direction **(b)** A Fourier transform of **(a)**, where the red circles indicate the areas chosen for the mask. **(c)** An IFFT has been applied to the mask chosen in **(b)**, showing the atomic planes perpendicular to the growth direction. In this images, dislocations can easily be distinguished and a few examples are shown in the red boxes.

2.4.2 Scanning transmission electron microscopy

Scanning transmission electron microscopy (STEIM) is a technique which utilizes a convergent beam, instead of a parallel beam as in ordinary TEM images, to form the images. In STEM mode this convergent beam, called a probe, is used to scan across the sample in a raster pattern. The probe is kept parallel to the optical axis in



(a) (b)

Figure 2.25: **(a)** A schematic showing the incident probe on the specimen, and the names of the different scattering ranges. The scattering range for the BF detector is $\theta < 10mrad$, for LAADF $10 < \theta < 10mrad$, and $\theta > 50mrad$ for HAADF. The difference between HAADF and LAADF is shown in **(b)**. The same detector is used for both HAADF and LAADF, but the camera length is changed. By decreasing the camera length electrons scattered to higher angles can be detected. Image adapted from figure 22.13 in [7].

order to ensure that the probe hits the sample at the same angle. The scattered electrons are then collected by a annular detector, which is centred around the optical axis. The image is formed by integrating the signals from the detector into a total intensity for each pixel of the image. This is done without the need for lenses, reducing the aberrations in the final image.

STEM can also be used in conjunction with energy dispersive X-ray spectroscopy (EDS) or electron energy loss spectroscopy (EELS), giving chemical composition with a high spacial resolution. These two techniques are explained in further detail in section 2.5.1 and 2.5.3, respectively.

The amount of electrons scattered to high angles is much lower compared to those scattered to lower angles by Bragg scattering. In order to gather enough intensity in each pixel, the time the probe dwells on each pixel can be significant, depending on the magnification used. Because of this, vibrations, temperature fluctuations, electromagnetic

and acoustic noise must be kept at a minimum, as this would shift the sample or introduce noise in the image. Keeping these effects as low as possible requires building the room with vibration damping and electromagnetic shielding, which is very costly.

There are mainly three types of detectors used to form a STEM images. A BF detector which collects the electrons, that are scattered to low angles, $\theta < 10mrad$, an annular bright field detector (ABF), which collects low angle scattered electron in an annulus and an annular dark field (ADF) detector, that collects the electrons scattered to higher angles. The ADF detector is a annulus, described by an outer radius and a inner radius. This determines the range of scattering angles the detector can pick up at a given camera length. By moving the detector closer to the specimen, which decreases the camera length, electrons scattered to higher angles can be detected. The ADF imaging technique can be divided into two categories, determined by the camera length used. It is called low angle annular dark field (LAADF) when the detector is placed such that it picks up electrons scattered between $10mrad$ and $50mrad$. At angles higher than $50mrad$ it is called high angle annular dark field (HAADF). The various scattering ranges are shown in figure 2.25a and in figure 2.25b the difference between HAADF and LAADF is shown.

2.4.2.1 Contrast in STEM mode

There are various contrast mechanisms which will be present in a STEM image, depending on the angle of scattering. Diffraction and strain contrast are scattered to low angles, and so these effects will be visible in BF and LAADF imaging modes. As the scattering angle increases, the electrons scattered due to diffraction will decrease, but the amount of electrons scattered due to Rutherford scattering will increase, which is dependant on Z. As angle of scattering is increases beyond $50mrad$, into the HAADF regime, the contrast will be a (almost) pure Z-contrast. But due to an effect called channelling, the contrast in a HAADF image won't necessarily be a pure Z-contrast. This is a highly complicated effect, and a thorough explanation can be found in [39].

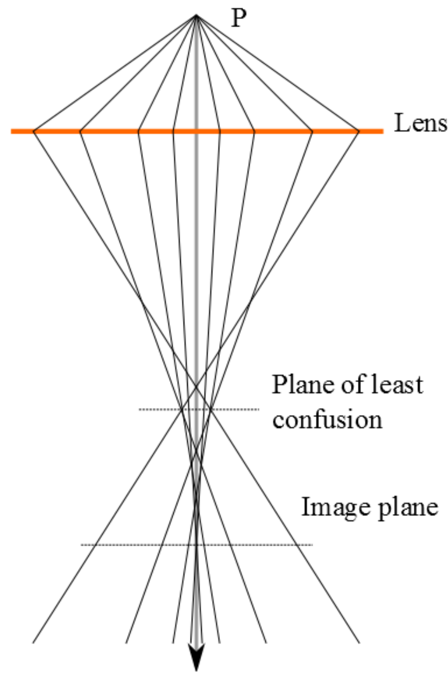


Figure 2.26: A schematic of the spherical aberration of an imperfect lens. The aberration causes a point in the specimen to appear to be a disk in the image plane. The plane of least confusion is where this disk is at a minimum, and the image plane is where the rays, closest to the optical axis, are focused to a point again. The radius of both of these disks, will be dependant on C_s and β. Adapted from figure 6.11 in [7].

2.4.2.2 Spherical aberration

Spherical aberrations occur for lenses that bend off-axis electrons more strongly towards the optical axis. As a consequence of this, a point imaged in the specimen, will appear as a disk in the final image. The intensity of this disk will be strong near the middle, surrounded by a halo of decreasing intensity. Spherical aberration is a major limiting factor as the magnification increases, and therefore represent a large decrease in resolution in HR(S)TEM images.

Figure 2.26 shows a schematic of this process. Here, the beams scattered of a point P, are collected by an imperfect lens. The image plane is defined by the near-axis electrons, as these are very close to the same behaviour as for a perfect lens. The plane of least confusion is where the spread of the beams is the smallest. The size of the disk,

at both the image plane and plane of least confusion, will be dependant on the angle of the beams collected by the lens, called the collection angle β, and a constant called the spherical aberration constant, C$_s$. This factor is constant for a particular lens. The radius of the disk in the image plane will be equal to

$$r_{sph} = C_s \beta^3 \tag{2.14}$$

where C_s and β are defined as above.

The spherical aberrations are corrected for by creating a concave lens, which spreads out off-axis beams, called C_s correctors. Adjusting this concave lens is a set of quadrupoles, hexapoles or octupoles electromagnets, where the current through them is controlled by a computer. These systems are very expensive, as the current needs to adjusted at very fine levels.

The C_s correctors work differently for HRTEM and STEM mode. In HRTEM they work by decreasing the size of the disk in the image plane and are therefore placed after the specimen. In STEM they work by decreasing the probe size and are placed before the sample. These are therefore called image correctors and probe correctors, respectively.

2.4.3 Electron diffraction

Electron diffraction can be done in mainly two modes in a TEM: With parallel incoming beam, and with a convergent beam. Parallel beam electron diffraction (ED) produces a DP with sharp spots, whereas the convergent beam electron diffraction (CBED) produces a pattern of disks. The ray diagrams for the two cases are shown in figure 2.27a and 2.27b, respectively.

The DP produced in the parallel beam mode can be indexed by the integers corresponding to the scattering plane, i.e. the Miller indices. For a cubic material, this can be done by the following method, here h, k, l will denote the Miller indices.

The atomic plane spacing for a cubic material is given as

$$d_{hkl} = \frac{a}{\sqrt{h^2 + k^2 + l^2}} \tag{2.15}$$

where a is the lattice spacing.

Figure 2.28 shows the geometry of the beam after it interacts with the sample until it reaches the screen (in diffraction mode). The length of the reciprocal lattice vector **g** (one the viewing screen) can then be

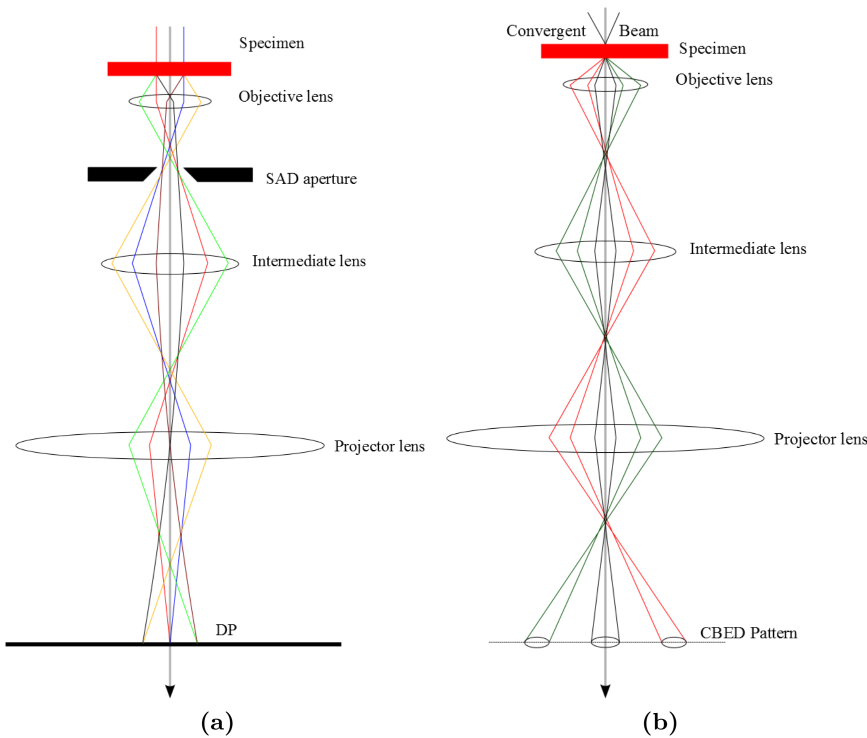


Figure 2.27: The ray diagrams for **(a)** parallel beam electron diffraction and **(b)** convergent beam electron diffraction. In **(a)** an aperture, which limits the area the DP is taken from, has been inserted. Images adapted from figure 9.12 B and 20.1 in [7], respectively.

found from purely geometrical considerations. As the angle θ is very small, the following approximation can be made

$$tan(2\theta) \approx 2\theta = \frac{|\mathbf{g}|}{L} \tag{2.16}$$

where L is the camera length, or the distance from the sample to viewing screen / projection lens.

Combining this with Bragg's law (equation 2.5) and using $sin(\theta) \approx \theta$, the following relation is found

$$|\mathbf{g}_{hkl}|d_{hkl} = \lambda L \tag{2.17}$$

The length $|\mathbf{g}_{hkl}|$ is what is measured on the screen and by taking the ratio between two diffracting vectors, the two constants vanish. Combining this with equation 2.15, gives

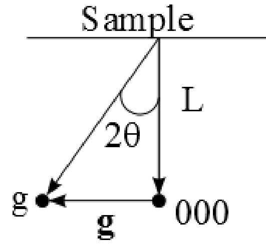


Figure 2.28: A schematic showing the geometry of the scattered electron beams after they interact with the sample. L is the camera lenght, which is the length from sample down to viewing screen or projection lens. 2θ is the scattering angle and the diffraction peak is labeled g, while the reciprocal lattice vector connecting the two spots are given in bold, i.e. **g**.

$$\frac{|\mathbf{g}_{h_1k_1l_1}|}{|\mathbf{g}_{h_2k_2l_2}|} = \frac{d_{h_1k_1l_1}}{d_{h_2k_2l_2}} = \frac{\sqrt{h_1^2 + k_1^2 + l_1^2}}{\sqrt{h_2^2 + k_2^2 + l_2^2}} \tag{2.18}$$

By using this relation, the magnitude of the hkl values for each diffraction peak can be found.

Taking the dot product of each reciprocal lattice vector, the sign of the indices can be found. The dot product between two vectos, $\mathbf{g_1}$ and $\mathbf{g_2}$, can be written as $\mathbf{g_1} \cdot \mathbf{g_2} = |\mathbf{g_1}||\mathbf{g_2}|cos(\theta_{12})$, where θ_{12} is the angle between the two vectors. As these two vectors are reciprocal lattice vectors, the expression for θ_{12} can be written as

$$cos(\theta_{12}) = \frac{h_1h_2 + k_1k_2 + l_1l_2}{\sqrt{h_1^2 + k_1^2 + l_1^2}\sqrt{h_2^2 + k_2^2 + l_2^2}} \tag{2.19}$$

Using this equation and the previous one in a trial and error method will produce a indexed pattern which matches the measured angles. Note that there are some ambiguity in these indices, depending on the symmetry present in the DP.

As the incoming beam will be parallel to all the diffracting planes on a zone axis, that direction of the beam and also the zone axis, can be identified. This is done by choosing two reciprocal lattice vectors in the DP, and taking the cross product. This will produce a vector that is parallel with the incoming beam. The direction of the zone axis is written as [UVW], where U, V, W are integers in x, y, and z direction respectively, e.g. the zone axis [110] will have an incoming beam travelling parallel to the 110 plane of the unit cell.

In the case of parallel beam ED, an aperture can be inserted in the image plane, in order to restrict the area of the sample used to form the DP. This aperture is called the selected area diffraction (SAD) aperture. Used in this way the DP is referred to as a selected area diffraction pattern (SADP). These are mainly used for determining the crystallographic orientation, lattice parameters, or dark field imaging. A problem here is that the smallest SAD apertures are still quite large, and it is therefore not possible to study single nano-particles or similar structures. This can be overcome by using CBED, as here the convergent beam can be made very small. CBED can be used for everything parallel beam ED can, except DF imaging, has a much higher spacial resolution. A CBED pattern is less useful in thin regions of the sample, as it is probable here that the electrons will only scatter once, making the information available here the same is for a SADP. This is because there are no dynamical effects in this region and all features in the CBED disks will therefore disappear.

The CBED pattern will contain features from different effects, such as Kikuchi bands and HOLZ lines. Here only the effect of the HOLZ lines on the CBED pattern will be explained.

In the CBED disks of a diffracted beam, there will be a bright line, corresponding to the beams with an angle equal (and very close

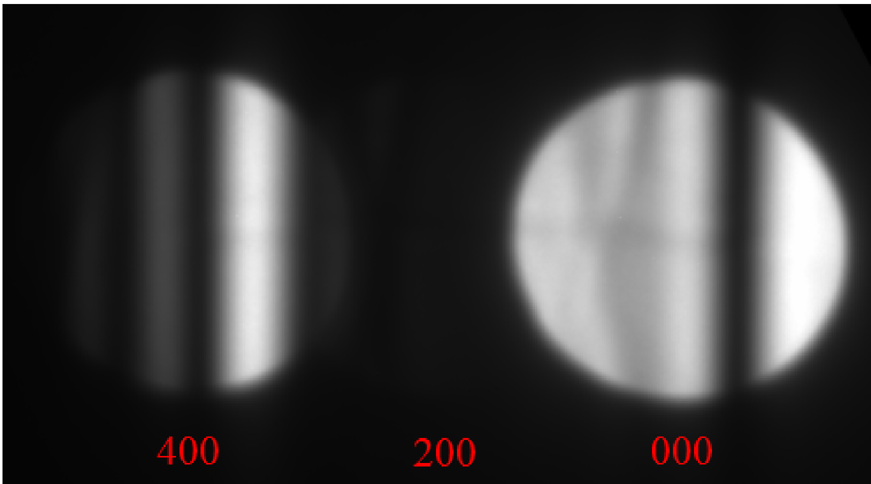


Figure 2.29: The CBED pattern in a 2-beam condition with the 400 disk and central disk. The fringes present in this disk are the K-M fringes. The 200 disk is barely visible in the image.

to) the Bragg angle. The central disk will contain corresponding dark lines, which is caused by other diffraction peaks being in Bragg, thus "stealing" some of the intensity. At a zone axis, the central peak will have several of these dark lines, intersecting each other. The origin of these are from HOLZ diffraction spots. The lines are therefore referred to simply as HOLZ lines. These HOLZ lines can produce very striking pattern in the central disk when the sample is tilted to a zone axis.

When tilting the sample to a 2-beam condition, multiple dark lines will appear in the disk of the diffracted beam. The 2-beam condition means that the sample is tilted in such a way, that mainly two spots are in Bragg. In SADPs this can be done to reduce dynamical scattering, but in CBED this is done to produce Kossel-Mollenstedt (K-M) fringes, which can be used to determine the thickness at the area where the probe has been placed. Figure 2.29 shows several of these dark lines. This technique can only be used as long as the thickness is above a certain size, and so this technique cannot be used in thin areas of the sample.

CBED can also be used to determine the polarity in non-centrosymmetric crystal structures. This technique will be explained qualitatively with GaAs as the example. For a more thorough discussion on the polarity determination in III-V semiconductors, see the paper *Non-centrosymmetric Effects and Polarity determination in III-V Semiconductors* by K. Marthinsen et. al [40].

By tilting the sample away from the [011] zone, around the (100) direction, to a conjugate four beam case, strong asymmetry effects can be observed between the ±200 disks, even though GaAs deviate very little from centrosymmetry. What is meant by the conjugate four beam case, is that the CBED pattern when imaging the $\bar{2}00$ disk is the conjugate of the CBED pattern when imaging the 200 disk. The four beam CBED pattern in these two cases is shown schematically in figure 2.30a and 2.30b, respectively. The basis for these simulations, were Ga atoms at (0,0,0) position and As at $(\frac{1}{4},\frac{1}{4},\frac{1}{4})$ and the zone axis used were [011].

The polarity is determined by looking at the intensity difference in the 200 disk and the $\bar{2}00$ disk. From the figures, it can be seen that one of the reflections has a white cross, and the other a black cross. This is caused by destructive and constructive interference from scattering into the 200 ($\bar{2}00$) reflections.

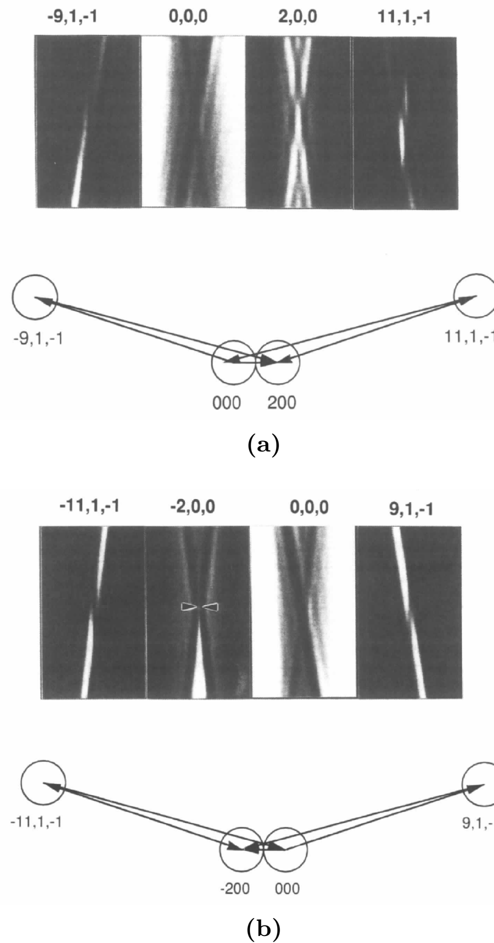


Figure 2.30: Calculated four beam CBED intensity distributions from [40]. **(a)** Shows the 200 disk and **(b)** the $\bar{2}00$ case. In both cases the scattering paths are indicated.

There are two main scattering paths into the 200 or $\bar{2}00$ reflection. The beam can scatter directly into both 200 and $\bar{2}00$ reflections, or by via reflections with h, k, l values whose sum is equal to $4n + 1$ for scattering into $\bar{2}00$, or $4n - 1$ for scattering into the 200 reflection. Due to a phase difference in these scattering paths, there will be constructive and destructive interference. This can be seen from figure 2.30a, where there is constructive interference in the 200 reflection, and in figure 2.30b, where the scattering causes destructive interference in

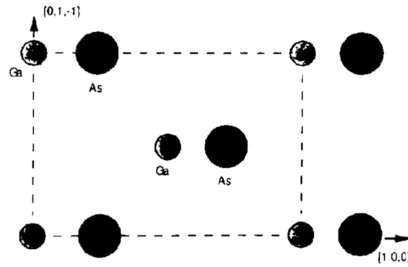


Figure 2.31: A projection of the unit cell of GaAs along the [011] zone axis, with Ga at (0,0,0) position and As at (($\frac{1}{4}$,$\frac{1}{4}$,$\frac{1}{4}$) [40].

the $\bar{2}00$.

The positions of the white and black cross, uniquely determines the polarity of GaAs. If the white cross is to the right of the direct beam (as seen in figure 2.30a, the As atoms will be to the right of the Ga atoms in the sample. In figure 2.31 the projected unit cell in the [011] zone axis is shown for this case.

2.5 Spectroscopy techniques

The inelastically scattered electrons are defined by the fact that they lose energy in the scattering process. This energy is lost through either the generation of characteristic X-rays, Auger electrons, secondary electrons, or bremsstrahlung X-rays. Here only the characteristic X-rays will be discussed, which are the basis for a technique called energy dispersive X-ray spectroscopy (EDS). The inelastically scattered electrons that are transmitted through the sample can be utilized in a technique called electron energy loss spectroscopy (EELS). Also electrons that have lost no energy and have been transmitted through the sample are important in EELS.

2.5.1 Energy dispersive X-ray spectroscopy

When an incoming electron scatters inelastically with a sample, the electron interacts with the electrons in the inner shell of the atom. If this interaction transfers a high enough energy to a inner shell electron, it can eject it from the atom, in a isolated atom case, or above the Fermi level for a solid. This leaves the atom in an ionized state, causing

an outer shell electron to fill this hole. By this transition a Auger electron or an characteristic X-ray is generated. The energy of the generated photon will have an energy equal to the difference between its originating shell and the one it fell into. As a consequence of this, the generated X-rays will be different for different atoms. By detecting these X-rays, the chemical composition of a sample can be determined. These X-rays are picked up by a detector, positioned above the sample stage. The EDS detector can be operated in both TEM mode, where it provides information of large areas, or in STEM mode, where it is possible to determine the atomic composition on an atomic scale. For a more thorough explanation of the working principles of EDS and experimental set-up, see volume 4 of Williams & Carter's book on transmission electron microscopy [7].

2.5.2 Electron-probe micro analysis

An electron-probe micro analyser (EPMA) is used for non-destructive chemical analysis of solids. The spatial resolution is around 1-2 μm, but the energy resolution is much higher compared to EDS in TEM. It operates on similar principles as the EDS, with an electron beam impinging on the surface of the material, causing the generation of

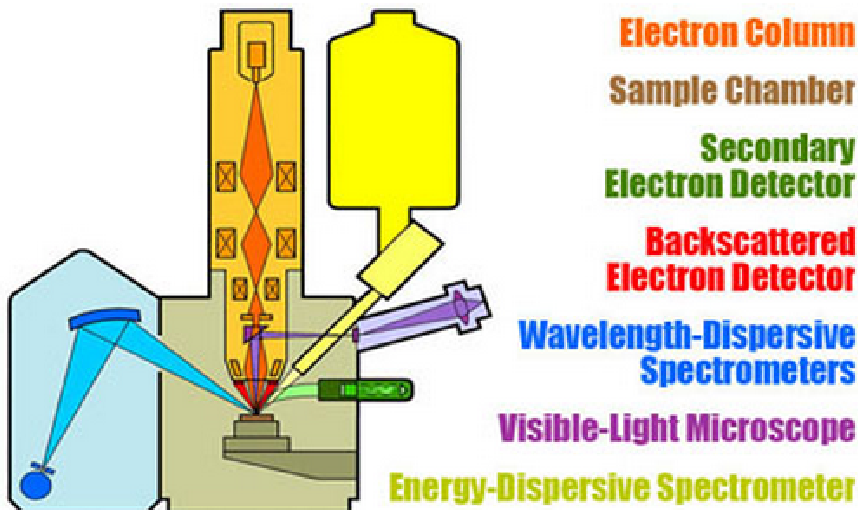


Figure 2.32: A schematic of a typical EPMA setup. The various generated signals are shown, with their respective detection area [41].

characteristic X-rays.

The electron beam is generated by a heated W filament, and a cathode, which accelerates the electrons. A set of electromagnetic lenses then focuses the beam on the surface of the sample. The sample is placed in a chamber under vacuum on a stage movable in all directions. A range of crystal spectrometers are then used to detect the characteristic X-rays, enabling quantitative chemical analysis with a high sensitivity. The probe in an EPMA extends about 1 μm into the sample, and it is necessary that the sample being probed is homogeneous over this region. The samples prepared for EPMA must also have a smooth surface and any irregularities must be removed by polishing.

The EPMA are able to perform a variety of techniques, but it was only used to perform quantitative chemical analysis in this thesis [42].

2.5.3 Electron energy loss spectroscopy

Electron energy loss spectroscopy uses the electrons that have been transmitted through the sample, both those that have lost no energy and the inelastically scattered ones, in order to provide information about the chemistry of a sample. EELS can extract a wealth of information, related to the composition of the sample, amongst these are the free electron density, specimen thickness, and the band gap just to name a few. EELS analysis can detect and quantify all the elements in the periodic table, compared to EDS which can only detect elements above Li. The spatial resolution and sensitivity is also superior to EDS.

In EELS the transmitted electron beam is limited by an entrance aperture, allowing only electrons scattered to very low angles through. These electrons then enter a magnetic prism, which alters the path of electrons by at least 90°. The electrons that has lost no energy are deflected less than those who have, creating a spread in the incoming beam as a function of their energy loss. These are then passed through an optics system, before being recorded on a CCD or a photodiode (see figure 2.33). The energy the electrons has lost will depend on their composition and other factors. In this thesis, EELS were only used for thickness determination, and therefore the principles of EELS won't be explained in any further details.

Thickness determination by EELS is done by recording an energy

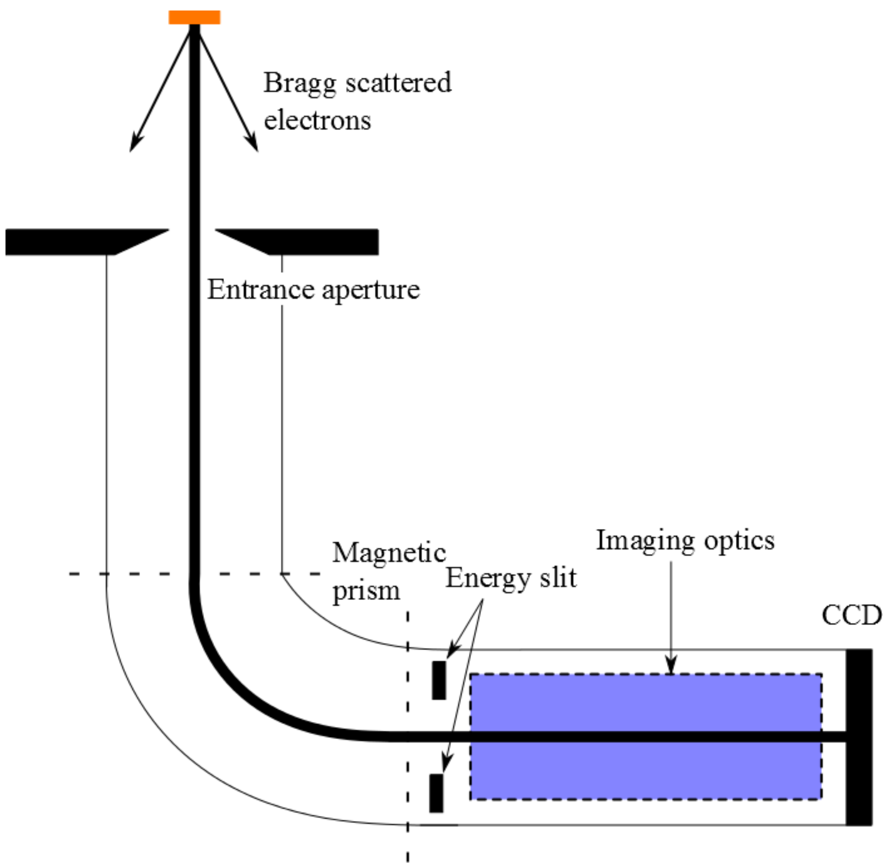


Figure 2.33: A schematic of the GIF system, which is a EELS system with an energy slit for Energy-filtered TEM. The incoming beam is deflected by a magnetic prism, before it is sendt through an imaging system and recorded on a CCD camera. An entrance aperture blocks most of the Bragg scattered electrons, but this is dependant on the camera length used, and which mode the GIF/TEM is operated in. Adapted from figure 37.15 A in [7]

loss spectrum and then integrate the total area under the zero-loss peak and comparing this to the integrated area of the whole spectrum. The zero-loss peak (ZLP) is the peak in the spectrum associated with the electrons that lose no energy. By this method, the thickness, t, is given by [43]

$$t = \lambda ln(\frac{I_t}{I_0}) \tag{2.20}$$

Here, λ is the mean free path, I_t is the intensity of the whole

spectrum and I_0 is the intensity of the ZLP. This equation provides a relative thickness measurement and in order to find an absolute value, the λ value must be calculated or found experimentally. By scattering theory, the value for λ can be approximated by the following equation [43].

$$\lambda \approx \frac{106F(E_0/E_m)}{ln(2\beta E_0/E_m)} \tag{2.21}$$

where F is a relativistic factor (0.618 for a voltage of 200kV), E_m is the mean energy loss, E_0 is the energy of the incident electrons, and β is the collection angle. E_m is given in eV, E_0 in keV, β in mrad and λ in nm. For electrons with $E_0 = 100keV$ this approximation is valid for β up to 15mrad. E_m can be calculated from the approximate formula [43]

$$E_m = 7,6Z^{0.36} \tag{2.22}$$

where Z is the atomic number. If the material is a composite an effective atomic number can be used, defined as

$$Z_{eff} = \frac{\sum_i f_i Z_i^{1.3}}{\sum_i f_i Z_i^{0.3}} \tag{2.23}$$

where f_i is the atomic fraction of element Z_i.

For more information on the working principles of EELS and applications see R. F. Egertons book on *Electron energy-loss spectroscopy in the electron microscope* [43].

2.6 Molecular beam epitaxy

Molecular beam epitaxy is a method for depositing thin single crystals on a substrate. The technique consists of a ultra high vacuum (UHV) chamber and several sources, which depends on the experimental setup. These sources are controlled by mechanical shutters, and provide a high precision in deposition and doping. The technique is used to grow thin epitaxial structures made of semiconductors, metals or insulators. The crystallization is done by keeping a substrate at high temperatures while opening the source shutters. The atoms from the sources adhere to the substrate layer by layer, creating a very even surface. The growth rate depend on the evaporation rates of the sources, but generally it's around 1 monolayer (ML) per second [44].

In order to have epitaxial growth the crystal structure should be similar, but it is still possible to grow structure with small lattice mismatch. Whether this strain will be tensile or compressive will depend on the order the materials are grown [44]. If a material with smaller lattice constant is grown atop a material with higher lattice constant, the strain in the growing film will be tensile. The reason for this is that the in-plane lattice constant will be the same as the bottom material, forcing the lattice constant in the growth direction to become smaller. In the other situation, the strain will be compressive, as the in-plane lattice constant of the film will be smaller, forcing the lattice constant in the growth direction to stretch.

The materials in this thesis were grown by MBE in the Stranski-Krastanov (SK) growth mode [44]. This utilizes the lattice mismatch of two materials to create 3D islands. Strain between the two materials causes the deposited material to relax into 3D islands, or QDs depending on the size, after a certain amount has been deposited, in order to reduce the strain energy. Some of the material deposited don't contribute to the 3D islands, but instead form a layer called the wetting layer.

The lattice mismatch between InAs and GaAs is about 7%, making this method suitable for growing InAs QDs in a GaAs matrix. Under optimal growth conditions, these QDs can be grown defect free as long as they are kept under a certain size. Several studies also report that the volume of the QDs are larger than the amount of InAs deposited [45–47] indicating that Ga atoms diffuse into the QDs [45, 47]. Other studies indicate a varying critical thickness in the transition between deposited film and QD formation, indicating an In-Ga intermixing as the wetting layer forms [48, 49].

For a more detailed explanation of MBE and the

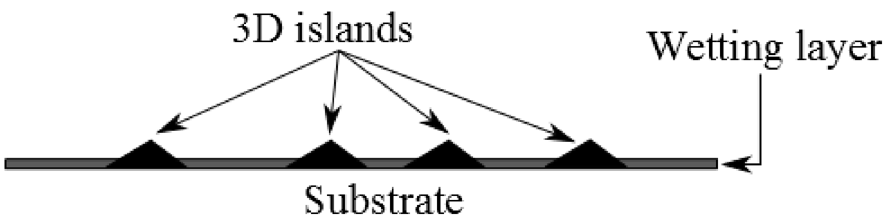


Figure 2.34: The formation of 3D islands, or QDs, by the SK growth mode. This leaves behind a wetting layer, exagerated in the image, and the self-assembled QDs

self-assembled growth of InAs/GaAs QDs see M.Herman and H. Sitter's book *Molecular beam epitaxy - Fundamentals and current status* [44], and the chapter on the InAs/GaAs (001) Quantum dot transition in *Self-assembled quantum dots* [46].

2.7 Notation

The different notations used in the context of TEM can often be slightly confusing. This is a brief summary of the notation conventions used in this thesis.

Diffraction spots in a diffraction pattern are indexed by Miller indices hkl, without brackets, and if the indice is negative it is denoted either by a minus sign or a bar above the integer. When the indices are enclosed by curly brackets, the sign of the indices haven't been uniquely determined, e.g {111} spot could be 111 or $\bar{1}11$. The curly brackets can also be used to denote a set of planes in real space. When they are surrounded by parenthesis, i.e. (hkl), they denote uniquely determined planes in real space and when they are enclosed by square brackets, [hkl], they denote directions in real space which corresponds to a plane in reciprocal space. Enclosed by triangular brackets, i.e. <hkl>, they denote the set of all hkl values that are equivalent to [hkl] by symmetry, e.g <100> could either be along [100] direction or [$\bar{1}00$].

Chapter 3

Method and experiment

3.1 Samples

The samples studied in this thesis were grown by MBE in the SK growth mode at Chalmers University of Technology in Sweden. The full structure needed for an IBSC were grown, including the p- and n-emitters, and the QD layers.

The layer structure can be summarized as follows: First a buffer

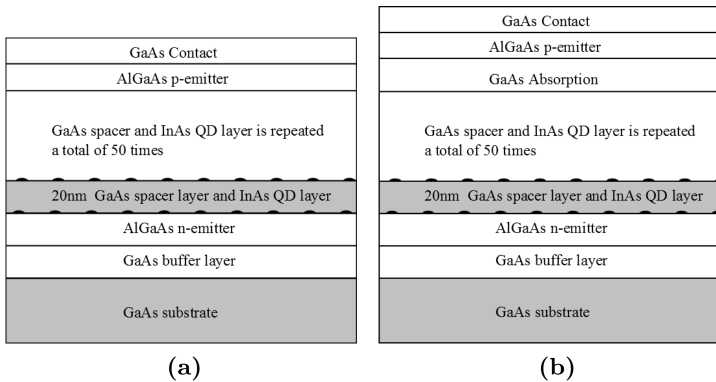


Figure 3.1: A schematic of the structure for **(a)** sample A and **(b)** sample B. The samples are grown on a substrate of GaAs with a buffer layer above and a contact on top of the sample, of the same material. The n- and p-emitter are AlGaAs and the QDs are InAs. The spacer layer is the main difference between the two samples, GaAs in sample A and GaNAs in sample B. **(b)** Sample B also contains an absorption layer. The QDs are grown by the Stranski-Krastanov growth mode.

Table 3.1: A table showing the material parameters of samples A and B. The (**A**) and (**B**) denotes which sample the layer belongs to, and are separated from the rest by a horizontal line. Those with no notation are the same for both samples. The concentration column indicates the amount of dopant atoms in the specific layers.

Layer	Material	Thickness [nm]	Dopant atom	Concentration [1/cm^3]
Substrate	GaAs	$350 \cdot 10^3$	Si	$2\text{-}4 \cdot 10^{18}$
Buffer	GaAs	100	Si	$3 \cdot 10^{18}$
N-emitter	$Al_{0.25}Ga_{0.75}As$	100	Si	$1 \cdot 10^{18}$
QD	InAs	2 ML	-	-
P-emitter	$Al_{0.25}Ga_{0.75}As$	30	Be	$2 \cdot 10^{18}$
Contact	GaAs	50	Be	$3 \cdot 10^{19}$
Spacer (**A**)	GaAs	20	-	-
Spacer (**B**)	GaNAs	20	-	-
Absorption (**B**)	GaAs	100	-	-

layer was grown on a GaAs (100) substrate, in order to ensure a perfect crystal structure. On top of this, a n-emitter of $Al_{0.25}Ga_{0.75}As$. Next a repeating structure consisting of 50 layers of QDs, each separated by a spacer layer, making the total number of spacer layers 49. Lastly a p-emitter of $Al_{0.25}Ga_{0.75}As$ and a Be doped GaAs contact were grown. The difference between the two samples studied in this thesis lies mainly in the spacer layer, where one sample had a GaAs spacer layer and the other a GaNAs spacer layer. The other difference was that the sample with GaNAs spacer layer also had a 100nm thick absorption layer of GaAs between the 50 QD layers and the p-emitter. Besides these difference, the thicknesses of each layer were the same. These two samples will from here on out be referred to as **sample A for the one with GaAs spacer**, and **sample B for the one with GaNAs spacer**. The schematic for the structure of sample A and B can be seen in figure 3.1a and 3.1b, respectively.

The various parameters related to the structure are summarized in table 3.1, and the growth temperatures corresponding to the different layers are given in 3.2.

The InAs QD layers were δ doped with Si, with a concentration equal to $3 \cdot 10^{10}$ cm^{-2}. This was done in order to half fill the electron states in the intermediate band, introduced by the QDs.

In both samples the beam equivalent pressure (BPE) during deposition of Ga was $4.6 \cdot 10^{-7}$ Torr, $2.4 \cdot 10^{-7}$ Torr during deposition of In, and $1.1 \cdot 10^{-7}$ Torr during deposition of Al. The BPE during deposition of As was dependant on what was being grown. During GaAs growth,

Table 3.2: The different temperatures during the growth of the various layers.

Sample	Layer	Growth temperature [°C]
Both	N-emitter	612
	P-emitter	612
A	Buffer	604
	Spacer	530
	Contact	607
B	Buffer	594
	Spacer	539
	Absorption	610
	Contact	603

it was $1.3 \cdot 10^{-6}$ Torr, while during AlGaAs growth it was $1.5 \cdot 10^{-6}$ Torr. In sample B the pressure of N were $6.95 \cdot 10^{-5}$ Torr.

3.2 Sample preparation

In TEM sample preparation it is important to make the sample "electron transparent". This means that the sample needs be prepared in such a way that most of the electrons hitting the sample are transmitted.

Electrons interact with both the electron cloud around the nucleus and the nucleus. The interaction between the electron beam and the material is therefore very strong and much of the electron beam is backscattered or scattered to large angles. It is therefore important to give the beam a small as possible "interaction" volume. This is accomplished by thinning the sample, since the area the beam covers is controllable.

How thin a sample needs to be will vary depending on the material. For certain aluminium materials thin would be a thickness up to $\backsim 1\mu$m, whereas for steel it would be around a couple hundred nm [7, p. 11]. In general though, thinner is always better. There are various ways of doing this, but in this thesis only one method was used.

The samples studied in this project was as explained above, and the area of interest in the sample was therefore a cross-sectional view of the sample. This makes it possible to look at the interfaces of the layers and also study the size and density of the QD's, as long as the

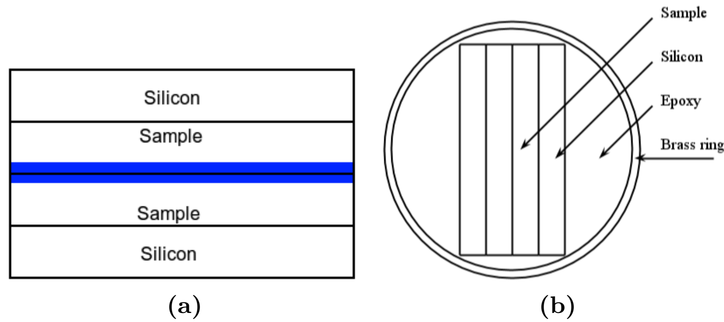


Figure 3.2: (a) Schematic view of the sandwich after gluing the rectangular pieces together. The area of interest, i.e. the film, is indicated in blue. **(b)** Schematic of the sandwich after it has been glued inside the tube and cut into disks.

area under study is sufficiently thin (so there is no overlapping of the QD's). The method used to make these cross-section samples is the so-called sandwich method. It mainly consists of gluing the sample into a tube with a inner diameter of 2 mm, cutting this into disks and then thinning it sufficiently. In more detail the steps are as follows: First a piece of glass is heated to about 130°, wax is then melted onto the glass piece. The sample, which in this case was a quarter circle wafer with a 2 inch diameter, is then placed on the melted wax with the film side pointing down on the glass. After making sure there are no air pockets under the sample, the glass piece is removed from the heater and allowed to cool.

After the wax is solidified the wafer is cut into rectangular pieces, 2 mm wide and of varying length, with a 150μm Allied High Tech diamond saw. The same process is then done for a Silicon wafer, to fill the extra space within the tube. The reason this is done is to reduce the amount of glue inside the tube. This will reduce the drift in the sample, due to there being less isolating material.

These rectangular pieces are then glued together with EpoxyBond 110 from Allied High Tech in the following order: first a silicon piece, then a sample piece with the film side pointing up, then a sample piece with the film pointing down, and lastly a silicon piece. Glue is applied between each of these pieces. The schematic of this is shown in figure 3.2a.The sandwich is then placed into a holder, which holds the sample under pressure while it is being heated. The trick here is to add enough pressure that the glue layers becomes quite thin, but

not too much as that could result in cracking of the sample.

The next step is then to polish the sandwich until it fits inside a brass tube with 2 mm inner diameter. A brass tube is cut into an appropriate length and then glued to a small piece of silicon and hardened by applying heat. The glue used for this step and the next is the same is in the previous step. After this the whole brass tube is filled with glue. To avoid air bubbles inside the tube, the glue is inserted by using a pipette. The pipette is filled with glue, allowing for no air bubbles in the lowest portion. It is then lowered into the bottom of the brass tube. While squeezing the pipette to extract the glue, the pipette is slowly raised. After this process the polished sandwich is cleaned with ethanol, and dried, before it is inserted into the tube. The glues is then hardened by heating it.

After the glue has hardened, the cylinder is cut into disk of about 1 mm thickness. This is done by the same saw as mentioned previously. The disks will then look like figure 3.2b. The disk that looks best, i.e. the one with the least cracks and air bubbles, is then glued to a stub and polished down to about $70-80\mu$m using successively fine grinding paper, from grain density 2500 cm^{-2} to 4000 cm^{-2}. After polishing one side until it's free of cracks, scratches and air bubbles the sample is flipped and polished down to about $70-80\mu$m. During the polishing process water is used as a grinding solvent. The thickness is found by comparing the focus point of the sample and the stub with a VLM.

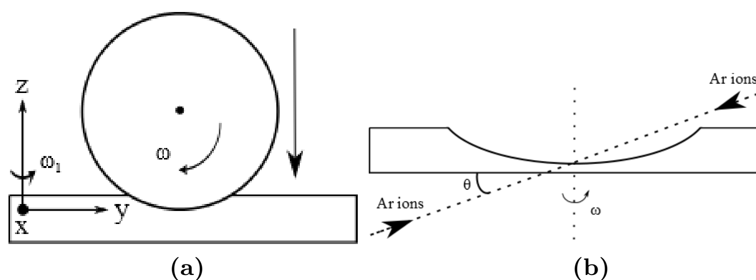


Figure 3.3: (a) The geometrical set up for the dimpling process. A spinning grinding wheel is slowly lowered into the sample, indicated by the arrow, while grinding away material. The sample is also rotating around the center of the sample. **(b)** The thinned sample from the dimpling is being bombarded with Ar ions. These sputter away material, leaving a thin area near the center of the sample. The angle θ is the incident angle of the Ar ions. Also in this case the sample is rotating around its center.

The sample is then transferred to a Fichione EAF Instruments Model 2000 dimpler. This machine makes a dimple in the center of the sample. After dimpling, the bottom of the dimple should be $20-30\mu$m for these samples. For less brittle materials the bottom of the dimple can be made even thinner. The dimple is created by having a rotating wheel "dig" into the sample while the sample stage is rotating. This process is shown schematically in figure 3.3a. Grinding diamond paste is applied to the wheel during this process. The sample preparation done in this thesis utilized grinding paste with 1μm grain size.

The last step of the process is to ion mill the sample. The thinned disk is then put into a Precision Ion Polishing System Model 691 from Gatan, cooled with liquid Nitrogen. The sample material is sputtered by sending a beam of Argon atoms at the sample. The Argon atoms are accelerated by a voltage of 3.5kV. The angle of the beam used for theses samples were $4°$. This can be seen in figure 3.3b. Our goal of a sufficiently thin sample is achieved by milling a small hole in the bottom of the dimple. If the dimple is centered on the film, the resulting hole will have four possible areas of interest that are thin enough. After the hole has formed, the accelerating voltage is decreased in steps to remove as much of the amorphous layer, that formed during the high voltage part of this process, as possible.

In the case of samples preparation of EPMA, all that was done was to cut a small piece from the wafer, and then wash it with ethanol. This was sufficient, as the surface of the sample was smooth enough for analysis, and polishing it with even 1 μm polish would have ruined the interesting area.

3.3 Experimental equipment

The instruments used in this thesis were the three recently installed through the NORTEM project [50]. A Jeol 2100 with a thermionic source was used for low magnification imaging, CBED, thickness determination and some EDS work. This microscope is equipped with an orius 2k CCD camera for imaging, a Gatan GIF system with a 2k CCD detector, and Oxford X-max 80 SDD EDS detector. The microscope is also equipped with an ADF detector and a BF detector for STEM.

HRTEM, STEM and EDS-STEM were perfomed on a Jeol 2100F and a Jeol ARM-200F. The Jeol 2100F is equipped with a schottky

FEG, a bottom mounted Gatan 2k ultrascan CCD camera, and the same EDS detector as the one on 2100. The STEM detectors are also the same as for 2100.

The ARM-200F is a C_s corrected instrument for both TEM and STEM mode, capable of sub-Ångstrøm resolution, with a CFEG source. A bottom mounted 2k UltraScan XP CCD and a side mounted 2k Orius CCD camera were used for standard imaging. The instrument is fitted with an Centurio SDD EDS and a Quantum GIF with DualEELS and 2k CCD. It is also equipped with two BF detectors, one annular bright field (ABF) detector, a ADF detector and two HAADF detectors for STEM work.

All three microscopes were operated at 200 kV, and a double-tilting sample holder, capable of tilting $\backsim \pm 30°$ in both directions, were used for all imaging and spectroscopy. The ARM-200F was operated by Dr. Per-Erik Vullum, adjunct associate professor at the Department of Physics at NTNU and research scientist at SINTEF.

The EPMA analysis were perfomed on a JEOL JXA-8500F, operated by Morten P. Raanes at the department of Materials Science and Engineering at NTNU.

Chapter 4

Results and discussion

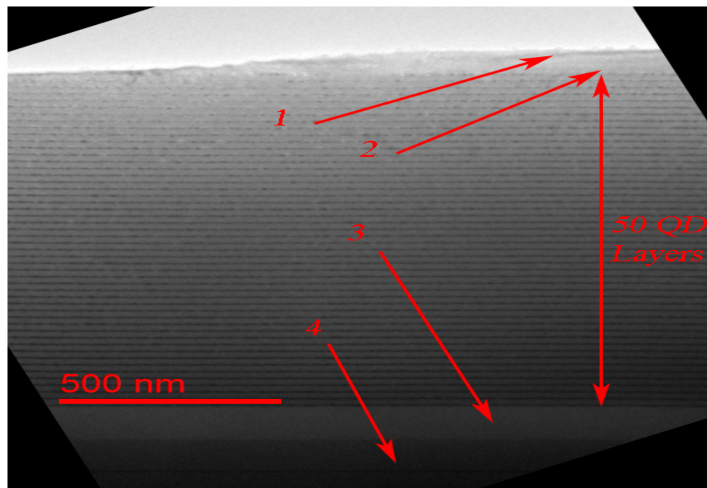
In this chapter the results obtained during the thesis work will be presented. After each result is presented, a detailed discussion will follow.

4.1 Structure

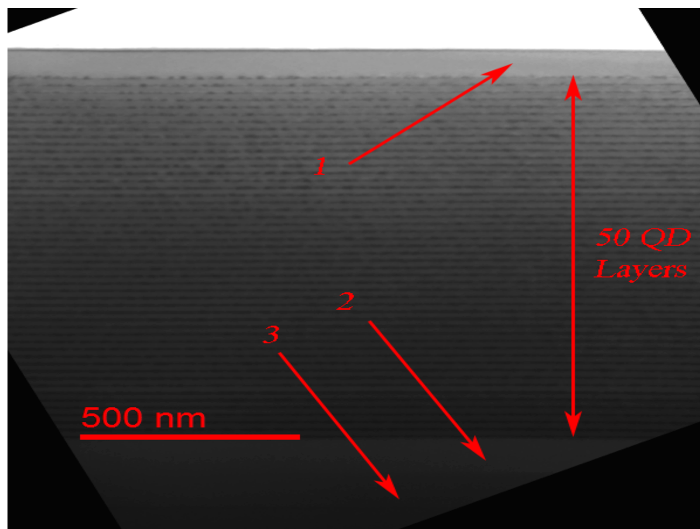
Images of both samples were taken at relatively low magnification in order to confirm that the materials were structured according to the growth specifications. Figure 4.1a shows the structure for sample A. Here the substrate is located at the bottom of the image, and above this is the buffer layer of GaAs (arrow labelled 4 in the figure) and the n-emitter of AlGaAs (arrow 3). These layers are easily differentiated as the contrast between them is quite large due to a difference in atomic number, giving a sharp interface. Since these are cross-sectional TEM samples, measuring the length of the layers in the vertical direction (assuming the images are rotated as in figure 4.1b) will yield the thicknesses of these layers.

The thicknesses were measured to be approximately 104.3 nm and 107.7 nm for the buffer layer and n-emitter respectively. This is close to the expected thickness of 100 nm.

The layers were then counted, giving the expected number of 50 QD layers. The thickness of all the spacer layers combined gives a value of 1084.1 nm. From this, assuming that the thickness of the spacer layers are the same (justification for this will be given later in this section), the spacer layer thickness is found to be approximately 22.1 nm, similar to the expected value.



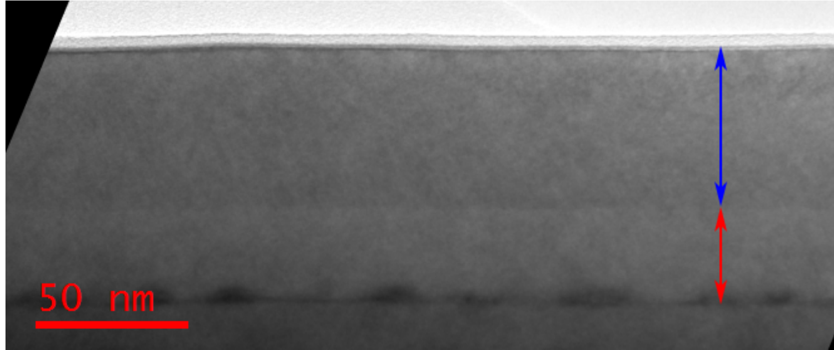
(a)



(b)

Figure 4.1: Low magnification BF-TEM images taken taken on the Jeol 2100. These images are taken from areas with no defects, showing the ideal situation for these materials. **(a)** Image from sample A, the arrows labelled 1, 2, 3 and 4 show the buffer, n-emitter, p-emitter, and contact layers, respectively. **(B)** Image from sample B, the arrows labelled 1, 2, 3 and 4 show the buffer, n-emitter, and top layers, respectively.

(a)



(b)

Figure 4.2: BF images taken at different magnification, for both samples, showing the top layers. **(a)** sample A, here the top QD layer can be seen, followed by two layers. **(b)** sample B, showing the two top layers and one QD layer. There is a thin layer on top of these two layers, which is an amorphous layer introduced during sample preparation. In both images, the red arrow indicates the p-emitter and the blue arrow indicates the contact.

Measurement of the thickness of the top layers, were found from figure 4.2a, as the contrast between the two layers are quite poor in figure 4.1a. The contrast between these two layers are quite poor, even though the atomic mass is different. This is because the sample is very thin in this area, reducing the ratio between scattered beams and direct beam. For figure 4.2a the contrast is a little better, as the image is taken at higher magnification, reducing the total intensity. Thus giving a higher ratio between scattered beams and direct beam. The thickness were found to be 29.8 nm for the p-emitter (red arrow in the image) and 48.4 nm for the contact layer (blue arrow in the image).

Figure 4.1b shows the structure of sample B. It is clear from the image that the GaAs absorption layer the sample should have contained is not there, as the total thickness of the layers above the 50 QD layers, is only 83.1 nm, whereas the absorption layer should have been 100

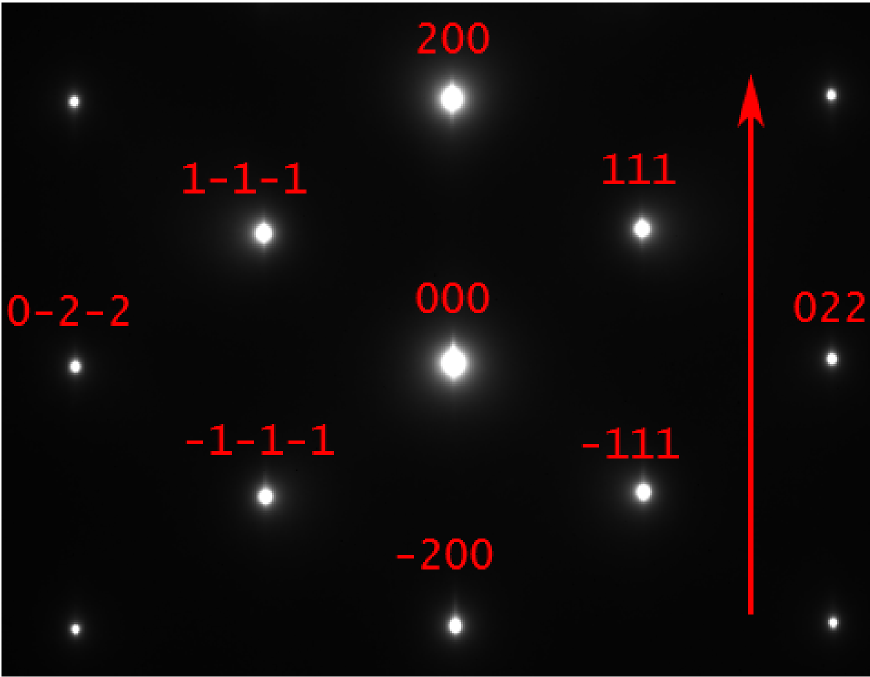


Figure 4.3: Diffraction pattern from sample A, showing the $[0\bar{1}1]$ zone axis, with some indexed spots. The streaking of each spot along the 100 direction are due to the superlattice created be the repeating pattern of QD and spacer layer. The red arrow indicates the growth direction.

nm.

The thickness of the buffer layer (arrow 3) and the n-emitter (arrow 2) were found to be 104.3 nm and 105.1 nm, respectively. These layers should both be 100 nm thick, and the measured thickness's are therefore close to the expected values.

The total number of QD layers were found to be 50, and the total thickness of the spacer layers were measured to be 1081.6 nm . Thus the thickness of each spacer layer were found to be 22.1 nm.

The thickness of the top layers were measured from image 4.2b, and found to be 29.9 nm for the p-emitter (red arrow in the image) and 50.3 nm for the contact (blue arrow in the image).

The diffraction pattern for both samples are exactly alike, as both are grown epitaxially on the same type of substrate. An indexed DP at the zone axis $[0\bar{1}1]$ is given in figure 4.3. The zone axis was identified by the method described in section 2.4.3 and the DP was indexed by

the procedure described in the same section. This diffraction pattern is taken from the same area as figure 4.1a, confirming that the growth directions is in the 100 direction.

In figure 4.3 there is streaking of the diffraction peaks along the <100> direction. This is an effect caused by long-period superlattices and since the direction of the reflections are along the growth direction, the repeating structure must be the InAs wetting and GaAs spacer layer. This would only happen if the repeating structure were regular, and thus the assumption that each spacer layer is more or less the same is well founded. By taking an DP at higher camera length, thus enlarging the DP, the streaks can be resolved, and it is shown that these appear as spots in figure 4.4a and 4.4b. Using the spacing between these spots, the thickness of the repeating structure can be found. As the thickness of the wetting layer is very small, the thickness found will correspond to the thickness of the spacer layer.

To calibrate the diffraction image, giving a measure of how large

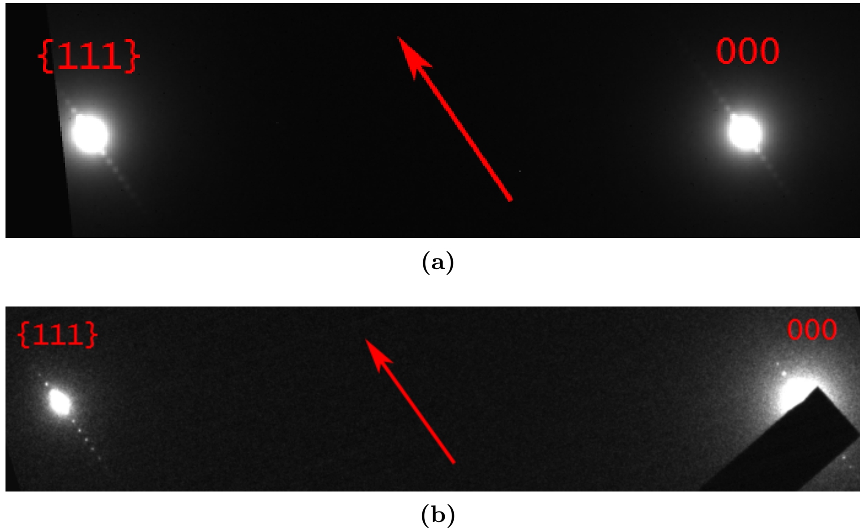


Figure 4.4: DP from both samples, taken with a high camera length. The images shows reflections from the superlattice formed by the repeating structure of InAs wetting and GaAs spacer layers. **(a)** Sample A, taken on Jeol 2100 where it is not possible to damage the camera by over saturation. **(b)** Sample B, taken on Jeol 2100F, where the camera is much more sensitive, and a beam stopper must be inserted to reduce the intensity from the central spot. In both images the red arrow indicates the growth direction.

each pixel is (measured in Å^{-1}), a DP containing a known peak must be found for the same camera length. By using the distance between the atomic planes resulting in this chosen peak, the distance spanned by each pixel can be found.

As the magnification in a TEM image is usually subject to a ±10 % uncertainty [7, Volume 2], there will be some uncertainty in the measurements from a TEM image. This is the case for all the measured values so far. The spacer layer thickness calculated from the superlattice reflections will be a more accurate measure of the thickness, as it is calibrated by the lattice spacing of the material.

The thickness of the spacer layer can be found by the following equation.

$$d_{spacer} = \frac{1}{|\mathbf{g}\text{spacer}[\text{Å}^{-1}]|} = \frac{|\mathbf{g}_{hkl}[\text{px}]|d_{hkl}[\text{Å}]}{|\mathbf{g}\text{spacer}[\text{px}]|} \tag{4.1}$$

where $|\mathbf{g}_{hkl}[px]|$ is the distance measured to some hkl spot in the DP, given in pixels, and $\mathbf{g}_{spacer}[px]|$ is the distance between two consecutive spots from the superlattice, measured in pixels.

The thickness of the spacer layer was calculated for both samples by this method, using a {111} peak, giving a thickness of 20.7 nm for

Table 4.1: The measured values for sample A and B. The layer with - thickness here indicates that this layer was not present in the sample.

	Layer	Thickness [nm]
Sample A	Buffer	104.3
	N-emitter	107.7
	Spacer (BF image measurement)	22.1
	Spacer (Superlattice measurement)	20.7
	P-emitter	29.8
	Contact	48.4
Sample B	Buffer	104.3
	N-emitter	105.1
	Spacer (BF image measurement)	22.1
	Spacer (Superlattice measurement)	21.1
	Absorption	-
	P-emitter	29.9
	Contact	50.3

sample A, and a thickness of 21.1 nm for sample B. These values fit with both the previously calculated and the expected values.

There are some uncertainty related to these measurements, as the diffraction spots are quite big, making it hard to find the center. In order to reduce this uncertainty, the distance $|\mathbf{g}_{111}[px]|$ were instead measured from the first reflections from the superlattice of each spot. These are smaller, making it easier to find the center, and the distance between them is the same as the distance between the 111 and 000 reflection.

All the measured thicknesses for both samples are summarized in table

4.1.1 Defect structures and QD stacking

Two HAADF-STEM images of the structure taken with a higher magnification is given in figure 4.5a and 4.5b for sample A and in

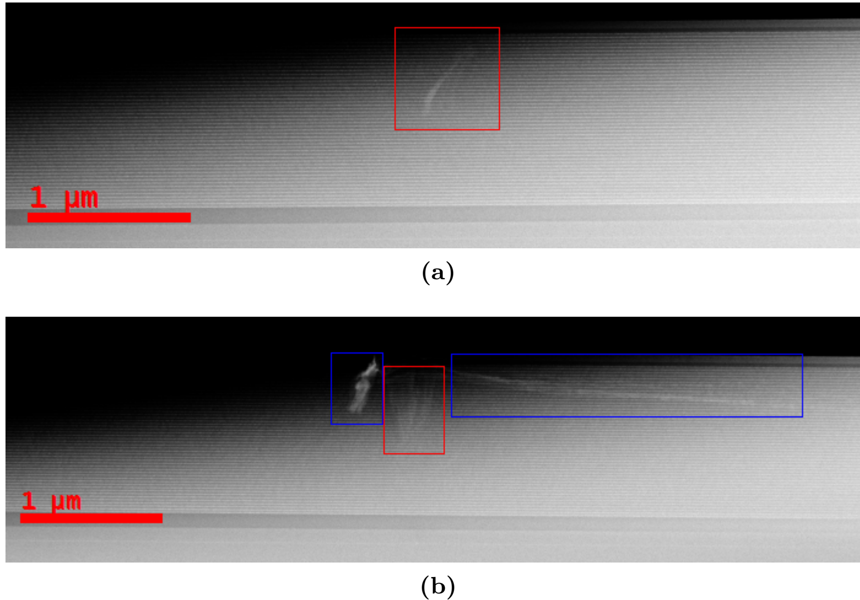
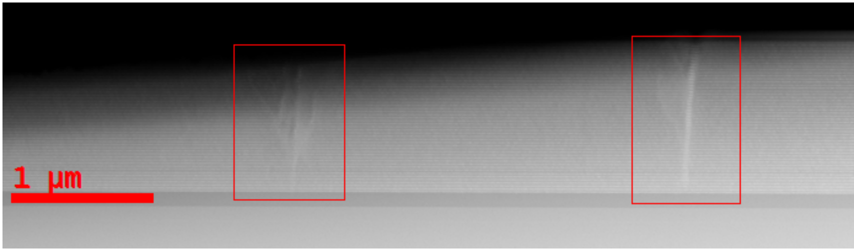
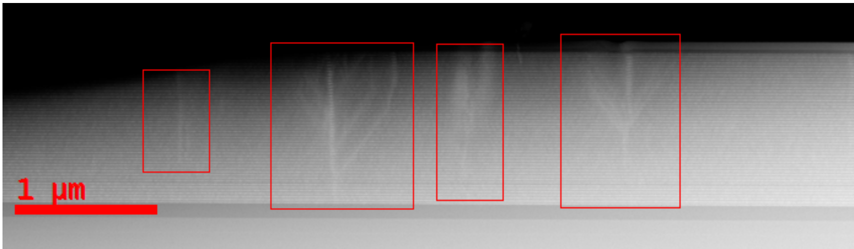


Figure 4.5: HAADF-STEM images from sample A, taken at low magnification. The images in **(a)** and **(b)** are taken from two different sides. **(a)** The feature highlighted in the red square is a stacking of QDs, which are very strained **(b)** the features highlighted by the blue squares are contaminations introduced during sample preparation, most likely glue, and the red square shows a cluster of defects



(a)



(b)

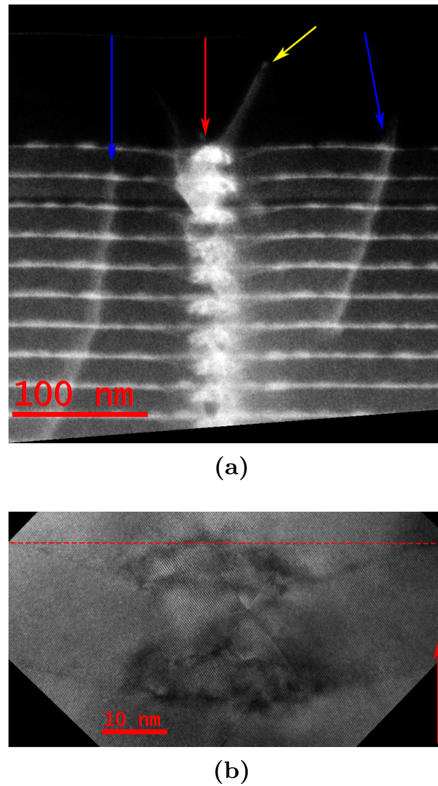
Figure 4.6: HAADF-STEM images from sample A, taken at low magnification. The images in **(a)** and **(b)** are taken from two different sides. The highlighted red squares in both **(a)** and **(b)** show various defects structures.

figure 4.6a and 4.6b for sample B. These provide an overview of the material, giving a better sense of how these materials look.

The defect structures present in the images must contain In, as they are clearly distinguishable from the background of GaAs. The contrast present in the images is most likely a combination of Z and strain contrast, where the Z-contrast is dominant. The reason the glue in figure 4.5b is so bright is because the extra amount of material, causes more of the beam to scatter, i.e. thickness contrast.

From the two sets of images, it is clear that the sample that contains N, which should have the better crystal quality, is the one with the most defects. This is contrary to what is expected, as the N should decrease the strain in the sample, and thereby reduce the number of defects.

In order to determine if the samples had somehow gotten mixed during sample preparation or before, the samples were studied by EDS to confirm the presence of N. The results from this is presented in section 4.2. The N content was the only difference between the two samples, as the 100 nm absorption layer, which should have been



(a)

(b)

Figure 4.7: (a) LAADF image of a defect structure in sample B. The large defect, indicated by the red arrow, is highly strained due to the large QDs. The contrast in the dislocations, indicated by the blue arrows, on either side of the large defect is most likely from strain contrast. It can also be seen that the strain due to the large defect causes dislocations to extend into the top layers (indicated by the yellow arrow). **(b)** HRTEM image of the top QDs in a defect, showing how the spacer layers bend as consequence of the strain introduced by the large QDs. The red dotted line indicate the position of the wetting layer without this bending effect, and the red arrow indicate the growth direction

present in the sample with N, were not existent in the structure. The reason that the samples weren't suspected to be from the same wafer, were the dark spacer layers present in the images from sample B (figure 4.6a and 4.6b), which aren't in the images from sample A (figure 4.5a and 4.5b). These layers are discussed more in detail in section 4.2.

A magnified defect is shown in figure 4.7a. The image shows one large line defect going parallel with the growth direction (indicated by

the red arrow), and a number of dislocations going along some plane (indicated by the blue arrows). This plane is found by measuring the angle it makes with the <100> direction of the material. This was found to be about 55 °, making this a {111} plane. From the image it can also be seen that due to the large build up of strain from the large defect, dislocations will extend into the top layers of the material (indicated by the yellow arrow).

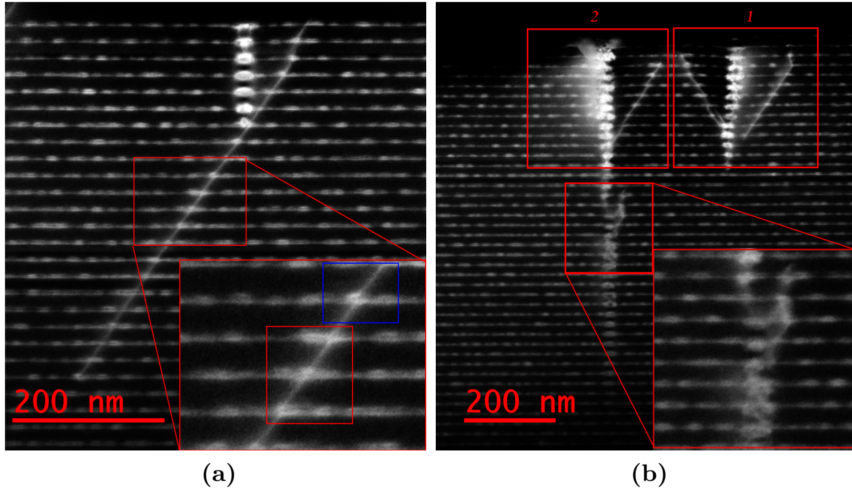
It can also be seen from this image that the layers start bending downwards at the top of the defect. This is a result of the increasing strain in the spacer, causing the lattice here to stretch in a direction perpendicular to the growth direction. This straining reduces the lattice parameter in the growth direction, causing the unit cell of the spacer to compress. The compression of the unit cell will depend on the strain around the QD, giving a varying degree of compression around the QD, causing the bending seen in figure 4.7b for the top QDs. This was also reported by Ng et. al. for similar samples [51].

When a large QD is formed, the strain introduced by it will propagate to the next layer. This causes the lattice parameter here to be slightly closer to the lattice parameter of InAs. As a consequence of this, InAs will "prefer" to diffuse to these areas and segregate here, compared to non-strained areas, as this lowers the strain energy.

The defects going parallel to the growth direction are formed by this mechanism. The defects start with one large QD, which is large enough that the "strain field" is felt by the deposited InAs, causing a QD to form directly above the first. In these defects, the first QD is large enough that the strain field causes the second QD to grow bigger, and thereby increasing the strain. This causes more and more In to diffuse into the column, during growth, increasing the QD sizes, which again increases the strain making it more likely that the defect will keep propagating until the top layers.

The small dislocations on both sides of the LAADF image (indicated by the red arrow in figure 4.7a) work by a similar mechanism. A start QD is large enough to introduce dislocations along a {111} plane. This dislocation will then terminate on the next layer, where during the deposition of InAs, a QD forms. The QD is likely to segregates here, as the dislocation line will propagate some of the strain and, by the same reasoning as previously, it will be energetically favourable for the InAs to segregate into a QD here.

This QD can now either create dislocations in the same plane,



(a) (b)

Figure 4.8: LAADF-STEM images taken from sample B. **(a)** Shows a dislocation line along the 111 plane and at some point it spawns a defect along (100) direction. The magnified part shows how the dislocations terminate and start at different QDs. **(b)** Shows two line defects which splits of into dislocation/defect lines along 111 planes. The magnified area shows a point in the defect splits in two.

causing the defect to continue along this plane (figure 4.7a), or if the QD is large enough, the defect can start propagating in the growth direction (rectangle number 1 in figure 4.8b). Another option is a combination of the two, which can be seen in figure 4.8a.

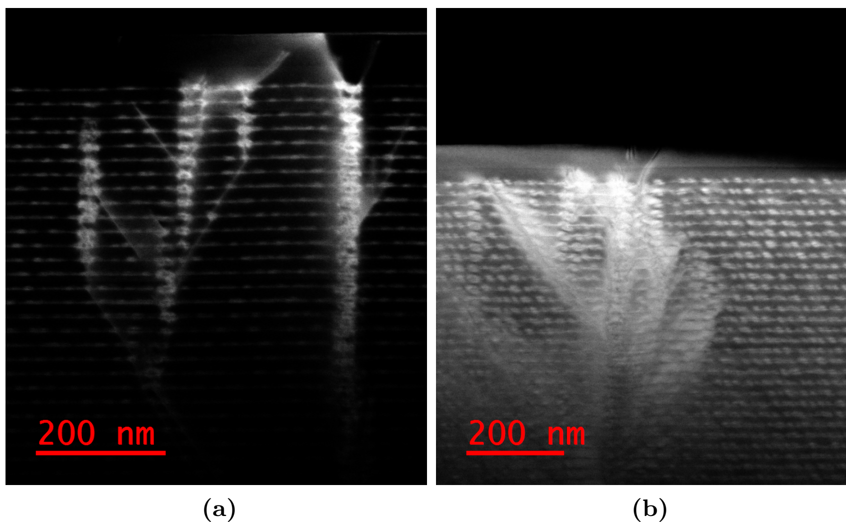
The dislocation along the {111} plane will terminate on a QD, and if it continues on, it can do this in two ways. This is shown in the magnified part of 4.8a, where the red square shows the first case and the blue the second. The first case is when the dislocation continues along the same direction after terminating on a QD, in this case the dislocation line seems to pass right through the QD. In the second case the dislocation terminates on a QD, and a new dislocation going along the same plane, but shifted from the original dislocation, is created.

The dislocation created at a QD can also go along planes that are perpendicular to the dislocation that is terminated on that QD. This means that these dislocations along {111} planes can go along all the 4 main 111 directions, i.e. (111), $(\bar{1}11)$, $(1\bar{1}1)$ and $(11\bar{1})$. Only two of these will be present in the sample geometry studied in this sample, as the other two will be in directions parallel to the zone axis. When

the sample is tilted to the [011] zone axis, or an equivalent direction, these two types of dislocations will just look like a short stack of QDs. This is because the spacing of between each QD along a {111} plane is $\frac{20\ \mathrm{nm}}{cos(54°)} = 34$ nm, and the sample thickness is on the order of $\backsim 100$ nm. Though as the TEM image is a projection, it will seem like the distance between the QDs in the stack is 20 nm, when imaged at the [011] zone. It might be possible to identify these stacks by tilting the sample.

These dislocations along the {111} planes, reduces the strain of the material. This can be seen in the LAADF image in figure 4.8b. This image was taken at quite a high camera length and the strain contrast is therefore very prominent.

Looking at the dislocation on the left in rectangle number two in this image, it can be seen that the diffuse intensity on the left side increases as the defect progresses towards the top. This effect is not seen on the right side of this rectangle, and the reason for that is the strain reduction by the creation of a dislocation along a {111} plane. This dislocation originates at one of the QDs in the defect. The same



(a) **(b)**

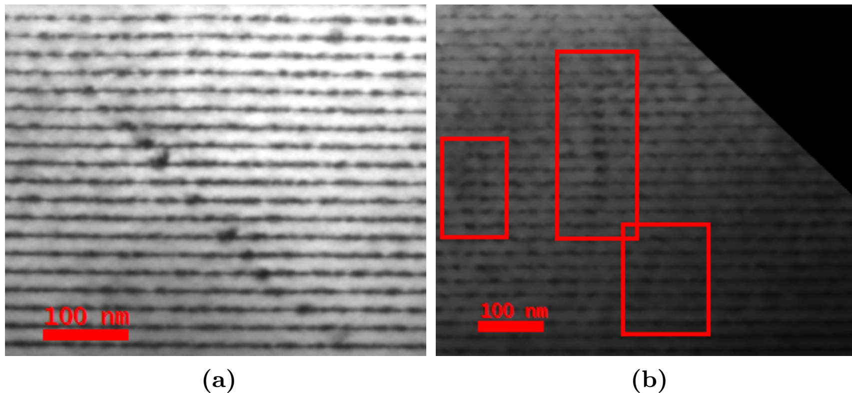
Figure 4.9: Two LAADF-STEM images of examples of the branching caused by the combination of defects going along the growth direction and the ones going along 111 planes. Both images are from sample B. **(a)** is taken at a relatively thin area

effect can be seen in the defect on the right (rectangle number 1). The left side of this defect is dark, which indicates that this area is less strained compared to the right side. The left side has a defect (maybe two) originating from QDs in the defect, which reduces the strain. The right side (in rectangle 1) also has a dislocation going along a {111} plane, but this doesn't originate at a QD in the defect and so the strain isn't reduced.

The defects in both samples are similar in structure and are usually a combination of the line defects and the dislocations along {111} planes. This causes some defects to branch out as they propagate through the layers. This can be seen in figure 4.8b, where, after some point defects along the 111 planes are created. The same figure also shows a magnified part of the image of where the defect splits into two. Only one of these two defects continue to propagate towards the top. This is most likely caused by the diffusion of InAs into the left defect, and thus there is not enough InAs left for the QDs in the right one to become large enough for the defect to continue to grow.

Two larger examples of this branching is shown in figure 4.9a and 4.9b.

If a QD is not large enough to start a defect, but still large enough that the strain field is "felt" in the next layer, a stack of QDs can be created. These stacks usually end before the top of the material. This is because the strain from the first QD, doesn't increase the diffusion



(a) (b)

Figure 4.10: BF-TEM images taken from the two samples. **(a)** Angled stacking of QDS in sample A **(b)** Three different QD stackings, indicated by the red rectangles, along the growth direction in sample B.

Figure 4.11: A schematic of the attempt to measure the defect density in each sample. The numbers indicate the 4 positions were the thickness was measured, and the red squares is the area being imaged.

of InAs in any significant way, resulting in a similarly sized QD on top of the first. Eventually the stacking tapers off, leaving behind a stack of QDs spanning a few layers. These types of structures are stacked in the same directions as the defects, i.e. along 111 planes or along the growth direction. An example of both types of stacking are given in figure 4.10a and 4.10b.

A more detailed analysis on the dislocations in QDs will be given in section 4.4.

When comparing the number of stacks between the two samples, it becomes clear that sample B contains more of these structures than sample A. Again this is contrary to what is expected for the sample containing N.

From these images a clear indication that the crystal structure of sample A is better than sample B, is found, but as the thickness of the two samples might differ, the results aren't directly comparable.

In order to remedy this, attempts were made at determining the defect density of both samples, but this proved to be difficult.

The attempt at measuring the density was done by taking a low magnification image of the sample, on an area containing all 50 layers. EELS were then used to measure the thickness in the corners of the camera. Using these thicknesses and the length of the imaged area, the volume were found. In order to get an accurate density, the idea was that the areas imaged were continuous along each side, meaning no overlap or gaps in the areas imaged. Figure 4.11 shows a schematic of how this process would have been done. This would have been done for several sides of each sample, giving a total volume imaged. The defect density could then be found by counting the number of defects and dividing this by the total volume.

This proved to be impossible to accomplish, as the thickness mea-

surement by EELS were not valid for thick areas. When the measurements were done in the bottom corners, from left to right in figure 4.11, the thickness's found didn't change with the evident increase in thickness from the image. This would have caused the measurements found to be lower than the exact value, but not by the same amount. As a consequence of this it would not have been possible to compare between samples by this method and the uncertainty in the values found would be high.

In order to be within the limit of the EELS measurement, only one area could be imaged for each side, as this was the only area thin enough. This would have resulted in a very low number of measurements, and the density found would have been very uncertain and not necessarily representative of the sample (it might have been by some stroke of luck, but this is highly unlikely).

The defect density was instead determined qualitatively by comparing sides which were more or less equal in thickness. As the thickness of the bottom layers were still quite thick, a measurement of two points, parallel to the growth direction were taken at intervals of 1μm on each side. By comparing the thickness at these two points, a rough comparison of the defect density in each area could be given. The two chosen points were from the last layer, i.e. the thinnest area on the sample, and from layer number 30, as the thickness here was below the limit of EELS.

For sample A, the areas measured are from the same side as the areas shown in figure 4.5a and 4.5b, but only figure 4.5b was thin enough for the EELS measurement. Over a length of 5μm, only one small defect structure was found in this area. This is the same one as the one in the red square from this image. The area that was too thick for EELS measurements, contained no defects, only a small angled stack of QDs, indicated by the red square in figure 4.5a.

For sample B, the areas measured were from the same side as 4.6a (area 1) and figure 4.12 (area 2). Area 1 contained two large defect structures, while area 2 contained two large line defects. In both cases the defects are indicated by the red squares. The central red square in figure 4.12 is a stacking of QDs, and it appears darker than the surrounding matrix because strain in the spacer layer causes less electrons to be scattered to higher angles by the channelling effect. As the channelling effect is dependant on the thickness, this effect will only be seen in highly strained areas which are also thin.

Table 4.2: The measured thicknesses for one area of sample A, and two areas of sample B. The thicknesses are measured from left to right in each area, where the orientation of these areas are the same as in figure 4.12. The measurement at the 50 layer point and the 30 layer point, were done such that a line between these two points was parallel to the growth direction. The N/A value is from an area of the sample too thick for thickness measurement

	Area	Layer #	Thickness [nm]				
Sample A	1	50	9	4	210	252	297
		30	370	258	221	260	326
Sample B	1	50	8	16	370	149	110
		30	298	322	295	346	N/A
	2	50	318	334	337	306	351
		30	362	369	381	401	428

The measured thicknesses for both samples are shown in table 4.2. The thicknesses going left to right in the table, are from areas in the samples, oriented as in figure 4.12, going left to right. For each thickness column, there are two measurements. One from the top layer and one from layer number 30. These measurements were done such that the line formed by these two points are parallel to the growth direction.

In sample A, it can be seen that the area has a high thickness gradient, near the left side of the imaged area, while towards the right it is much flatter. It can also be seen that the thickness isn't strictly increasing from left to right, but varies more unevenly. This is caused

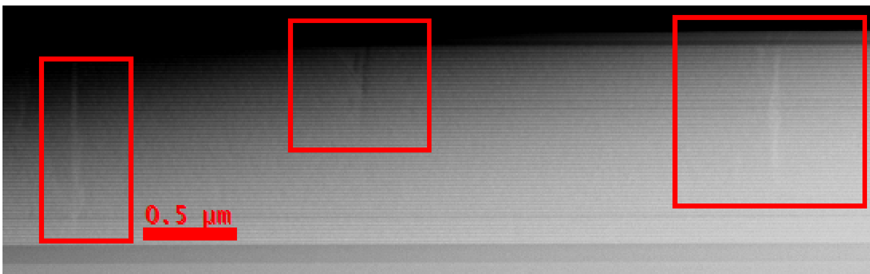


Figure 4.12: HAADF overview image from sample B, showing three different defects. The middle defect, is most likely a highly strained dislocation, causing a strong channeling effect.

by a rough surface, which exhibits "valleys" and "peaks" introduced by the sample preparation. The same things can be said for area 1 of sample B. Area 2 of sample B on the other hand exhibits a much flatter surface, but the second effect is still seen here.

The thicknesses of area 1 of sample A and area 1 of sample B, are very similar, and so these areas can be used to compare the defect density between samples. Area 1 of sample A only contain 1 small defect structure, whereas area 1 of sample B contain two very large defect structures. The defect density of area 1 in sample A will therefore be smaller than the density in area 1 of sample B. Area 1 is also comparable to area 2 of sample B, as this area is much flatter over all the 50 layers, while area 1 of sample A has a higher thickness gradient, making the thickness near the bottom of the two samples more or less the same. The number of dislocations starting in the bottom of both samples are therefore comparable, and this number is higher for sample B. Another thing that can be noted, is that area 2 of sample A, which were to thick for EELS measurement, contained no defects.

These results demonstrate that the defect density of sample A is lower than in sample B, but there are large uncertainties due to the lack of statistics. Nevertheless, it needs to be confirmed which of the two samples that contains N, as there is clearly some difference in the crystal quality of the two samples.

Another technique that could have been used to determine the thickness of the thicker areas of the sample, CBED. This wasn't done is there was not a sufficient amount of time available.

4.2 The hunt for the missing Nitrogen

In order to identify if the samples had been switched during sample preparation or before, EDS were performed. As the signal from Carbon is easily detected, it should be possible to detect N content in these samples, given that the content of N is above the detection limit, as the characteristic X-rays of N have higher energy than those of C.

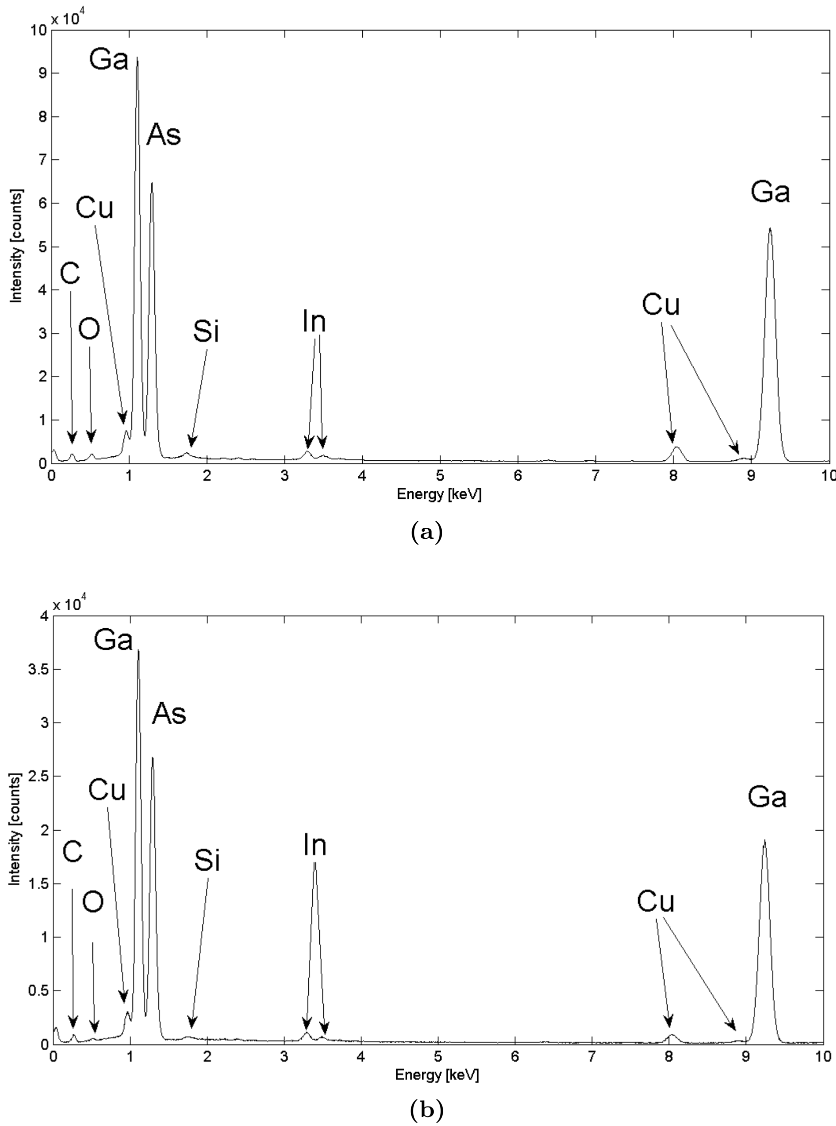


Figure 4.13: EDS results from **(a)** sample A and **(b)** sample B. In both samples the most prominent peaks are labelled by their corresponding atomic element. The energy range was set to 10 keV in order to increase the resolution. The difference in counts between the two spectra is only a factor of 2, which could be caused by a number of factors, such as thickness, sample tilt, or beam intensity. The difference here is not significant.

The EDS analysis were performed on the JEOL 2100, with the

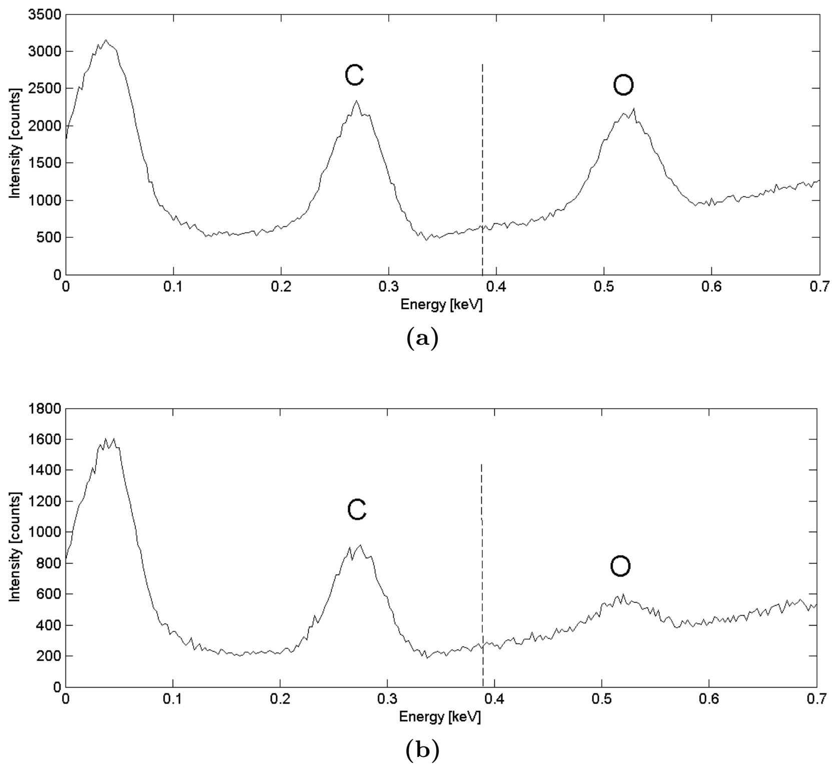


Figure 4.14: A magnified portion of the two previous spectra. **(a)** Sample A and **(b)** Sample B. The dotted line is approximately where the peak from N should be and in both images it can be clearly seen that this is non-existent

beam covering all 50 layers. The results from these measurements are shown in figure 4.13a and 4.13b, for sample A and B respectively.

From these spectra, it is seen that both samples contain the same elements. Gallium from the spacer layer, As from both the QDs and the spacer, and In from the spacer. The C peak is either from contaminations of hydrocarbons which gathers on the surface on the sample or from the glue used during the sample preparation. The O peak is due to oxides that form on the surface of the sample. The Si and Cu peaks originate from the method these samples were prepared. The Si signal is from the Si slices which were used to fill space within the ring, and the Cu signal is from the brass ring, which is an alloy of Cu and Zn. The Cu signal can also stem from materials present in the column of the microscope.

As the details of the area where the N peak should be isn't quite clear from these images, a magnified portion of these spectra are presented in figure 4.14a and 4.14b, for sample A and B respectively. From these two images it can be seen clearly that there is no peak at the position associated with N. This is indicated by the dotted lines in the images, and should correspond to an energy of 0.392 keV. The C and O peak in both cases are quite low and there is clearly no overlap from these two peaks that could cause the N signal to be hidden within the "tail" of one of them. Therefore the N content must be below the detection limit of the EDS detector on this instrument.

The samples were then tested with EPMA, which has a lower detection limit (about 200 ppm), compared to the EDS detector on the TEM. This technique has a spatial resolution of about 1-2μm, but this doesn't matter as the N should be evenly distributed through all 50 layers of the sample. The probe depth of this technique is about 1μm, which fits well with the sample structure in this thesis. The EPMA results are shown in table 4.3.

The table shows the atomic content in mass percent, so the sum of all the elements won't add up to 100%. The detection limit of 200 ppm, is equal to 0.02 % in mass percent. From the table it can be seen that both samples have measurements of N content which is above this limit. In sample A, this was the only measurement which found any

Table 4.3: The results from the EPMA study for both samples. The values are given in mass percent.

Measurement #		Ga	As	In	N	Al	Si
Sample A	1	51.53	58.88	0.66	0	1.73	0.040
	2	51.84	57.78	0.53	0	1.73	0.018
	3	51.83	58.42	0.63	0	1.73	0.041
	4	51.35	58.04	0.51	0	1.72	0.036
	5	51.73	58.45	0.50	0.026	1.70	0.020
	6	51.67	58.43	0.55	0	1.70	0,0040
Sample B	1	51.50	58.97	0.72	0.0090	1.73	0
	2	52.18	59.10	0.64	0	1.69	0.0090
	3	52.38	59.07	0.61	0.010	1.69	0.0020
	4	52.10	58.97	0.67	0.026	1,.68	0
	5	51.84	59.28	0.69	0.010	1.69	0
	6	51.94	59.22	0.46	0.0080	1.68	0.0010

N. In sample B the rest of the values were below the detection limit.

These results gives no clear indication of which sample contains N and because both samples have only one measurement above 200 ppm, there is large uncertainty in the N content. The only thing which can be concluded from these results is that both samples most likely have a nominal N content below 200 ppm or even zero.

Using the EPMA results, the atomic composition of both samples can be quantified. In table 4.4 the average atomic percent of each element in the sample is given, here the N content has been assumed to be insignificant. This table also shows the standard error associated with each measurement.

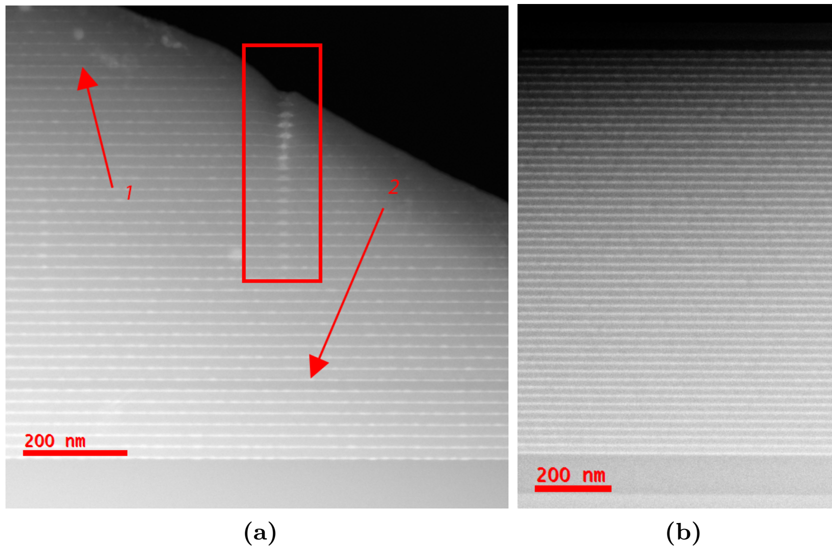
The Si content is highly uncertain, as from table 4.3 it can be seen that only two measurements give a value above the detection limit. The Si content is expected to be low, as it is only used as a dopant in these samples. The average atomic percent of Si in sample A is therefore most likely much lower, and probably below the detection limit.

In both samples it can be seen that the As content is higher than the Ga content, this is also what is expected as As is present in all the layers in these samples, while Ga is present in every layer expect the wetting and QD layers. The Ga content is also lower because the Al in the emitters replaces some of the Ga atoms.

The Indium content in the 50 layers should be about 1.4%, which is calculated by looking at the ratio of the total number of In atoms and the total number of atoms in the unit cells for a sample thickness equal to the probe depth. The discrepancy from the measured value is assumed to be a combination of two things. First, since the probe only goes 1μm into the sample, only around 46 InAs layers will contribute to the signal, as the top of the sample contains two layers with a thickness

Table 4.4: The atomic composition of both samples, expressed in atomic percent. The standard error for the measurement of each elements is also given.

	Ga	As	In	Al	Si
Sample A	46.61	48.99	0.31	4.01	0.060
Sample B	46.50	49.19	0.34	3.91	0.00
Standard error	0.030	0.040	0.010	≈ 0	≈ 0



(a) (b)

Figure 4.15: (a) A HAADF image taken on the JEOL-ARM200F from sample B, showing the first dark spacer layer. This is indicated by arrow number 2. In the red rectangle a defect is highlighted. The first arrow shows some contaminations introduced by the focused probe used in STEM. **(b)** A HAADF image taken on the Jeol-2100F showing all 50 layers from sample A. There is no dark spacer layers present here.

of about 80 nm. The decrease from this shouldn't be significant.

The second thing is the penetration depth of the probe, as the amount of electrons penetrating the sample decreases with increasing depths, there will be higher scattering intensity from atoms near the top. This effect can also be seen in the Al content from the EPMA results. As Al only makes up about 25% of the 30 nm thick p-emitter, the resulting Al content should be around 0.75%, by the same calculation as for In. The Al content found by EPMA is much higher, around 4%. This indicates that atoms near the surface of the sample, contribute more to the measured atomic content, and so the quantified atomic content cannot be trusted.

As previously noted are the presence of the a couple of darker spacer layers in figure 4.6a and 4.6b. These are not present in sample A, and so represent a very clear structural difference between the two samples.

In figure a 4.15a a HAADF image is shown, where the first of these

dark spacer layers can be seen. Also provided is a HAADF image showing all 50 layers of sample A (4.15b), which show no such layers. Clearly there is a structural difference between the two samples.

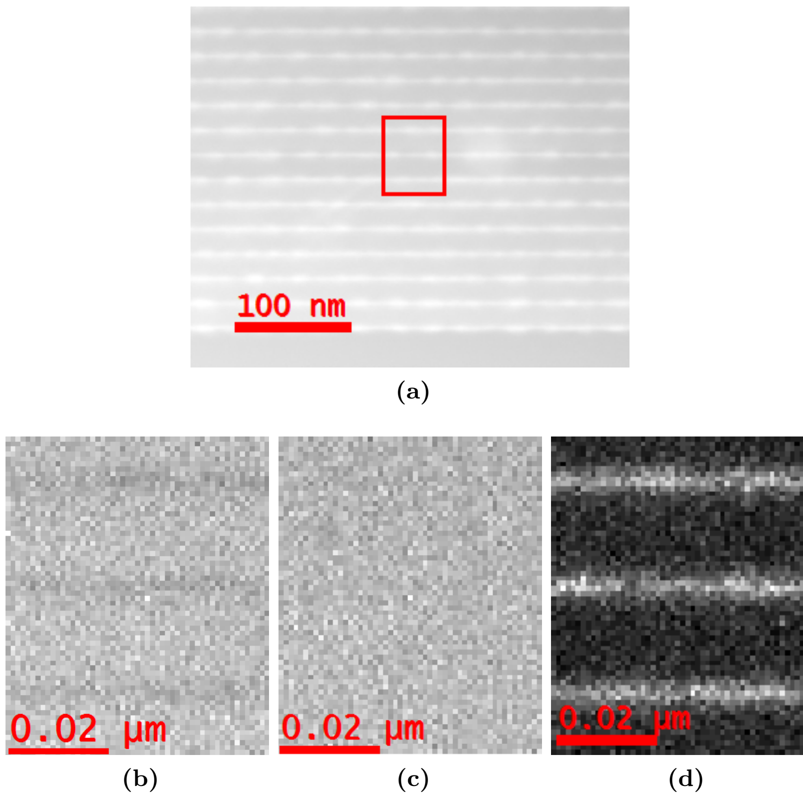
As figure 4.15a is a HAADF image, taken with a very low camera length, the Z-contrast will dominate, but some thickness and strain contrast might also be present. The strain contrast can be ruled out as the cause of these dark layers, as the contrast would have changed with thickness. As these layers are dark, if strain contrast is responsible, it must be because of the channelling effect. This effect diminishes as the thickness increases due to more material in the beam path, increasing scattering to higher angles.

This leaves either Z-contrast or thickness contrast. It is not possible that this is caused by a sudden change in thickness, as the interface is too sharp. This leaves Z-contrast, which is not as dependant on the thickness of the sample, and this is why there isn't a contrast difference for this layer in the thicker areas of the sample. In a HAADF image, the Z-contrast will cause lighter elements to appear darker, as they scatter less to higher angles. The presence of these dark layers is therefore a good indication that sample B, contain a few layers which consists of a lower Z-composite material. This could be the missing N.

In order to find out if these layers contained N, EDS-STEM were performed on the Jeol-ARM200F. The EDS-map were performed on the area indicated in figure 4.16a, and the corresponding elemental maps for Ga, As, and In are shown respectively in 4.16b, 4.16c, and 4.16d.

The Ga mapping shows less concentration of Ga atoms in the wetting layers and the QDs, as is expected. There will still be some intensity left in the QD and the wetting layer. In the QD this is because the map is a projection of the sample, which might be $20-30$ nm, while the QD is only around 10 nm, meaning there will still be Ga behind or in front of the QD. In the wetting layer, which penetrates through the sample, this might be caused by an unevenness in the layer, or by the intermixing of In and Ga atoms.

The As map shows a very even concentration in the entire map, and this is reasonable as the As content is the same in both the spacer and the QD/wetting layer. If the N content in the dark layer had been high, it might have been possible to see this in the In map. As N replaces As in the spacer and thus this layer would have appeared darker in the As map.



(a)

(b) (c) (d)

Figure 4.16: **(a)** HAADF image of sample B taken on the Jeol-ARM200F, showing the region which were mapped with EDS, indicated by the red rectangle. **(b)**, **(c)**, and **(d)** shows the maps of Ga, As, and In, respectively

The concentration of In atoms can be seen from the In map. Here it can be seen that the layers seem slightly diffuse, which is caused by the intermixing of In and Ga atoms. The scattered intensity in the spacer layer is most likely just noise, as the In concentrations here will be either zero or below the detection limit.

The spectra from the EDS map is shown in figure 4.17. Here the whole spectra from the mapped area is shown, combined with the magnified part surrounding the area where the N peak is. Besides the elements whose presence have already been explained, there can also be seen two extra peak from Fe and Co. The origin of these are from materials within the instrument.

From the magnified portion of this image, it can be clearly seen that there is nothing where the peak corresponding to the characteristic

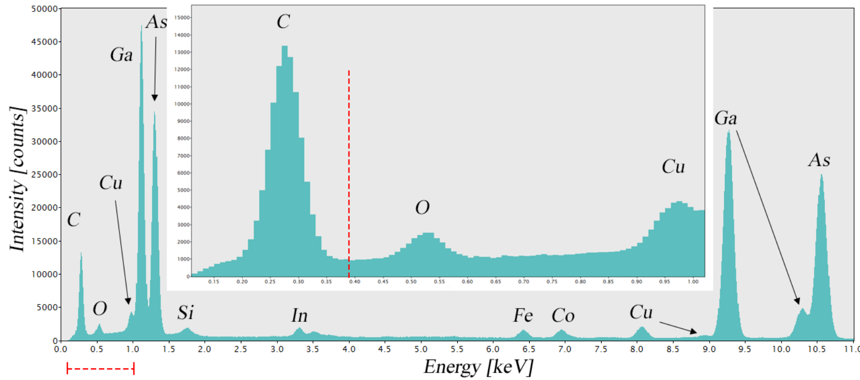


Figure 4.17: An EDS spectra from the area indicated in figure 4.16a. The inset spectra is a magnified part of the the region indicated by the interval in red in the lower left corner. The dotted red line in the magnified spectra indicates where the peak from N should have been

X-rays from N should have been. The C and O peaks are also quite low, so there is no overlap of either these peaks with the signal from N.

These results gives no indication of N content, and coupled with the EPMA studies, the N content must be below 200 ppm. Still, there is the presence of dark spacer layers in sample B, which must be caused by a lower Z-composite, as all other contrast mechanisms have been ruled out.

N can be ruled out as the cause of these dark layers, due to a different ionic diameter of N compared to Ga and As, which would cause scattering into the HAADF regime, actually increasing the intensity in the image [52]. The dark spacer layers are therefore most likely not N, but some other low Z element. The likely candidate here is Al, but this cannot be shown from the EDS data, as the tail of the first As peak, most likely covers the signal from Al. This could probably be seen in EELS, but there was not sufficient time to check this.

4.3 Quantum dot size and density

In order to find the QD sizes, two different techniques were used. First dark field imaging and then HR-STEM. The basic idea behind the first technique was a different structure factor for InAs and GaAs. The

selection rules for GaAs are

$$F_{hkl}^{GaAs} = \begin{cases} 4(f_{Ga} + f_{As}), & h+k+l = 4n \\ 4(f_{Ga} - f_{As}), & h+k+l = 2(n+1) \\ 4(f_{Ga} \pm i f_{As}), & h+k+l = \text{ odd} \\ 0, & h,k,l \text{ mixed.} \end{cases} \tag{4.2}$$

where f_{Ga} is atomic form factor for Ga, f_{As} is atomic form factor for, and n is an integer.

The structure factor for InAs will be the same, only the atomic form factor for Ga has been replaced by the atomic form factor for In. The atomic form factor for Ga and As will be very similar, as these two elements have similar atomic numbers, resulting in weak reflections for h, k, l values whose sum is equal to 2(n+1). This is because the intensity in a hkl peak is proportional to the structure factor for this peak. With InAs however, the difference will be greater, as the difference in atomic number is greater here, and so these reflections will be a lot higher in intensity.

By tilting the sample away from the zone axis, around the <100> direction, to a 2-beam condition with the 000 and 200 reflection, the dynamic effects present in the image will be reduced. As a consequence of this, most of the intensity in the 200 peak will be from InAs. This results in an image with good contrast between the InAs QDs and the GaAs spacer. The Al in the emitter layers, will also make these layers very bright compared to the surrounding GaAs matrix. This is because AlGaAs has the same structure as both GaAs and InAs, and the difference in atomic form factor is large compared to GaAs. This effect can be seen in figure 4.18a.

Some of the results from the attempts at finding the QD sizes DF images with the 200 reflection are shown in figure 4.18a, 4.18b, and 4.18c. In the first image the QDs are very faint, and they are surrounded by a dark layer, which is most likely caused by strain. The QD sizes can therefore not accurately be determined from this image.

In the second image the sample hasn't been tilted far enough away from the zone axis, causing the image to still retain some dynamical effects, which makes it hard to see the outlines of the QDs. In the top right corner some QDs can be barely seen, but these atleast look nicer. Again the problem with these is that due to strain, the surrounding area of the QD will appear dark, thus the size on the image won't correspond to the real size.

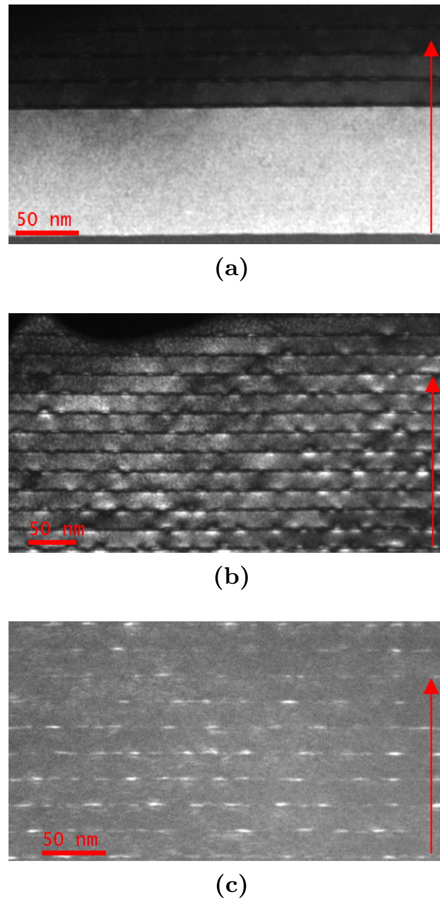


Figure 4.18: Three different dark field images of the 200 reflection, for different tilt angles. The red arrows in all the images indicate the growth direction. **(a)** Bottom emitter of AlGaAs, which is very bright in this DF reflection. The QDs on the other hand is very faint, and is almost not possible to see. **(b)** QD layers close to the edge of the sample. At this tilt angle, there is still a lot of dynamical effects, and so the outline of the QDs are hard to see. **(c)** Here the sample has been tilted very far away from the zone axis. This results in very few dynamical effects, but also a very low intensity, requiring long exposure times which causes the QDs to "smear" out. The size of the QDs are therefore not possible to determine accurately from this image.

Finally in the last image, the sample has been tilted very far from the zone axis. There is now very little dynamical effects, and there is little strain contrast. The problem with this image, is that due to the

large tilt angle the intensity in the image will be very low. Therefore, in order to image this area, a very long exposure time is needed. As there is drift in the sample. the QDs will "smear" outwards, and thus the size on the image won't be representative of the actual QD size.

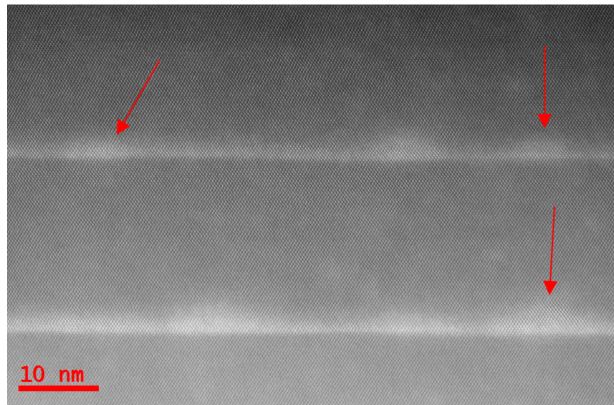
So the problem with this technique is that at low tilt angles, where the intensity is high enough to keep the exposure times low, the image is dominated by dynamical effects and strain. This causes the outlines of the QDs to either be very blurred or smaller than the actual size, as the QDs are strained more on the edges, decreasing the intensity here. At higher tilt angles, these effects disappear, but here the intensity is too low, increasing the exposure time needed to record an image. This causes the QDs to smear due to sample drift, making them appear larger than the actual size. This technique is therefore not good enough for finding the QD sizes and won't be used.

The other technique used is to image the sample in STEM, with the HAADF detector mode. As the contrast here is dependant on the atomic number, and In is a great deal larger than both Ga and As, it should be much brighter than the surrounding matrix. This was done for both samples on the Jeol 2100F. A typical image from which the QD sizes can be determined are shown in figure 4.19a.

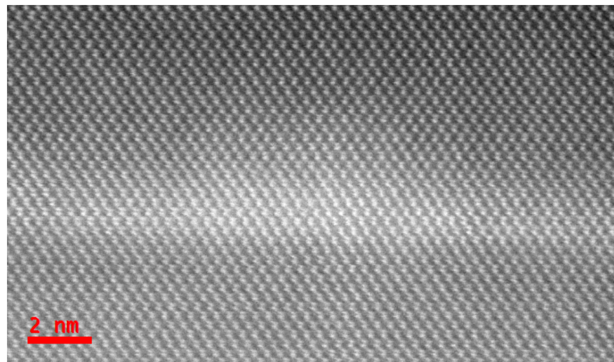
The QDs in these samples were lens shaped, i.e. described by a height (in the growth direction) and a diameter (parallel to the layers). This can be seen in figure 4.19b. In this image it can be seen that the QD is quite diffuse, and since the image is in focus (this can be seen from the fact that there is atomic resolution in the image), this must come from the diffusion of In atoms into the spacer and Ga atoms into the QD. The effect is negligible below the QD, and this is reasonable as the atoms here have had time to relax into a stable configuration before the QD layer were grown. This is not the case for when the spacer layer is deposited above the QD. As the QD is very strained, the atoms will not be in a particularly stable configuration, so during deposition of the spacer layer, there will be an intermixing of In and Ga atoms.

As these QDs have no well defined boundary, the measured sizes will have a slight uncertainty, as its hard to know exactly where the QD ends and the spacer starts. This is the same for both samples, so it won't affect the comparison in any significant way.

To determine if there were any difference in sizes between the layers, the measured QD sizes were placed into one of three categories,



(a)

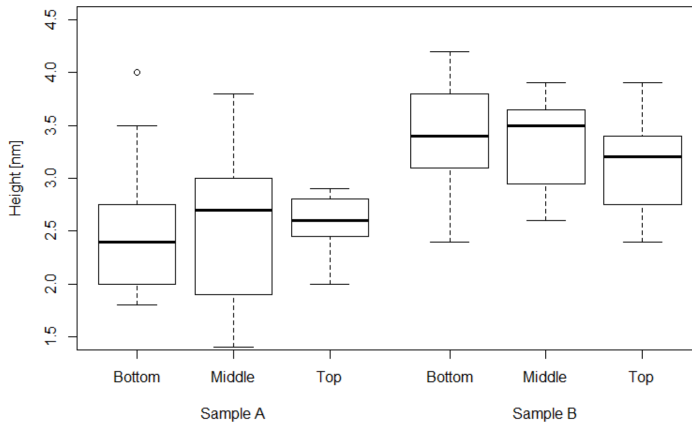


(b)

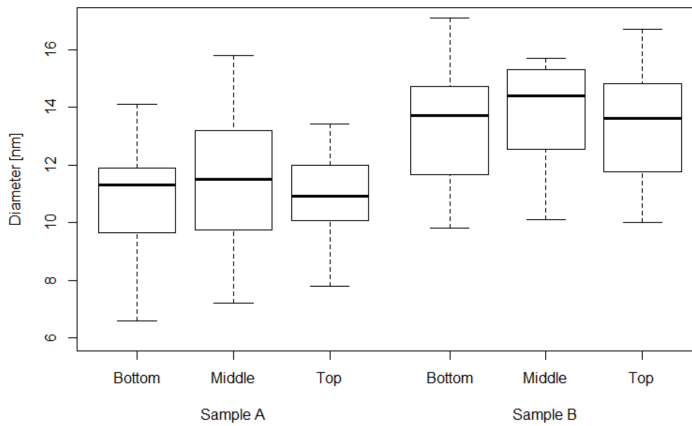
Figure 4.19: (a)A HAADF image taken on the Jeol-ARM200F showing QDs two layers of QDs. The QDs which are sharp enough for determining sizes are indicated by the red arrows. **(b)** A HAADF HR-STEM image of a single QD taken on the JEOL-ARM200F, showing the general shape of the QDs

depending on which layer the QD was in. These categories were simply bottom, middle, and top, where the bottom category consisted of measurements from layers 1 to 15, middle consisted of layers 16 to 30, and the top consisted of layers 31 to 50. In each category a total number of 15 QDs were measured.

The box plots in figure 4.20a and 4.20b, describe the variation in the measured QD sizes in all three categories for both samples. The thick dark line in these boxes represent the median. The first line above this is the upper quartile, 25% of the data lies above this line.



(a)



(b)

Figure 4.20: Box plots showing the variation in **(a)** height and **(b)** diameter of the QDs for both samples. These data show little variation within the layers of each sample, but a difference in both height and diameter between the samples can be seen. The lines in each of these boxes indicate, from top to bottom, the maximum value, upper quartile, median, lower quartile, and minimum value, respectively

The next line is the maximum value, excluding outliers. One of these outliers can be seen in the bottom category for sample A in figure 4.20a. These outliers represents data points that are more that 3/2 as large as the upper quantile. The first line below the median is the lower quartile, and 75% of all the data lies above this point. Further down is the minimum data value, excluding outliers. There aren't any

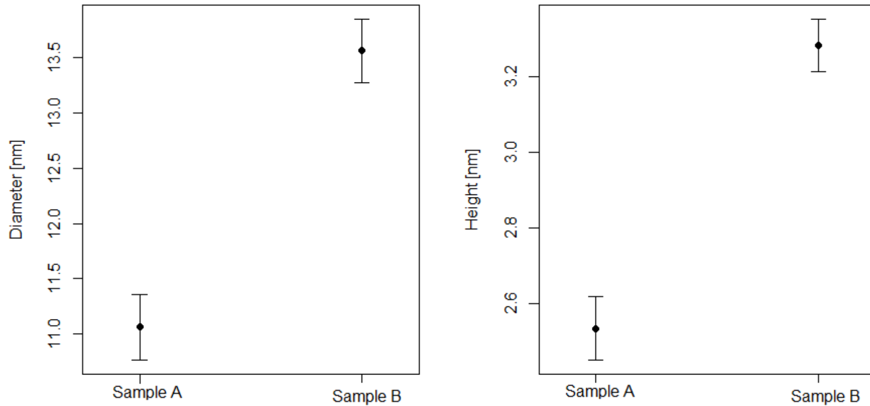


Figure 4.21: A plot of the average value of **(a)** diameter and **(b)** height of the QDs in both samples. Here the average value of all layers in each sample is given, as the variation within each sample was insignificant. The dark dot in these plots represent the average value, and the range given by the lines is determined by the standard error.

lower outliers in these images, but the lower outliers are defined as data points that are less than 3/2 of the lower quartile. These types of plots are a good way of showing the distribution of data within the measurements.

From these two figures, it can be seen that there is little variation between the height and diameter of the QDs within each sample. There is however a difference between the samples. This difference is even more evident in a lineplot, which shows the average values and a range determined by the standard error. As the difference between layers were insignificant, the plots in figure 4.21 are composed of all the values within each sample, making no distinction between the layers.

From this plot it is clear that there is a distinct difference in both diameter and height of the two samples, as there is no overlap of the ranges defined by the standard error. The average values and standard error for each sample for both height and diameter are given in table 4.5.

The sizes of the QDs will affect the crystal structure, as larger QDs are more likely to create dislocations, and propagate more strain. Previously, it was seen that sample A has a lower defect density than sample B, and this fits well with the measured QD sizes, as the average QD in sample A is smaller than in sample B. This is most likely caused

by the different growth temperatures during deposition of the InAs, as the amount of deposited InAs was the same in both samples, leaving this as the only likely cause. The dark spacer layers in sample B has no, or little effect, on the QD sizes, as the variation within the layers were non-significant. If they did affect the QD size, the values in the middle layers would differ from the other two layers, as only the bottom and top categories contain dark spacer layers.

Since the deposited In were the same in both samples, and since the QDs in sample A is on average smaller than in sample B, it is likely that sample A will have a higher QD density compared to sample B. It could also happen that the wetting layer is thicker in sample B, compared to sample A, but this wasn't measured in this thesis.

To check this, the thickness of the sample must be found. At first it was considered to use CBED to find this thickness, but since the areas where there is minimal chance of overlapping QDs are in thin regions of the sample, this technique is not applicable. EELS were therefore used to measure the thickness.

To measure the thickness for each QD, the following procedure were used. First the sample was tilted away from the zone axis, where the QDs were clearly visible. This tilt angle was kept low in order to not increase the measured thickness to much compared to the actual value (which is the value measured at zone). An image where then taken, and by focusing the probe, EELS measurements were done for all the corners of the camera. The thicknesses measured were then assumed to be varying linearly from bottom to top right corner, and the same for the left side of the camera. This can be seen in figure 4.22, where the scale on each side of the image is in factors of the mean free path, here each value is given at an interval equal to the spacer thickness.

Table 4.5: The average diameter and height of the QDs for both samples, including the standard error of the measurements. These values show a clear difference between the diameter and height in the two samples.

		Average value [nm]	Standard error [nm]
Sample A	Diameter	11.06	0.30
	Height	2.53	0.08
Sample B	Diameter	13.56	0.29
	Height	3.28	0.07

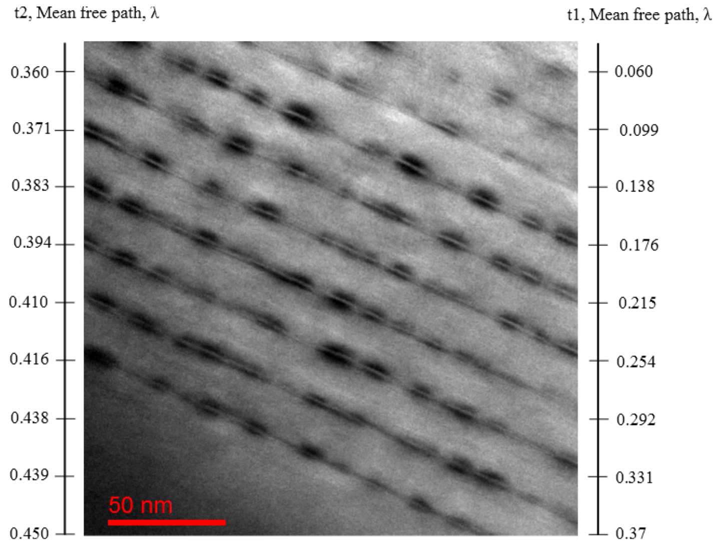


Figure 4.22: A tilted BF image of the bottom layers from sample B showing how the thicknesses in each layer were found. The measurement points are in the corners of the image, and by assuming a linear variation parallel to the horizontal direction, the scale is found. In each layer the number of QDs are counted, with the end and start thicknesses taken from the scale. The mean free path in this case were about 100 nm.

Another assumption made, was that the thickness would vary linearly within each layer. By this method the area spanned by each QD layer could be calculated by the following formula

$$A = \frac{1}{2}L(t_1 + t_2) \tag{4.3}$$

here L is the length of the layer as measured on the image, and t_1 and t_2 are the thicknesses at the end and start point of the layer, i.e at the image edge.

Since the area image, was extremely thin ($\approx 10 - 20$ nm), and the QDs had a diameter of around $11 - 13$nm, it could be the case that a counted QD was actually only half a QD. This is shown schematically in figure 4.23. In order to account for these effects, the thicknesses used in equation 4.3 were added one QD diameter, where the added value was the average found previously. This reduced the chance of overestimating the density.

Also in these measurements the layers were divided into three categories, the same as for the the case of QD sizes. In each category

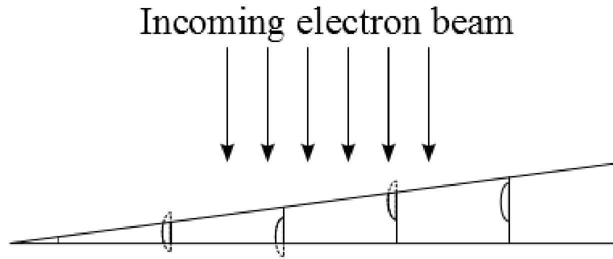


Figure 4.23: A schematic of the effect that could cause a too high QD density at thin areas. As the QD diameter is comparable to the thickness of the samples, a half dot might be counted as one dot, increasing the QD density beyond the actual value. The dotted lines of the QDs indicate that these portions of the QDs have been removed during sample preparation

Table 4.6: Average density for both samples, including the standard error in the measurements.

	Average density [10^{10} cm^{-2}]	Standard error [10^{10} cm^{-2}]
Sample A	23.8	0.5
Sample B	9.7	0.3

a total of 12 layers were counted. A comparison of the density in these categories for each sample is given in figure 4.24. From this it can be seen that the variation within the layers of each sample is not significantly different, but that the density between the two samples are. The top category of sample A, contains a number of outliers, which could be caused due to experimental errors or chance. Experimental errors in this case, would most likely be counting to many or to few QDs. This could be caused by small unstrained QDs, which would be hard to see from the BF image, or it could be caused by two QDs overlapping, and being counted as one.

The average density for all layers in both samples is plotted in figure 4.25, with the standard error. These values are summarized in table 4.6 for both samples. From this table it can be seen that there is about a a factor 2 difference between the samples, with sample A having the highest density. This fits well with the observed QD sizes in sample A and B, as the QDs in sample A are on average smaller than the QDs in sample B.

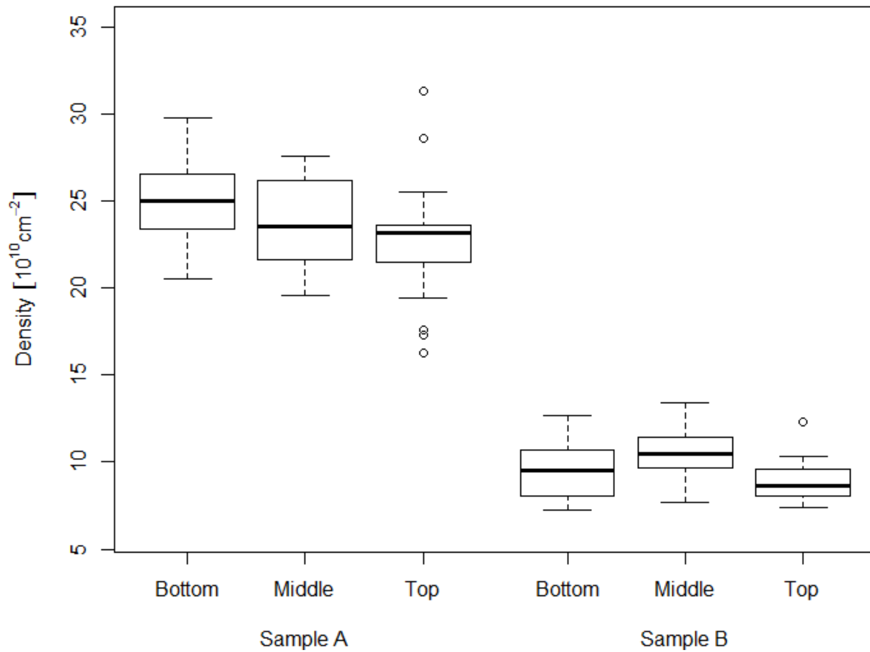


Figure 4.24: A box plot of the QD densities within the layers of each sample. The Top category of sample A has a number of outliers, which could be caused by experimental errors. One of these outliers is also seen in the top category for sample B. This plot shows that there is little variation within the layers of each sample, but that there is a difference between the samples. The lines in each of these boxes indicate, from top to bottom, the maximum value, upper quartile, median, lower quartile, and minimum value, respectively

4.4 HRTEM imaging of QDs

This section provides a more in depth look into what happens to the crystal structure for QDs in defects or in stacks and QDs which are not part of any strained structure. These dislocation and defect mechanisms are very complicated, so the analysis presented here will be very superficial and in places very simplified.

A look into what happens to the crystal structure for QDs in a defect, as the defect progresses, is shown in figure 4.26a-4.26d. Here, figure 4.26a and 4.26b, shows HRTEM images from the bottom and top of a defect. The highlighted areas indicate where the FFT in figure 4.26c and 4.26d were taken from. These are RGB images, which are

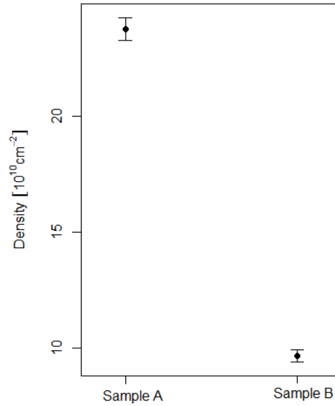


Figure 4.25: The average QD density for both samples, plotted with the standard error. This plot shows that the QD density in sample A is about a factor of 2 higher than sample B.

basically just a sum of three images, where each image is given a colour (either red, green or blue). If there is only two images, one image can be given two colors. This method can be used to differentiate between spots that overlap and those that don't, as the ones that overlap will have a color equal to the resulting sum, e.g. green and red will appear as yellow, and green, red, and blue will appear as white.

Here figure 4.26c is constructed from the FFT of area 1 (given the colors red and green), and the FFT of area 2 (given the color blue). From this image it can be seen that the lattice parameter for the two areas is relatively unchanged, as there is no distinct separation of the spots, i.e. the spots appear white. The strain in the QD causes the spots to become more "smeared" out, and this can be seen from the image, as each spot are surrounded by a diffuse cloud of blue.

The RGB image in figure 4.26d is constructed from the FFT of area 1 (red and green) and the FFT of area 3 (blue). From this image a clear separation of the spots in the FFT can be seen, with the blue spots, originating from the QD, having a shorter distance in reciprocal space, and the yellow spots originating from the unstrained spacer layer. This indicates that the lattice parameter of area 3, or the QD, is larger than the lattice parameter of the unstrained spacer layer of GaAs. This would only happen if the crystal structure here had relaxed to the lattice of InAs.

Thus, as the defect progresses, the QDs grows larger and larger,

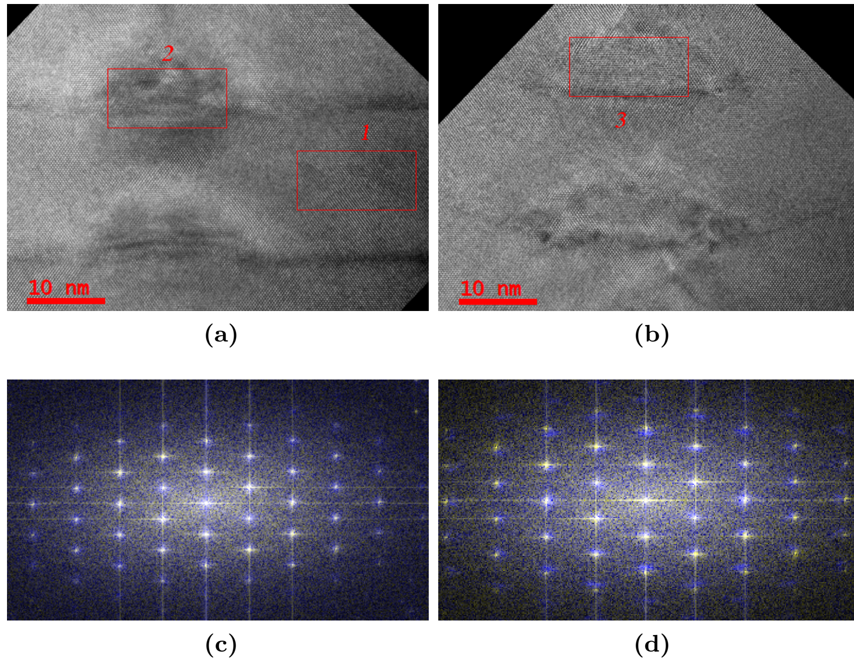
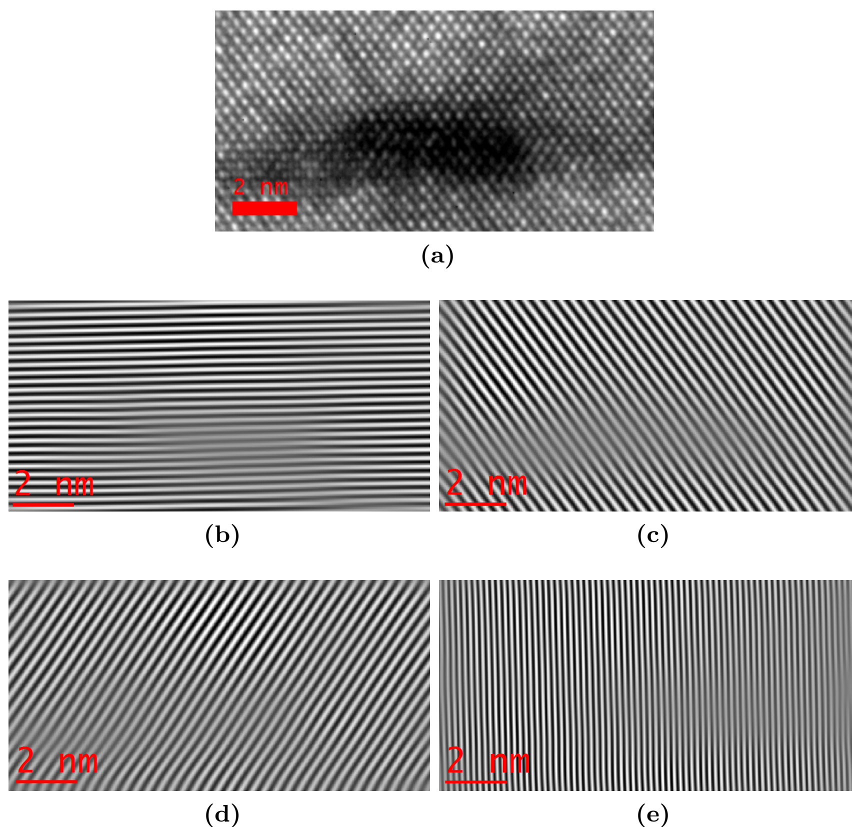


Figure 4.26: **(a)** HRTEM image from the bottom of a defect, where the red rectangles indicate the areas which have been Fourier transformed. Rectangle number 1 is positioned on a unstrained area in the spacer layer, and the second rectangle is positioned on a QD. **(b)** HRTEM image from the top of the same defect, the red rectangle is positioned on a QD, and indicates the area which has been Fourier transformed. **(c)** A RGB image constructed from the FFT of rectangle 1 (red and green) and rectangle 2 (blue). This shows that there is little difference in the lattice parameter between the QD and the spacer, but that the lattice in the QD is strained. This can be seen from the blue diffusivity around each spot. **(d)** A RGB image constructed from rectangle 1 (red and green) and rectangle 3 (blue). Here a clear difference can be seen, indicating a different lattice parameter. **(a)** and **(b)** were taken on the JEOL-ARM200F of sample B.

resulting in a relaxation of the lattice to InAs as the QDs reach a certain size. As these QDs have relaxed into the lattice of InAs, the spacer layer must compensate much more, and this is what causes the bending of the layers. This effect can be seen in the image in figure 4.26b. Note that this effect is not present in figure 4.26a, where the QDs haven't relaxed.

The results presented here aims to look at the difference in dislo-



(a)

(b) (c)

(d) (e)

Figure 4.27: **(a)** HRTEM image of a free QD. **(b) - (e)** shows the lowest index planes extracted from the FFT taken of the QD in **(a)**, showing the **(b)** (200) **(c)** (111) **(d)** $(1\bar{1}\bar{1})$ **(e)** and the (022) planes, respectively.

cations for free QDs, stacked QDs, and QDs in defect structures. Free here means that the no dislocations from other QDs terminate on the QD, and that it doesn't cause any dislocations which terminate on some QD. The difference between stacked QDs, and QDs in defect structures, are as explained in section 4.1.1.

The dislocations can be found by the method described in section 2.4.1. Imaging at a $[0\bar{1}1]$ zone axis, means that it will be possible to locate dislocations of planes which are perpendicular to this zone. The three types of planes which will be studied are the lowest order planes which are perpendicular to this zone axis. These are the (200), (111), $(1\bar{1}\bar{1})$ and (022) planes

These planes, for a free QD, are shown in figure 4.27a-4.27e.

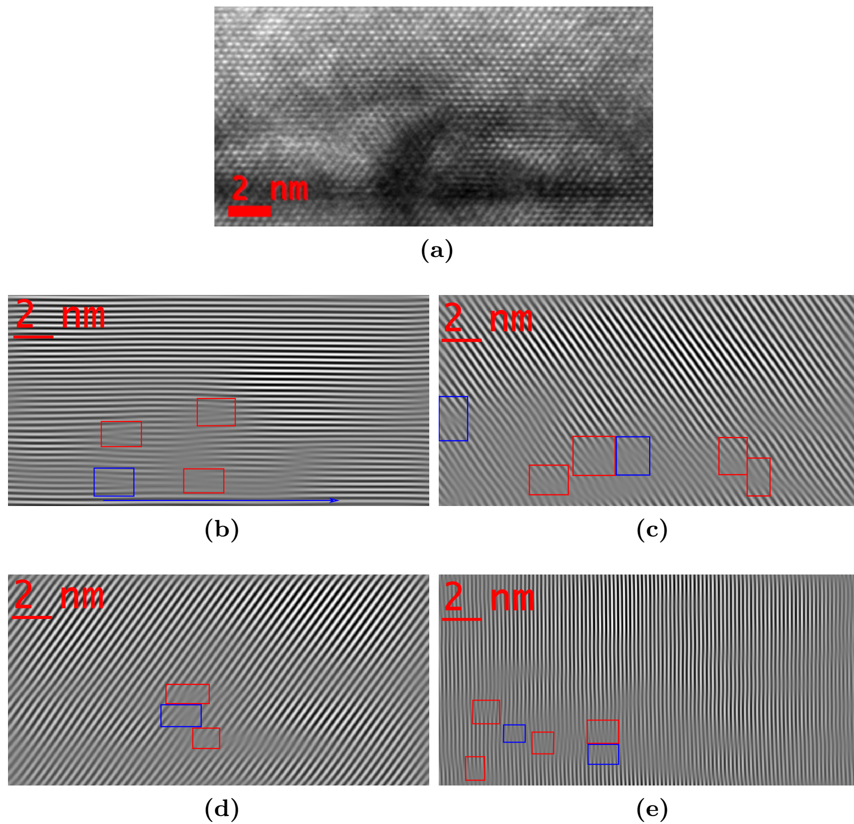


Figure 4.28: **(a)** HRTEM image of a stacked QD. **(b)** - **(e)** shows the lowest index planes extracted from the FFT taken of the QD in **(a)**, showing the **(b)** (200) **(c)** (111) **(d)** $(1\bar{1}\bar{1})$ **(e)** and the 022 planes, respectively. In image **(b)** - **(e)** the red rectangle indicate dislocations which introduces an extra plan, while the blue rectangle dislocations which are terminated. The blue arrow in **(b)** indicates a fictitious direction useful when counting the terminated and created dislocations

From these images, it can be seen that this QD doesn't contain any dislocations along these planes. This examination was done for a few free QDs. These results indicate that free QDs either contain no dislocations or in the (200) plane, no dislocations along the (111), $(1\bar{1}\bar{1})$, and (022) planes were seen.

This was also done for QDs in a stacking, and this is shown in figures 4.28a-4.28e.

From these images, a number of dislocations can be seen along all

the low index planes. The dislocations created by the QD are indicated by the red rectangles. These dislocations introduces an extra plane, which lowers the total strain energy. The area around this plane will still be strained, and this extra plane will propagate the strain until it is terminated. This QD is within one of these stacks, and so the dislocations terminated on the QD, most likely originate from the QD below. The terminated dislocations are shown in blue. Most of the terminated dislocations seems to be located on near the bottom of the QD. The dislocations terminating on the QD from the {111} planes seem to be centred more around the center of the QD, with the exception of one that is located on the upper edge of the QD on the right.

The dislocations along the (200) planes won't introduce defects perpendicular to the growth direction, as the InAs relaxes into isolated QDs which are only "connected" by the wetting layer. These dislocations most likely propagates a bit into the spacer layer, before terminating. Since the growth direction can't be used to determine if a dislocation is terminated or created in these planes, let's define the a "dislocation propagation" direction as going from left to right. This is indicated by the blue arrow in figure 4.28b. Using this, the number of dislocations that terminate and the number created can be counted. These are essentially the same, so the total strain won't be affected in any significant way by these dislocations in this case

Finally, these planes are also shown for a QD within a line defect. The chosen QD were located close to the bottom of the stack, as QDs further up will contain a lot of dislocations, as they have relaxed into the crystal structure of InAs. This relaxation process is by the creation of dislocations to reduce the strain energy.

The low index planes can be seen in figure 4.29b-4.29e for the QD in figure 4.29a, this QD is from near the bottom of a line defect. Again the red rectangles indicate created dislocations, and the blue rectangles indicate terminated dislocations.

From these images, it can be seen that the total number of dislocations in these planes doesn't differ much from the stacked QD shown previously. In the images from the (111), $(1\bar{1}\bar{1})$, and (022) planes, it can be seen that all the terminated dislocations are at the upper boundary of the QD/spacer. There is one dislocation along the (111) planes which seem to terminate at the lower edge of the QD.

In the (111) planes, it can be seen that this QD contains more

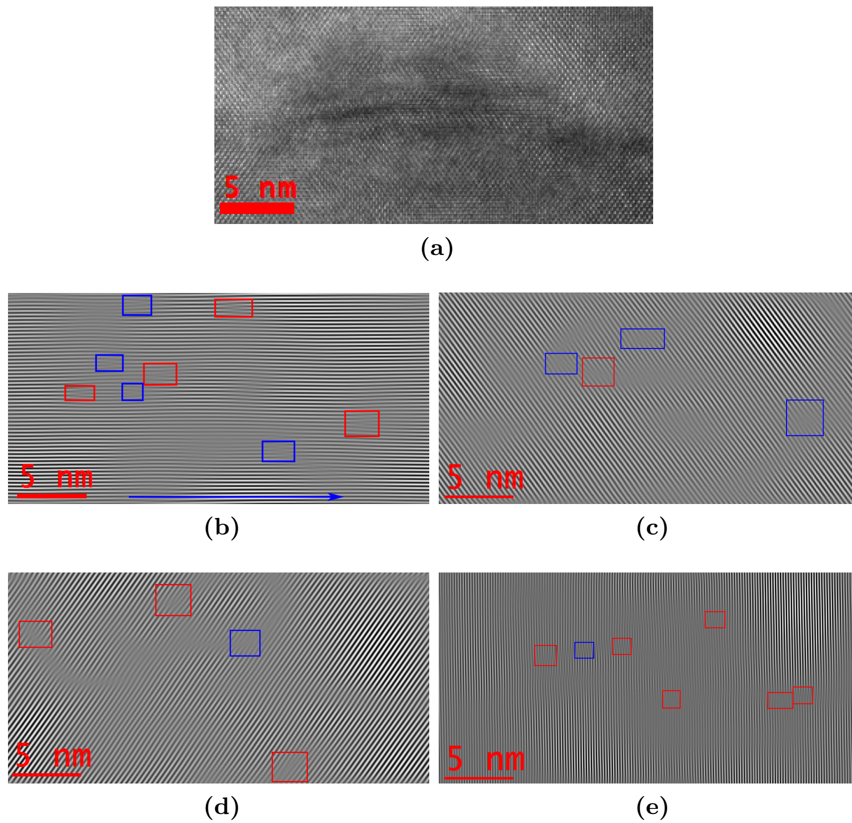


Figure 4.29: **(a)** HRTEM image of a QD at the near the bottom of a defect taken on the Jeol-ARM200F. **(b) - (e)** shows the lowest index planes extracted from the FFT taken of the QD in **(a)**, showing the **(b)** (200) **(c)** (111) **(d)** $(1\bar{1}\bar{1})$ **(e)** and the (022) planes, respectively. In image **(b) - (e)** the red rectangle indicate dislocations which introduces an extra plan, while the blue rectangle dislocations which are terminated. The blue arrow in **(b)** indicates a fictitious direction useful when counting the terminated and created dislocations

terminated dislocations, than created.

Again in the (200) planes, using the same definition as before, the number of created and terminated dislocations are more or less the same. This means that these only help decrease the strain at a local level, i.e. in the immediate area around the dislocation, and do not decrease the strain on larger scale.

The result that the amount of dislocations along the (111), $(1\bar{1}\bar{1})$, and (022) planes, are more or less the same as the stacked QD, seem

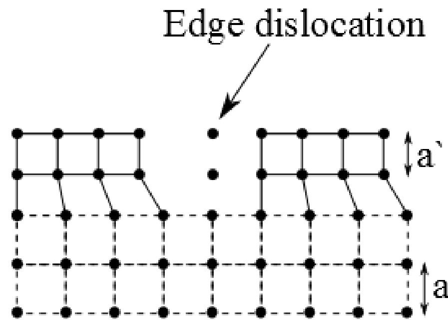


Figure 4.30: A schematic of a edge dislocation, created due to a lattice mismatch. Here a smaller lattice material, a‘, is grown on top a material with larger lattice constant, a. This straines the lattice planes, and in order to reduce the strain, these extra planes, or edge dislocations are created. The lattice mismatch is here very exaggerated

to indicate that the strain is mainly propagated elastically, i.e. not through dislocations, but by stretching the unit cell of the spacer, for these line defects. There is also a difference in the starting position of the created dislocations, and this might also explain why some QDs evolve into stacks and others into defects.

The explanation for the first cause, can be seen qualitatively by examining the lattice mismatch of the two materials. The lattice mismatch is about 7.2 % for InAs and GaAs, meaning that for every 14 unit cells of InAs there should be 15 unit cells of GaAs. An exaggerated schematic of this is shown in figure 4.30 for unit cells stacking in a <100> direction.

From this a critical length, or thickness, can be calculated for the different planes, before a dislocation is created. From this consideration, there is now a length range (or size range in 3D), which contains the same amount of dislocations. The situation will be more complex for a 3D system, especially for the lens shaped QDs, but the general idea behind it might explain why some QDs evolve into defects, while others just stack.

As the QDs in both the defect and the stack, contain more or less the same amount of defects, it can be said that both QDs have sizes within the same size range. Therefore, the reason why one QD evolve into a defect, while the other doesn’t, is most likely as a result of this QD being bigger. As the amount of dislocations are the same, the extra strain, introduced by the QD being bigger, must be absorbed

elastically within the spacer. This larger strain field then causes the QD in the next layer to grow larger than the one below, and this process is repeated upwards in the structure. The stacked QD is large enough to have the same amount of dislocations, but the size is smaller. This size is below a certain limit, which causes the QD above to remain the same size or become smaller. By this mechanism the stack of QDs eventually tapers off.

This is a very simplified explanation, and the defect formation process is most likely a combination of this effect, and strain propagated by the dislocations.

The idea of a critical size can be taken a step further. By looking at the distance between the stacking of unit cells along the different directions, an order of the critical QD sizes for the creation of dislocations along the different planes can be given. By the lattice mismatch, there should be 14 unit cells of InAs for every 15 unit cells of GaAs, but the distance between the unit cells will depend on which direction the stacking is regarded in. This means that the critical size for dislocations along the (200) planes will be the smallest, as the distance between unit cells is here a. The critical size for dislocations along the (220) planes are next, as these planes originate from a stacking of unit cells along the <110> direction, which has a distance between unit cells of $a{\cdot}\sqrt{2}$. Lastly is the critical size for dislocations along the 111 planes, which has a distance of $a{\cdot}\sqrt{3}$ between the unit cells stacking in this direction. Here a is the lattice parameter for a cubic unit cell.

This size will be even bigger for higher order planes, where the order is determined as the smallest prime number of the sum, |h|+|k|+|l|, e.g. the (111) planes are of third order. From this consideration, there should be more dislocations along the (200), compared to the (022) and (111) planes. There should also be more dislocations along the (022) planes compared to the (111) and $(1\bar{1}\bar{1})$ planes. This effect is observed in the QD in figures 4.29a - 4.29e, and a number of other QDs which were investigated by this method. For the stacked QD, figures 4.28a-4.28e, this isn't observed as the number of dislocations in the (200) direction is the smallest. This was attributed to a lower aspect ratio, which is the QD height divided by its diameter, of this QD, as a flat QD would probably have less dislocations along the (200) planes.

The strain propagating vertically is most likely a combination of dislocations and the elastically strained lattice. The dislocations

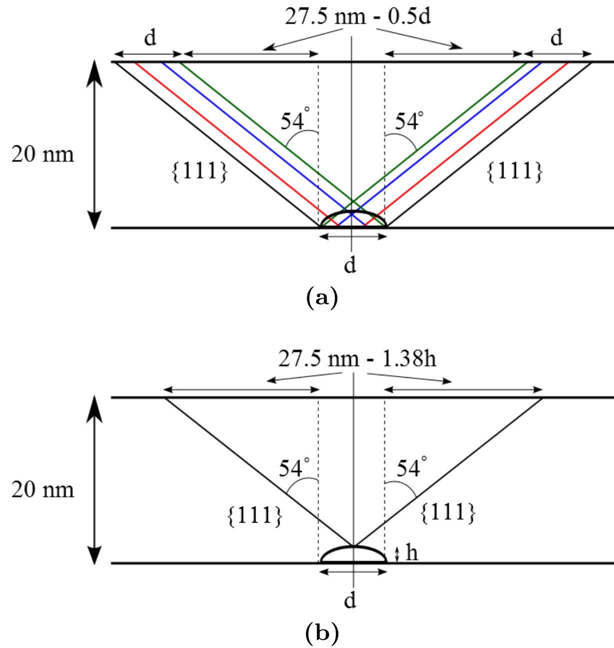


Figure 4.31: The two different types of dislocations along (111) planes. **(a)** Type 1. These dislocations start at the bottom of the QD and propagate through the QD towards the next layer. The different lines indicate that dislocations going from both left and right edges is possible for both dislocations going left to right and those going right to left. **(b)** Type 2. Here the dislocations start along the top edge of the QD, so where theses dislocations terminate on the next layer will depend on the height of the QD. Only one case is shown here, but it is possible to have dislocations on the entire outer edge.

responsible for this might be the (220) planes and certain configurations of dislocations along the {111} planes. The different cases of dislocations along {111} planes are shown in figure 4.31a and 4.31b.

The minimum limit for stacking directly on top of the first QD in a type 1 dislocation is a diameter of 27.5 nm. The configuration in this case can be seen in figure 4.32, where this configuration corresponds to case 3. This requirement is very strict, so instead define a line defect as any defect where two sequential QDs overlap. To make the calculations easier assume that the diameter of the two are the same. The minimum diameter is then found to be 13.8 nm for type 1 dislocations, here the configuration is as case 1 in figure 4.32. The midway point of these two cases are when the QDs overlap by one half

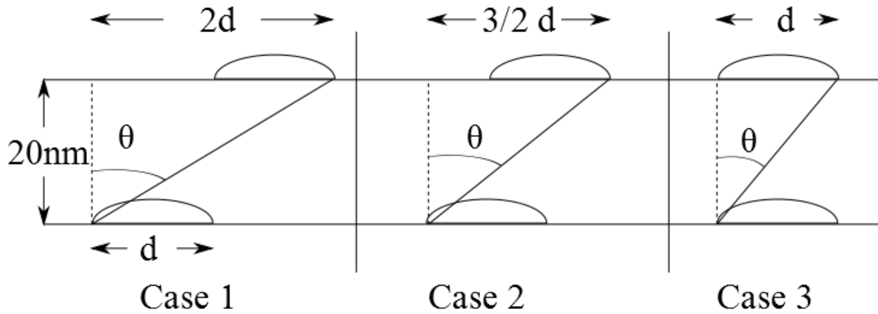


Figure 4.32: Three different QD stackings configurations, in a vertical stack. **(a)** The case when there is pratically no overlap. **(b)** QDs overlap by half a QD diameter **(c)** The QDs fully overlap. In all three cases θ is equal to 54° as these are (111) type dislocations

a QD diameter. This configuration is shown in case 2 in figure 4.32. The minimum diameter here is found to be 18.4 nm.

In the studied samples, case 3 is most likely to happen far up in a defect column, as here the QDs are very large. For smaller QDs, only case 1 and case 2 defects will be present. Case 1 dislocations are expected to happen at higher QD sizes than calculated as the dislocation isn't very likely to end at the exact edge of the QD. The average value of the diameter from sample B were 13.56 nm, and all the measured QDs, even the largest QDs measured (about 16.5 nm), were not part of any defect. This means that the calculated value, for case 1, is too low to create defects in the real case. This is most likely because the probability of having a dislocation as in case 1 is extremely low. These types of stackings most likely occur when the size of the bottom QD is bit larger, as the extra strain in the growth direction, will increase the diffusion of InAs towards the center line of the bottom QD, as the lattice parameter here will be closer to the lattice parameter of InAs. It is also probably that case 1 of these types of dislocations will instead evolve into a stacking or defect along a {111} plane for the smaller QDs.

Case 2 is the most probable configuration, as more of the elastic strain from the QD below is "felt" by the one above.

These cases should all be correlated to real cases, but this wasn't done in this thesis, due to a lack of time. Here the QD sizes of stacked and defected QDs should be measured by STEM, and from the same QDs HRTEM images should be taken. This should make it possible

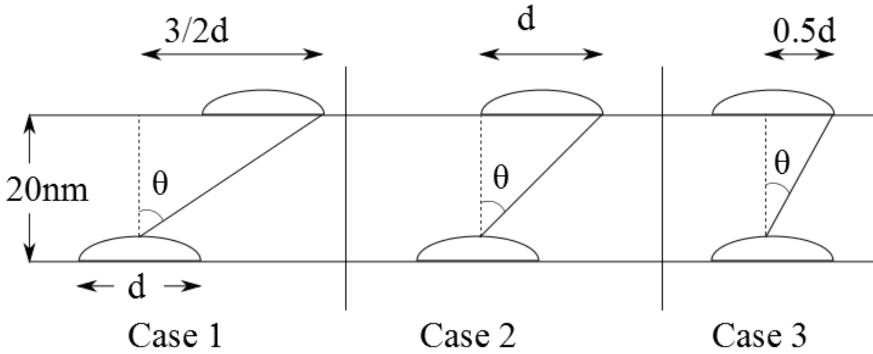


Figure 4.33: Three different cases for type 2 dislocations along the {111} planes. **(a)** The case when there is no overlap. **(b)** The case when the QDs overlap by half a diameter. **(c)**The case when the QDs fully overlap.

to correlate the type of dislocations to the size of the QD and how the QDs are stacked.

In the type two {111} dislocations, the minimum limit for vertical stacking will be dependant on the height and the diameter of the QD. By using the same assumptions as for the first type, i.e. QDs should overlap and that both QDs have the same diameter, a expression for the expected diameter can be found from the geometry of figure 4.31b.

$$d = \frac{20\text{nm}}{\frac{x}{tan(54°)} + \frac{h}{d}} \tag{4.4}$$

here h is the height, d the diameter, the fraction h/d is defined as the aspect ratio, and x is defined as either 1.5, 1, or 0.5, for case 1, case 2, and case 3, respectively.

As for type 1 dislocations, three different configurations are shown in figure 4.33.

From the measurements of the QD sizes, the average, minimum and maximum aspect ratios can be found. The diameter and height for all three configurations shown in figure 4.33 were calculated, and this is summarized in table 4.7.

From these values, it can be seen that the real values for case 1 and case 2 dislocations to occur, must be higher, as the values calculated are below the critical size for dislocations along the (111) planes to occur. This critical size for unit cell stacking along the <111> direction is approximately 14.6 nm from the simple considerations explained previously.

Table 4.7: The calculated QD sizes for the different aspect rations for type 2 dislocations along the (111) planes, for the cases shown in figure 4.33.

		Aspect ratio	Diameter [nm]	Height [nm]
Case 1	Average	0.24	8.1	1.9
	Maximum	0.35	7.8	2.7
	Minimum	0.11	8.6	0.9
Case 2	Average	0.24	11.6	2.7
	Maximum	0.35	10.9	4.1
	Minimum	0.11	12.6	1.2
Case 3	Average	0.24	20.4	4.9
	Maximum	0.35	18.3	6.4
	Minimum	0.11	23.5	2.6

The values that are left, in case 3, indicate that only the largest QDs will exhibit these types of dislocations in a vertical defect.

These calculations are extremely simplified, and the real cases will be much more complicated. The {111} type dislocations in vertical defects, will be a combination of the cases explained, depending on where in the defect the QD is, and the size. These results suggest that QDs stacking directly on top of each other (case 3), will exhibit type 1 and type 2 dislocations only if they are large enough, while the requirement is not as strict if they do not stack immediately on top of each other.

From this, small QDs stacking directly on top of each other in the start of a defect or QD must be caused by either dislocations along the (022) plane or elastically absorbed strain in the lattice of the spacer. This fits well with the previous explanation of why some QDs develop into defects, while some only stack for a couple of layers before tapering off.

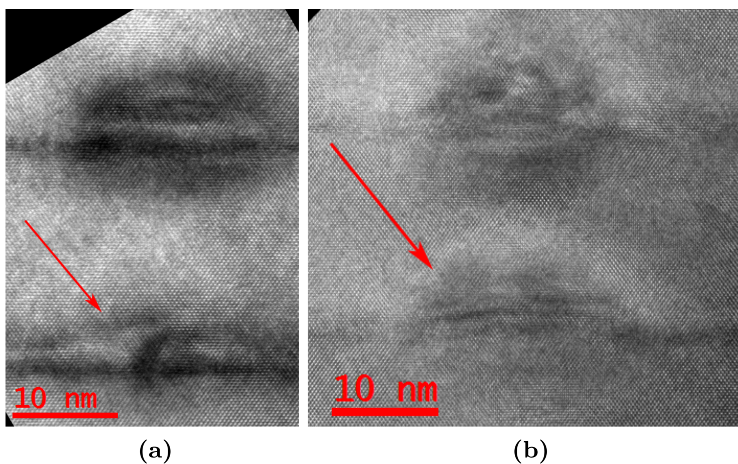
For a bit larger QDs, such as the stacked one in figure 4.28a and the one in the defect in figure 4.29a, there is a difference in which type of {111} that can be seen in each QD. For the stacked QD, the dislocations are mostly type 1, while for the QD in the defect, the dislocations are mostly type 2. The dislocations originating from these QDs can be shown to terminate on the next QD in the layer, where this QD is directly on top the first one in both cases. This can be seen in by correlating the QD in figure 4.34a and 4.34b to where the dislocations occur from figure 4.28a and 4.29a.

This could indicate that more strain is propagated by the type 2 dislocation, compared to the type 1. This could be because type 2 dislocations occur closer to the next layer, meaning that the dislocation have a shorter distance to "travel".

For the bigger QDs, then this stacking is most likely a combination of all types of dislocations and strain absorbed by the lattice elastically.

In figure 4.35a and 4.35b, a couple of both types of {111} dislocations can be seen. Here the blue lines indicate the type 1 dislocations and the red lines indicate type 2.

These dislocations are visible on the unmarked image (figure 4.35b) as well, but the two top (V-shaped) dislocations are very faint. These V shaped type 2 dislocations have been previously shown to eliminate all the strain in the originating QD [53]. The indication is that this also happens here, as this QD has relaxed into the crystal structure of InAs, which can be seen from figure 4.36. This is a RGB image, composed of an FFT from an unstrained spacer layer (given the colors red and green) and the FFT from this QD (given the color blue). As before, it can be seen that the blue dots, are are closer to the center in reciprocal space, indicating a higher lattice parameter.



(a) (b)

Figure 4.34: (a) HRTEM image from the same QD as shown in figure 4.28a, and the red arrow indicates the QD where the FFT were taken from. **(b)** HRTEM image from the Jeol-ARM200F from the same QD as shown in figure 4.29a, and the red arrow indicate the QD where the FFT were taken from.

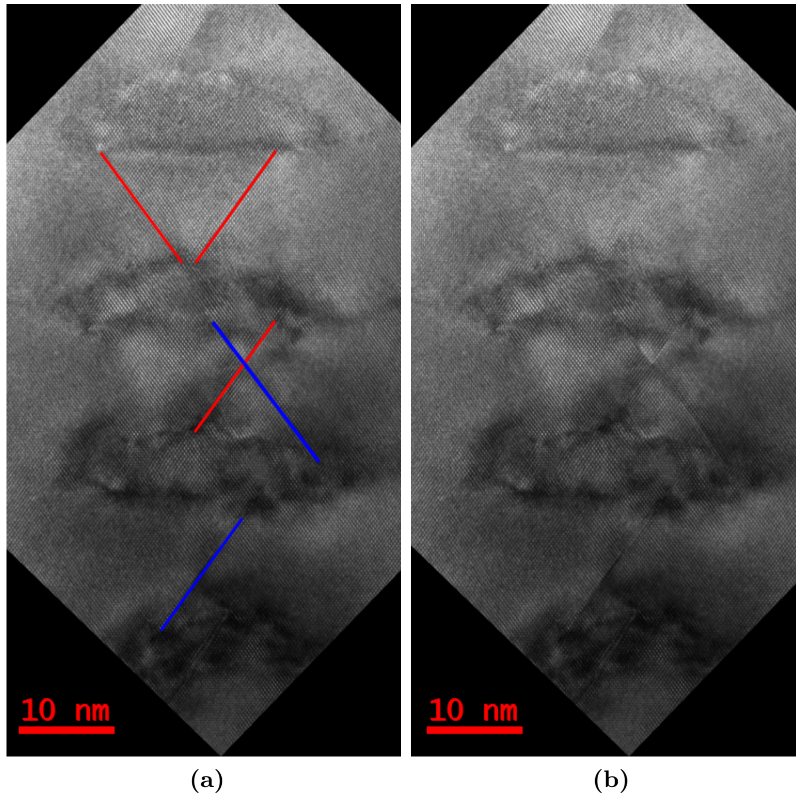


Figure 4.35: HRTEM image taken on the Jeol-ARM200F of the top QDs of a defect. In **(a)** the dislocations along the (111) planes are marked, where blue lines indicate that the dislocation is a type 1 and the red lines indicate that the dislocation is a type 2. **(b)** An unmarked image of the same image as in **(a)**, some of the dislocations marked in **(a)** are visible here. The two top type 2 dislocations are slightly harder to see.

The QDs which create the type 2 dislocations were measured, but since this is a HRTEM image, the measured sizes will only be a rough estimate. The diameter and height of the two QDs were found to be, 27.7 and 7.6 nm for the first QD and 29.1 and 9.5 nm for the second. These values fit well with the estimated values shown in table 4.7.

The defects and stackings going in the <111> direction, where not studied in depth. Therefore, the only thing that can be said about them is that, as in the case for the difference between stacking and defects in the vertical arrangement, is that the difference is most likely

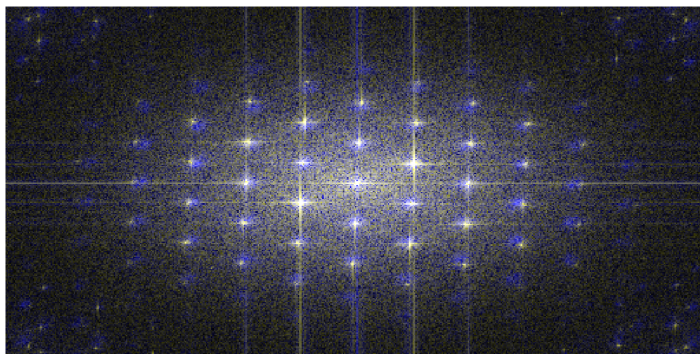


Figure 4.36: The RGB image composed of the FFT from a unstrained area (given the colors red and green) and the FFT from the QD (given the color blue).

due to the QD size. This causes a difference in the elastically absorbed strain and the number or type of dislocations between stackings and defects.

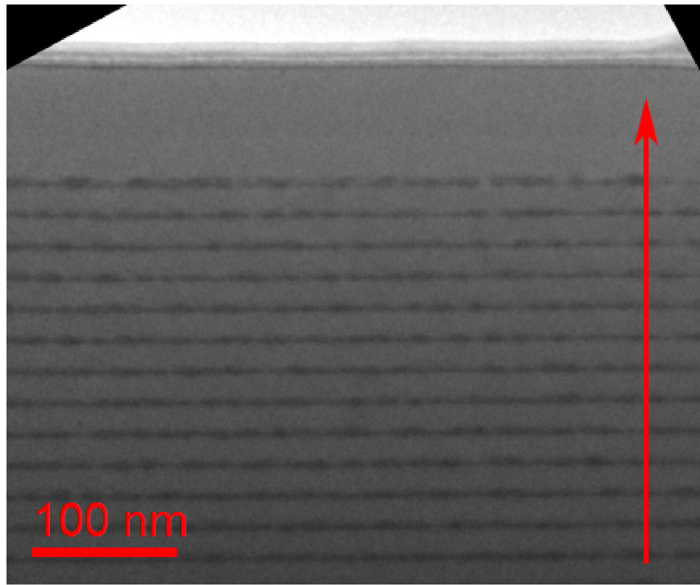
4.5 Polarity determination

The polarity of both samples were determined by the method described in section 2.4.3. This was then correlated to the atomic positions in the dumbbells present in a HR-STEM image taken on the JEOL-ARM200F. Here the pattern will be indexed the same way as in section 2.4.3, so the reflection with the white cross will be the 200 reflection, and the one with the dark cross will be the $\bar{2}00$ reflection. The basis of the unit cell are defined as Ga at (0,0,0) and As at $(\frac{1}{4},\frac{1}{4},\frac{1}{4})$ and the zone axis is [011].

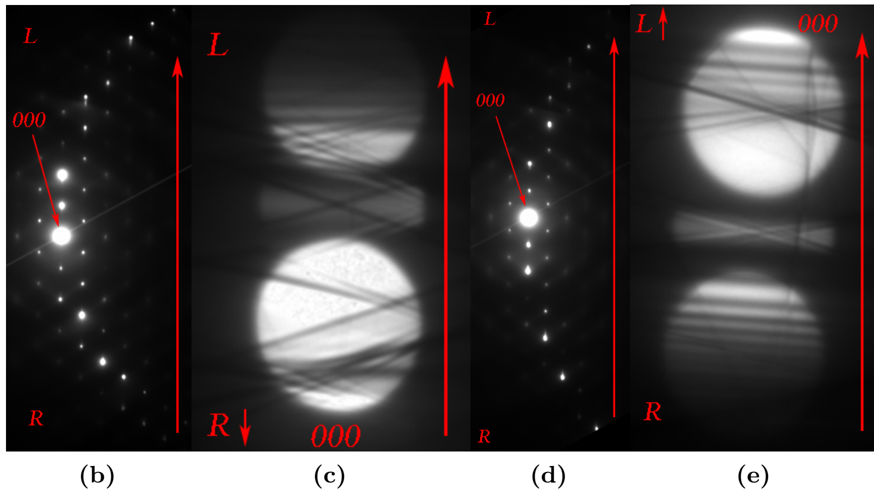
The results from sample A, is presented in figure 4.37a-4.37e. From this it can be seen that the white cross is on the right side of the direct beam, while the dark cross is on the left side. This indicates that the As atom is to the right of the Ga atom in the unit cell, shown in figure 2.31.

The results from sample B, is presented in figure 4.38a-4.38e. Here the white cross is of the right side, meaning that the As atom will be to the right of the Ga atom. This is the same is in the projected unit cell shown in figure 2.31.

As the reflection with the white cross is the 200 reflection, the growth direction both samples must be along (-100) with the chosen

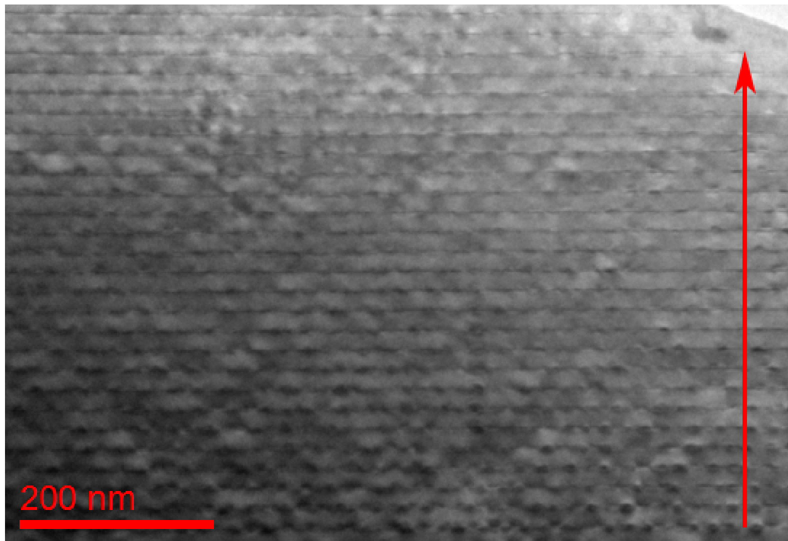


(a)

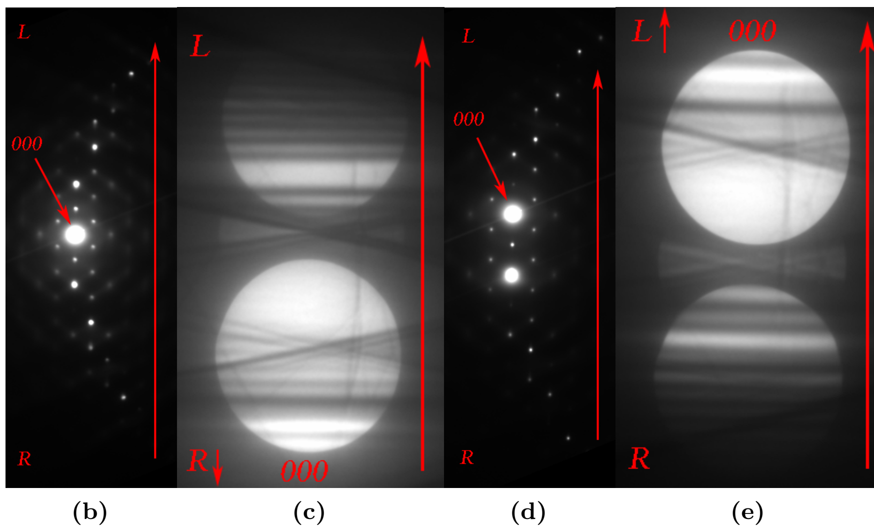


(b) (c) (d) (e)

Figure 4.37: (a) A BF image showing the area where the diffraction and CBED where performed for **sample A**. **(b)** Diffraction pattern in the case corresponding the the CBED pattern in **(c)**. **(d)** Diffraction pattern in the case corresponding the the CBED pattern in **(e)**. In the the images, the large red arrow indicate the growth direction, R and L indicates the right and left side of the direct beam



(a)



(b) (c) (d) (e)

Figure 4.38: **(a)** A BF image showing the area where the diffraction and CBED where performed for **sample B**. **(b)** Diffraction pattern in the case corresponding the the CBED pattern in **(c)**. **(d)** Diffraction pattern in the case corresponding the the CBED pattern in **(e)**. In the the images, the large red arrow indicate the growth direction, R and L indicates the right and left side of the direct beam

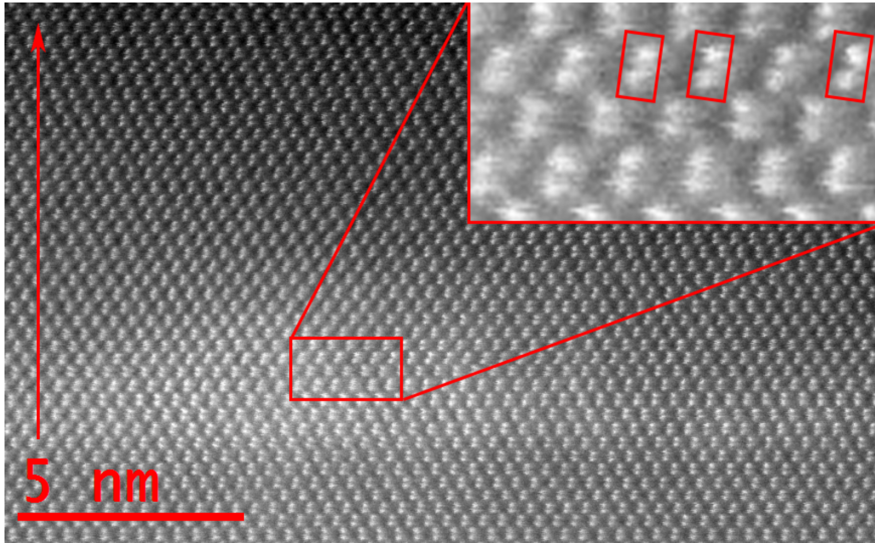


Figure 4.39: A HAADF HR-STEM image taken on the JEOL-ARM200F of a single QD from sample A. The magnified parts show the dumbbells of InAs / GaAs. In the dumbbells highlighted by the red rectangles a difference in intensity can be seen, with In atoms having the highest intensity on top, and the As atoms below. The slight rotation of the dumbbels is due to sample drift, or instrument related causes. The red arrow indicates the growth direction.

basis. Therefore when viewing a HR-STEM image, the Ga atom is expected to be at the top in the dumbbell. Since the Ga and As are so similar in atomic number, it won't be possible to differentiate between the two in a STEM image, but what can be seen is the difference between In and As. The Ga position can be inferred from this, as the In atom will replace Ga in the unit cell in InAs. A HAADF HR-STEM image of a single QD from sample A is shown in figure 4.39, and in the inset part it can be seen that the In atoms are on top in the dumbbell.

These results correspond to what was found previously with CBED for sample A, and so the CBED technique provide a easy method for determining the polarity in epitaxially grown samples with GaAs. The paper describing this technique also shows it's applicability to other III-V semiconductors [40].

The position of the Ga atoms have to be constant throughout the material, i.e. Ga atoms are always on top of the dumbbell, or In in InAs, or Al in AlAs. If this wasn't true, there would be chemical bonds

between As-As and Ga-Ga, which isn't possible. Therefore, from this method, the polarity of the entire sample, for both sample A and B, has been determined.

Chapter 5

Conclusion

In this work, two samples of quantum dot based intermediate band solar cells were compared. The basic structure of these samples where a repeating structure of 50 layers of InAs and a spacer layer, where one sample had been added N for strain compensation. Using BF TEM imaging, LAADF and HAADF STEM imaging, the crystal structure of the two samples were investigated. By comparing the number of defects qualitatively, it was found that the sample without N showed a higher quality crystal structure.

EDS and EPMA analysis were then performed in order to determine the N content of the two samples, to see if they had been switched. From this investigation it was determined that the N content for both samples was below the detection limit of 200 ppm from the EPMA study.

Using high resolution HAADF STEM images, typical sizes of the quantum dots were found. The sizes were compared between the two samples, and also between the different quantum dot layers. It was found that there was no difference in size between the different layers, but that the QDs were larger in the sample which nominally contained nitrogen. The typical values for the sample supposedly containing N were an height and diameter of 3.28 ± 0.071 nm and 13.56 ± 0.29 nm, respectively. For the sample without N, the height and diameter were respectively 2.53 ± 0.084 nm and 11.06 ± 0.30 nm.

This were then used to explain the larger defect density in the sample supposedly containing N, because the larger quantum dots introduce more strain in the sample. The larger quantum dot sizes were attributed to different parameters used during the growth process.

The quantum dot density of both samples were also found using EELS, to calculate the volume, and BF-imaging to count the quantum dots. The sheet density was found to be higher in the sample containing no N, with an average of $23.8 \pm 0.5\ 10^{10}\ \text{cm}^{-2}$, whereas the other sample had an average of $9.7 \pm 0.3\ 10^{10}\ \text{cm}^{-2}$.

The defect structures, present in both samples, were studied by LAADF and HAADF STEM imaging. It was here found that the defect structures consisted of defects along the growth direction and along the $\{111\}$ planes. It was shown that $\{111\}$ dislocations originating from defect lines, reduced the amount of strain in the structure.

A more in depth analysis of the dislocations from various quantum dots were also performed using HRTEM imaging. The various types of dislocations along $\{111\}$ planes were presented, and a explanation for why some quantum dots evolve into defects while others evolve into stacks that tapers off after a while, were also given. It was shown that in large quantum dots the crystal structure will relax into the lattice of InAs by the creation of V-shaped dislocations along $\{111\}$ planes.

Lastly, the polarity of both samples were found using CBED. This was checked against what is seen in the dumbbells of InAs from a high resolution HAADF-STEM image.

Chapter 6

Further work

A further line of study based on concepts studied on this thesis would be a more in depth study of the dislocations occurring in the QDs.

The investigation on why some QDs evolve into defects, and other just stack, should be done in greater detail. For each of these QDs the size should be measured by STEM, and this should be correlated to the dislocations seen in HRTEM images. This should be done for both vertical defects and defects along $\{111\}$ planes.

The strain from QDs could also be measured quantitatively, and by comparing the two different cases, for both vertical and defects along $\{111\}$ planes, a relation between the dislocations and strain could be established.

It might also be interesting to examine the critical size for the QDs to start to relax into the lattice of InAs, and the typical sizes for QDs to develop into defects and stacks.

Bibliography

[1] J. L. Sawin, “Renewables 2012: Global status report,” tech. rep., REN21, 2012.

[2] “Solar generation 6: Solar photovoltaic electricity empowering the world,” tech. rep., European Photovoltaic Industry Association, 2011.

[3] G.Masson, M. Latour, M. Rekinger, I. T. Theologitis, and M. Papoutsi, “Global market outlook: For photovoltaics 2013-2017,” tech. rep., European Photovoltaic Industry Association, 2013.

[4] W. Shockley and H. J. Queisser, “Detailed Balance Limit of Efficiency of pn Junction Solar Cells,” *Journal of Applied Physics*, vol. 32, pp. 510–519, 1961.

[5] M. A. Green, *Third generation photovoltaics: Advanced solar energy conversion* . Springer, 2003.

[6] F. ISE, “World record solar cell with 44.7% efficiency,” September 2013. Accesed march 2014.

[7] D. B. Williams and C. B. Carter, *Transmission Electron Microscopy*, vol. 1, 2, 3, 4. Springer Science, 2nd ed., 2009.

[8] J. Nelson, *The physics of solar cells.* Imperial College press, 2003.

[9] A. Luque and A. Martí, “Increasing the efficiency of ideal solar cells by photon induced transitions at intermediate levels,” *Physical Review Letters*, vol. 78, no. 26, p. 5014, 1997.

[10] A. Martí, L. Cuadra, and A. Luque, “Partial filling of a quantum dot intermediate band for solar cells,” *Electronic Devices*, vol. 48, no. 10, pp. 2394–2399, 2001.

[11] R. Strandberg and T. W. Renaas, "Optimal Filling of the Intermediate Band in Idealized Intermediate-Band Solar Cells," *Electron Devices, IEEE Transactions on*, vol. 58, no. 8, pp. 2559 – 2565, 2011.

[12] A. Martí, C. Stanley, and A. Luque, *Nanostructured Materials for Solar Energy Conversion.* Elsevier, 2006. Chapter 17.

[13] A. Luque, A. Martí, C. Stanley, N. López, L. Cuadra, D. Zhou, J. L. Pearson, and A. Mckee, "General equivalent circuit for intermediate band devices: Potentials, currents and electroluminescence," *Journal of Applied Physics*, vol. 96, p. 903, 2004.

[14] A. Martí and A. Luque, *Next Generation Photovoltaics - High efficiency through full spectrum utilization.* Institute of Physics Publishing, 2004.

[15] A. Martí, L. Cuadra, and A. Luque, "Quantum dot intermediate band solar cell," *Photovoltaic Specialists conference. Conference Record of the Twenty-Eigth IEEE*, pp. 940–943, 2000.

[16] A. Luque, A. Martí, and C. Stanley, "Understanding intermediate-band solar cells," *Nature Photonics*, vol. 6, pp. 146 – 152, 2012.

[17] G. Brown and J. Wu, "Third generation photovoltaics," *Laser & Photonics Reviews*, vol. 3, no. 4, pp. 394–405, 2009.

[18] S. M. Hubbard, C. D. Cress, C. G. Bailey, R. P. Raffaelle, S. G. Bailey, and D. M. Wilt, "Effect of strain compensation on quantum dot enhanced GaAs solar cells," *Applied Physics Letters*, vol. 92, p. 123512, 2008.

[19] X. Q. Zhang, S. Ganapathy, H. Kumano, K. Uesugi, and I. Suemune, "Photoluminescence study of InAs quantum dots embedded in GaNAs strain compensating layer grown by metalorganic-molecular-beam epitaxy," *Journal of Applied Physics*, vol. 92, no. 11, p. 6813, 2002.

[20] X. Q. Zhang, S. Ganapathy, I. Suemune, H. Kumano, and K. Uesugi, "Improvement of InAs quantum-dot optical properties by strain compensation with GaNAs capping layers," *Applied Physics Letters*, vol. 83, no. 22, pp. 4524–4526, 2003.

[21] R. Oshima, Y. Nakamura, A. Takata, and Y. Okada, "Optical properties of multi-stacked InAs/GaNAs strain-compensated quantum dots," *Journal of Crystal Growth*, vol. 310, pp. 2234–2238, 2008.

[22] Y. Okada, R. Oshima, and A. Takata, "Characteristics of InAs/GaNAs strain-compensated quantum dot solar cell," *Jornal of Applie Physics*, vol. 106, p. 024306, 2009.

[23] Y. Okada, T. Morioka, K. Yoshida, R. Oshima, Y. Shoji, T. Inoue, and T. Kita, "Increase in photocurrent by optical transitions via intermediate quantum states in direct-doped InAs/GaNAs strain-compensated quantum dot solar cell," *Journal of Applied Physics*, vol. 109, p. 024301, 2011.

[24] A. Takata, R. Oshima, Y. Shoji, K. Akahane, and Y. Okada, "Fabrication of 100 Layer-Stacked InAs/GaNAs Strain-Compensated Quantum Dots on GaAs (001) for Application to Intermediate Band Solar Cell," *Photovoltaic Specialists conference (PVSC), 35th IEEE*, pp. 1877–1880, 2010.

[25] T. Morioka, Y. Shoji, and Y. Okada, "Current enhancement in direct-doped InAs/GaNAs strain-compensated quantum dot solar cell," *Photovoltaic Specialists Conference, 37th IEEE*, pp. 3499–3502, 2011.

[26] A. Takata, R. Oshima, Y. Shoji, and Y. Okada, "Optical studies on InAs/InGaAs/GaNAs strain-compensated quantum dots grown on GaAs (0 0 1) by molecular beam epitaxy," *Journal of Crystal Growth*, vol. 311, pp. 1774–1777, 2009.

[27] R. Oshima, A. Takata, and Y. Okada, "Strain-compensated InAs/GaNAs quantum dots for use in high-efficiency solar cells," *Applied Physics Letters*, vol. 93, p. 083111, 2008.

[28] A. Martí, E. Antolín, E. Cánovas, N. López, P. Linares, A. Luque, C. Stanley, and C. Farmer, "Elements of the design and analysis of quantum-dot intermediate band solar cells," *Thin Solid Films*, vol. 516, no. 20, pp. 6716 – 6722, 2008. Proceedings on Advanced Materials and Concepts for Photovoltaics EMRS 2007 Conference, Strasbourg, France.

[29] W. H. Strehlow and E. L. Cook, “Compilation of energy band gaps in elemental and binary compound semiconductors and insulators,” *Journal of Physical and Chemical Reference Data*, vol. 2, p. 163, 1973.

[30] S. Adachi, “GaAs, AlAs, and $Al_xGa_{1-x}As$: Material parameters for use in research and device applications,” *Journal of Applied Physics*, vol. 58, p. R1, 1985.

[31] V. Popescu, G. Bester, M. C. Hanna, A. G. Norman, and A. Zunger, “Theoretical and experimental examination of the intermediate-band concept for strain-balanced (in,ga)as/ga(as,p) quantum dot solar cells,” *Physical Review B*, vol. 78, p. 205321, 2008.

[32] I. Vurgaftman and J. R. Meyer, “Band parameters for nitrogen-containing semiconductors,” *Journal of Applied Physics*, vol. 94, p. 3675, 2003.

[33] `http://en.wikipedia.org/wiki/File:Sphalerite-unit-cell-depth-fade-3D-balls.png`, Accessed May 2014.

[34] http://en.wikipedia.org/wiki/File:Wurtzite-unit-cell-3D-balls.png. `http://en.wikipedia.org/wiki/File:Wurtzite-unit-cell-3D-balls.png`, Accessed May 2014.

[35] `http://www4.nau.edu/microanalysis/microprobe-sem/instrumentation.html`, Accessed May 2014.

[36] `http://advressys.com/analytic/sem.htm`, Accessed May 2014.

[37] `http://en.wikipedia.org/wiki/File:Scheme_TEM_en.svg`, Accessed May 2014.

[38] C. Kittel, *Introduction to Solid State Physics*. Wiley, 8 ed., 2005. Chapter 2.

[39] D. C. Joy, D. E. Newbury, and D. L. Davidson, “Electron channeling patterns in the scanning electron microscope,” *Journal of Applied Physics*, vol. 53, no. 8, pp. R81–R122, 1982.

[40] K. Marthinsen, T. Lindheim, and R. Hier, "Non-centrosymmetry Effects and Polarity Determination in III–V Semiconductors," *Acta Crystallographica Section A*, vol. 53, pp. 366–375, May 1997.

[41] J. Goodge. `http://serc.carleton.edu/research_education/geochemsheets/techniques/EPMA.html`, Accessed May 2014.

[42] R. Castaing, "Electron probe microanalysis," vol. 13 of *Advances in Electronics and Electron Physics*, pp. 317 – 386, Academic Press, 1960.

[43] R. F. Egerton, *Electron Energy-loss spectroscopy in the electron microscope.* Plenum press, 2nd ed., 1196.

[44] M. A. Herman and H. Sitter, *Molecular beam epitaxy - Fundamentals and current status.* Springer, 1996.

[45] P. B. Joyce, T. J. Krzyzewski, G. R. Bell, B. A. Joyce, and T. S. Jones, "Composition of InAs quantum dots on GaAs(001): Direct evidence for (In,Ga)As alloying," *Phys. Rev. B*, vol. 58, pp. 15981–15984, Dec 1998.

[46] Z. M. W., *Self-assembled quantum dots, ch. The InAs/GaAs(001) Quantum Dots Transition: Advances on Understanding.* Springer New york, 2008.

[47] P. B. Joyce, T. J. Krzyzewski, G. R. Bell, T. S. Jones, S. Malik, D. Childs, and R. Murray, "Effect of growth rate on the size, composition, and optical properties of inas/gaas quantum dots grown by molecular-beam epitaxy," *Phys. Rev. B*, vol. 62, pp. 10891–10895, Oct 2000.

[48] C. Heyn, A. Bolz, T. Maltezopoulos, R. Johnson, and W. Hansen, "Intermixing in self-assembled InAs quantum dot formation," *Journal of Crystal Growth*, vol. 278, pp. 46 – 50, 2005. 13th International Conference on Molecular Beam Epitaxy.

[49] C. Heyn, A. Schramm, T. Kipp, and W. Hansen, "Kinetic model of intermixing during self-assembled InAs quantum dot formation," *Journal of Crystal Growth*, vol. 301, pp. 692–696, Apr. 2007.

[50] `http://www.ntnu.edu/geminicentre/nortemmic`, Accessed May 2014.

[51] J. T. Ng, U. Bangert, and M. Missous, "Formation and role of defects in stacked large binary InAs/GaAs quantum dot structures," *Semiconductor Science and Technology*, vol. 22, no. 2, p. 80, 2007.

[52] T. Grieb, K. Möller, O. Rubel, R. Fritz, C. Gloistein, N. Neugebohrn, M. Schowalter, K. Volz, and A. Rosenauer, "Determination of Nitrogen Concentration in Dilute GaNAs by STEM HAADF Z-Contrast Imaging," *Journal of Physics: Conference Series*, vol. 326, no. 1, p. 012033, 2011.

[53] P. E. Vullum, M. Nord, S. F. Thomassen, C. Boothroyd, R. Holmestad, B. O. Fimland, and T. W. Reenaas, "Quantitative strain analysis of InAs/GaAs quantum dot materials," *Unpublished*, 2013.