Effect of amorphous Si buffer layer on the ZnO/Si interface

Emil Blix Wisborg
© 2016 Emil Blix Wisborg

Effect of amorphous Si buffer layer on the ZnO/Si interface

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo
Abstract

Thin films of amorphous silicon were deposited as buffer layers between n- and p-type silicon and aluminium-doped zinc oxide, in order to reduce growth of insulating SiO$_2$. The buffer layers have been deposited by PECVD at substrate temperatures in the range 180-220$^\circ$, with thicknesses in the range 5 – 20 nm. In addition, buffer layers with similar parameters were deposited on both sides of double-side polished substrates, for lifetime measurements by the QSSPC method. Characterization of the Si/aSi/ZnO-structures was focused on electrical characteristics. Both current-voltage (IV) and capacitance-voltage (CV) measurements show deviations from thermionic emission theory. These deviations are supported by broad peaks observed in Deep Level Transient Spectroscopy (DLTS), indicating several high-concentration defect levels throughout the band gap. In addition, the difference in IV and CV results is considered to be due to inhomogeneity in the junction properties. QSSPC measurements show a higher minority carrier lifetime in samples with buffer layers deposited at higher temperatures, with a possible shift in interface structure at the highest temperature.
Acknowledgements

First of all, I would like to thank Prof. Edouard Monakhov, for guiding me through the work this thesis has led to. Per Lindberg has helped with all the instruments employed and given great advice when I was lost in the wilderness that is science. For a fantastic introduction to semiconductors and inspiring me to choose this venue for my Master’s Degree, a big thanks goes out to Bengt G. Svensson.

I would also like to thank Ilia Kolevatov for showing me the ropes in the electrical lab, and everyone at the lunch table at LENS for good company during meals.

Last but not least, I would like to thank my wife Ina and my children, for supporting me through this long process.
Contents

1 Introduction 1

I Theory 2

2 Physical principles 2

2.1 Crystalline materials 2
2.1.1 Crystal structures 2
2.1.2 Crystal defects 4

2.2 Semiconductors 5
2.2.1 Energy bands 5
2.2.2 Charge carriers 8
2.2.3 Amorphous semiconductors 10
2.2.4 Deep levels 10

2.3 pn-junctions 13
2.3.1 Equilibrium properties 14
2.3.2 The junction under a bias voltage 15
2.3.3 Depletion capacitance 17

2.4 Schottky- and heterojunctions 18
2.4.1 Metal-semiconductor junctions 18
2.4.2 Heterojunctions 19

2.5 The solar cell 20
2.5.1 Operating principles 20
2.5.2 Theoretical efficiency limit and third-generation solar cells 22

II Processing and characterization methods 25

3 Synthesis methods 25
3.1 Plasma in semiconductor synthesis 25
3.2 Magnetron sputtering 26
3.3 Plasma-enhanced chemical vapour deposition 27
3.4 Contact deposition 28

4 Sample preparation 30

5 Characterization methods 34
5.1 Profilometry 34
5.2 Deep-level transient spectroscopy (DLTS) 34
5.2.1 Principle of operation 34
5.2.2 Signal generation 36
5.2.3 DLTS signal analysis 38
III Results and discussion

6 Results and discussion
   6.1 Current and capacitance measurements
   6.2 DLTS measurements
   6.3 Lifetime measurements (QSSPC)

7 Summary and Conclusions
   7.1 Suggestions for further work

8 Attachments

References
1 Introduction

From 1973 to 2014, the world’s energy consumption has increased from 54 000 TWh to more than 100 000 TWh. The top three sources of energy production are coal, oil and natural gas, all fossil fuels. In 2014, these three energy sources combined supplied more than 80% of the world’s total energy production. [1] Fossil fuels, being a limited energy source, cannot sustain the world indefinitely. In order to meet our continuously rising energy demand, the output from renewable energy sources, such as photovoltaic (PV) solar cells, must increase.

Compared to fossil fuels, the energy production from PV solar cells is vanishingly small. However, the PV energy market has experienced rapid growth in the last five years [2]. The main reason the PV market is so small is the high cost of PV energy compared to other sources. PV solar energy is not yet competitive with fossil energy, cost-wise. In order to increase the energy/cost ratio, solar cell efficiency has to increase, while manufacturing costs have to decrease.

One way of increasing the efficiency of PV solar cells is reducing the shade from the front contacts. This can be done by using a transparent conducting oxide (TCO) instead of metallic contacts. Zinc oxide has a band gap of over 3 eV, and is thus viable as a TCO. However, when zinc oxide is deposited on silicon, a layer of insulating SiO$_2$ is created, which is detrimental for the interface. An amorphous silicon buffer layer can prevent the growth of SiO$_2$, and thus improve the interface.

The aim of this work is to explore the effect of a buffer layer of amorphous silicon as a means to alter the interface between zinc oxide and crystalline silicon. The buffer layer has been deposited at different temperatures and thicknesses, in order to investigate how these variables affect the electrical properties of the interface.
Part I
Theory

2 Physical principles

The purpose of this chapter is to give an overview of the basic theory needed to understand the concepts discussed in this thesis. The reader is assumed to have some knowledge of solid state physics. The chapter starts with a short introduction to crystal structures and defects, and then introduces the properties of semiconductors and semiconductor devices and their application in photovoltaics.

2.1 Crystalline materials

2.1.1 Crystal structures

The following review of crystal structures is based on textbooks by Tilley [3] and Streetman [4].

Crystallography is the study of crystalline solids. Until the early 1900’s, crystallography developed via two independent routes. The first of these was based on measurements of the angles and planes of crystalline solids. The regularity of crystals was believed to rise from an internal order, and their shapes were classified into seven different crystal systems. Crystalline symmetry, along with the observation that many crystals could be cleaved into smaller and smaller units, was what gave rise to the concept of the unit cell. The second route based itself on the mathematical description of the arrangement of arbitrary objects in space, and gave rise to the Bravais lattices. After the discovery of X-rays, the X-ray diffraction method developed quickly, following Bragg’s law in 1913 [5], which describes the relation between the scattered radiation and the distance between evenly spaced planes within a crystal. This technique, along with electron and neutron diffraction, unified the observational approach with the mathematical one, and gave a higher understanding of crystalline materials.

Crystals are distinguished by having their atoms arranged in a periodic fashion. That is, there is a basic three-dimensional arrangement of atoms which is repeated throughout the entire solid. This basic arrangement of atoms, the smallest unit that can be used to describe the larger structure of the crystal, is called the unit cell. A primitive unit cell is a unit cell constructed in a way that causes it to contain only a single lattice point, which is an atom or a molecule, and translation vectors representing the lattice. The seven crystal systems mentioned earlier are based on the different angles and For increased visual clarity, larger, non-primitive unit cells containing
several atoms or molecules are often used.

Which crystal structure a solid will adopt depends among other things on the bonding type in the solid. As the metallic bonding often is non-polar, most metals and alloys adopt a close packed structure, in which the atoms are considered as hard spheres. In pure elemental materials with only one type of atoms, the highest packing density is achieved by two almost similar structures: the face-centred cubic (FCC) structure and the hexagonal close-packed (HCP) structure in the Bravais system. In both structures, each atom has twelve nearest neighbours in three dimensions, and their difference arise in the plane stacking.

Covalent crystals, such as diamond and silicon, consist of atoms held together by a network of covalent bonds. As such, their adopted structure is affected by the number of valence electrons in the atoms, and energetically favorable bond angles. Silicon, having four valence electrons, crystallizes in a tetrahedral coordination, called the diamond structure. This is simplest described as an FCC lattice with an extra atom placed at $[1/4 \ 1/4 \ 1/4]$ from each of the FCC atoms. The FCC structure and the diamond structure can be seen in Figure 1. The extra atoms in the diamond structure are colored for visibility.

The zinc blende structure is a diatomic variation of the diamond structure, with different atoms on alternating sites. In this structure, each element occupies its own FCC orientation, and it is adopted by many compound semiconductors. The wurtzite structure is similar to zinc blende, but is based on the HCP structure. Zinc oxide, when grown as a sputtered film, adopts the wurtzite structure, as done in this work.

In addition to all crystal structures are materials without a crystal structure, i.e. amorphous materials. Most commonly observed in covalent-bonding materials, these materials exhibit covalent bonding on the microscopic scale. However, they show no long range ordering, due to deviations to bond angling and unbonded valence electrons, called dangling bonds. As a result, no planes or crystal directions are present. This causes diffra-
tion analysis results of these materials to become smeared out rather than forming clear patterns.

2.1.2 Crystal defects

No real crystalline solid has perfect periodicity. Deviations from the idealized structure described in the previous section are called defects. Defects modify many important properties of the crystals, and are thus important to semiconductor science.

There are two main ways to categorize defects. One way is into intrinsic and extrinsic defects (explained in the following subsections), the other is from their dimensionality in the material. One-dimensional defects, often called line defects, are usually found as dislocations around which some of the atoms of the crystal lattice are misaligned, and can arise as a result of strain on the crystal. Two-dimensional defects often appear as stacking faults or grain boundaries. Stacking faults are disruptions in the repeating sequence of atom layers, while grain boundaries are inevitable inter-crystallite borders in multi-crystalline materials. In addition, surfaces are considered two-dimensional defects. Three-dimensional defects, which extend in all three dimensions, can be cavities in the material, areas of a different phase, or a precipitate of a foreign material.

Particularly important in the scope of this work are zero-dimensional defects, commonly referred to as point defects. These are often electrically charged compared to the rest of the material, and they thus influence the electrical properties of the system. As such, point defects are given the most consideration in this work, and in most cases, the term defects will refer to point defects.

Intrinsic defects

Intrinsic defects exist in all crystals, even pure materials and stoichiometric compound crystals. Two types of point defects can occur in a pure crystal: vacancies and self-interstitials. The existence of these defects means the Gibbs free energy $G$ is at a minimum at a non-zero number of defects. The Gibbs free energy is expressed by the enthalpy $H$ of formation of the defect and the entropy $S$ associated with the defect distribution:

$$G = H - TS$$ \hfill (2.1)

The temperature dependence in this relation is characteristic for the concentration of intrinsic defects in crystals.

The simplest intrinsic defect is a Schottky defect. This is a vacancy, a missing ion or atom, in the crystal lattice. In ionic crystals, the number of positive and negative ion vacancies are usually roughly equal, as this is more energetically favorable. The formation of vacancy pairs keeps the crystal electrostatically neutral on a local scale.
In addition, a defect pairing known as a Frenkel pair can occur. This is composed of a vacancy and an interstitial of the same element. While both Schottky and Frenkel pairs are present in most materials, one tends to dominate over the other, depending on various factors including packing density and, for ionic materials, size ratio between anions and cations.

**Extrinsic defects**
Extrinsic defects arise from foreign atoms in the crystal structure. As point defects, these can manifest as interstitials or substitutions. A very high local concentration of foreign atoms can precipitate as a volume consisting of another phase, creating a three-dimensional defect. This precipitation is more thermodynamically favorable than a solid solution of point defects, but it is possible to avoid by rapidly cooling the material in order to freeze in defects.

Introducing controlled concentrations of specific foreign elements is called *doping*, and is done to alter the electronic properties of the material. Even trace amounts of foreign atoms can be enough to significantly change the electronic properties of semiconductors. This is used widely in the industry in order to make small-scale electronic devices like transistors and diodes. In the same way, even low concentrations of the wrong atoms in the semiconductor can make the material unusable for the same applications. This will be further described in section 2.2.2.

### 2.2 Semiconductors

Solid materials are commonly categorized by their ability to conduct electricity, and are divided into three main categories: metals, semiconductors and insulators. While metals and insulators have important uses in electrics and electronics, the electrical properties of semiconductors form the foundation of nearly all modern electronics, mostly through the transistor and the diode. In addition, solar cells have their foundation in semiconductor physics, which is relevant in this work. In order to understand the electrical properties of semiconductors, the fundamentals of electronic states in atoms and groups of atoms must be considered. In this review, crystalline structures will be used, while the key deviations of amorphous structures will be given in section 2.2.3. This review is based on textbooks by Streetman et al. [4], Nelson [6], and Campbell [7].

#### 2.2.1 Energy bands
In isolated atoms, electrons occupy atomic orbitals, formed at discrete energy levels inside the coulombic potential well of the nucleus. As two atoms form a bond, their atomic orbitals combine to form molecular orbitals, each
formed at a discrete energy level. As more atoms are included in the structure, the molecular orbitals grow larger, and the overlapping states form continuous regions on the energy scale that are available to electrons, called energy bands. These bands are separated by regions on the energy scale without available states, called band gaps. By exploiting the periodic nature of crystals, the structure of the energy bands can be calculated. Solving the single-electron Schrödinger equation for an electron in a lattice-periodic potential gives Bloch waves as solutions on the form
\[ \psi_{nk}(r) = e^{i\mathbf{k} \cdot \mathbf{r}} u_{nk}(r), \] (2.2)

where \( \mathbf{k} \) is a wave vector in reciprocal space, and is proportional to the crystal momentum. From the Bloch waves, a dispersion relation \( E_n(\mathbf{k}) \) can be calculated for each band \( n \), and the band structure can be represented by plotting \( E_n(\mathbf{k}) \) against \( |\mathbf{k}| \) for a selection of points in reciprocal space corresponding to important directions in the crystal. This is shown in Figure 2 for Si and ZnO.

At absolute zero temperature, every electron will occupy the lowest available energy level, and all higher energy levels will be left unoccupied. The boundary between the highest occupied and the lowest unoccupied energy level is known as the Fermi energy, and its location relative to the energy bands determines whether the material is considered a metal, a semiconductor or an insulator, as can be seen in Figure 3. If the Fermi energy is within a band, electrons can be excited to conducting levels with the slight-
Figure 3: A simplified band diagram, showing only the valence band and the conduction band at their smallest band gap.

The Fermi level is the highest energy at which an electron can be excited into conduction, and is located within the energy gap between the valence and conduction bands. In a material where the Fermi level lies above the valence band, there is no energy gap, and the material is then classified as a metal. For semiconductors and insulators, the Fermi energy is located in the energy gap between two bands. In this case, the band below the Fermi energy is called the valence band, and the band above the Fermi energy is called the conduction band. Whether the material is considered a semiconductor or an insulator depends on the size of the band gap, and the distinction varies depending on context and work area.

It is important to note that Figure 3 is a simplified band gap diagram, considering only the smallest value of the band gap at the Fermi energy. This allows the x-axis to represent the thickness or depth into a device, but some of the properties of the $E(k)$ relationship are lost. First among these is whether the band gap of the material is direct or indirect. This depends on what value of $k$ the extremes on the valence band and the conduction band are located. Zinc oxide, as seen in Figure 2, has the valence and conduction band extremes at the same $k$ value, and as such has a direct band. Silicon, on the other hand, has the band extremes located at different values of $k$, and has an indirect band gap. This affects how the electrons interact between the valence band and the conduction band.

In a material with a direct band gap, the emission of a photon can take place directly by relaxation of an excited conduction electron across the band gap. In a material with an indirect band gap, an electron cannot relax across the smallest value of the band gap without a momentum change as well. As photons are massless particles, this change in momentum must be provided from interaction with a lattice vibration wave, called a phonon. This extra demand implies that direct transitions will be less likely, and thus the lifetime of excited electrons higher, in materials with an indirect band gap, which is often the case. In addition, the curvature of the energy bands carries information about the physical properties of the material. The periodic potential influences the charge carriers in the material. This influence is usually incorporated as an effective mass, a pre-factor to the electron rest mass used in calculations. Lastly, the density of states (DOS) in the bands is affected. For calculations, parabolic functions are used to approximate
the band curvature in the material.

2.2.2 Charge carriers

A pure, undoped semiconductor is called an intrinsic semiconductor. At low temperatures, all electrons are bound in the completely filled valence band, and no electrical current can take place. Energy equal to or larger than the band gap is needed in order to excite an electron to the conduction band, where it can move freely, thus making an electric current possible. As the electron is excited, it leaves behind an unoccupied state in the valence band. A neighbouring valence electron can move into this state, effectively moving the unoccupied state in the other direction. This state is considered a positively charged quasi-particle called a hole, and contributes to the electrical current in the semiconductor. Electrons and holes are the charge carriers considered in semiconductor devices.

At non-zero temperatures, a carrier generation rate is provided by the thermal energy in the semiconductor exciting electrons to the conduction band, and a carrier recombination rate is caused by electrons from the conduction band relaxing to the valence band and re-occupying the energy state of a hole. At thermal equilibrium, the generation rate and the recombination rate are balanced, resulting in a net density of electron-hole pairs in the material.

The equilibrium charge carrier density at a given temperature is calculated by first considering the Fermi-Dirac distribution function:

$$f(E) = \frac{1}{1 + e^{\frac{E - E_F}{kT}}}$$ (2.3)

where $k$ is Boltzmann’s constant. This function describes the probability that an available energy state at $E$ will be occupied by an electron at the absolute temperature $T$. The quantity $E_F$ is called the Fermi level, and represents the thermodynamic work to add one electron to the body. It can be noticed that at the Fermi level, that is, at $E = E_F$, an energy state has a probability of 50% of being occupied by an electron. For intrinsic (undoped) semiconductors, the Fermi level is close to the middle of the band gap. For doped semiconductors, the Fermi level is shifted toward the appropriate band. From the dependence on the thermal energy in equation (2.3), the product $kT$, it can be seen that the shape of the Fermi-Dirac function changes as a function of temperature. At zero K, it takes the shape of a step function, and it smears out at higher temperatures.

By integrating the product of the Fermi-Dirac distribution $f(E)$ and the density of states $N(E)$ over different energies, the charge carrier concentration is found. For electrons, the density is given as:

$$n = \int_{E_c}^{\infty} f(E)N(E)\,dE,$$ (2.4)
and for holes:

\[ p = \int_{-\infty}^{E_v} (1 - f(E))N(E)dE. \] (2.5)

For energies a distance away from the Fermi level, where the exponential term in the denominator of equation (2.3) dominates, the Fermi-Dirac distribution can be approximated by the Maxwell-Boltzmann distribution:

\[ f(E) = e^{-\frac{E-E_F}{kT}} \] (2.6)

In addition, for relatively low doping levels, the charge carrier distribution drops rapidly for energies away from the edges of the valence and conduction band. Thus the charge carrier distribution can be evaluated at the band edges, rather than through the whole band. This simplifies the equations for the charge carrier concentrations:

\[ n = N_C f(E_C) = N_C e^{-\frac{E_C-E_F}{kT}} \] (2.7)

for electrons, and

\[ p = N_C (1 - f(E_C)) = N_C e^{-\frac{E_F-E_V}{kT}} \] (2.8)

for holes. \( N_C \) and \( N_V \) are here the effective densities of states at the conduction and valence band edges, respectively. These depend on the effective mass \( m_n^* \) and \( m_p^* \) of electrons and holes in the material:

\[ N_{C,V} = 2 \left( \frac{2\pi m_n^* m_p^* kT}{\hbar^2} \right)^{\frac{3}{2}} \] (2.9)

for the respective bands. \( h \) is the Planck constant.

In doped materials, dopants are compensated by an electron or a hole, which causes an uneven ratio of electron and hole concentrations. Silicon will here be used as an example. When a silicon atom is substituted by a phosphorus atom, which has one more valence electron, the new atom will have one unbound electron after completing the bonds to its neighbour atoms. This electron is only weakly bonded to its host atom, and will be delocalized by thermal energy at low temperatures, leaving a stationary positive ion. Phosphorus, along other atoms with one more valence electron, is called a donor dopant. In a semiconductor doped with donors, electrons are the dominating charge carrier, and the material is referred to as n-type.

If the substituting atom is boron, which has only three valence electrons, one of the bonds in the structure is left unsatisfied. In order to satisfy this bond, the boron atom accepts an electron from its surroundings. The captured electron leaves behind a hole in the valence band. Boron is in this case an acceptor dopant, and an acceptor-doped semiconductor is known as p-type and has holes as the dominating charge carrier. If both donor and
acceptor dopants are present in the material, the dopant with the highest concentration determines the type.

The number of charge carriers present in a semiconductor also effects its conductivity, which is given by

\[ \sigma = q(n\mu_n + p\mu_p) \]  

(2.10)

where \( q \) is the elementary charge, \( 1.6 \times 10^{-19} \) coulomb, and \( \mu_{n,p} \) is the mobility for electrons and holes, respectively. The electron and hole mobility is a largely material-dependent property describing the drift velocity of the carriers in an electric field.

2.2.3 Amorphous semiconductors

Amorphous semiconductors have a similar structure to crystalline semiconductors on the short range, but have no long range order. While each atom is bound to its nearest neighbours in the symmetry typical of the crystal, the arrangement of atoms further away is more or less random. In the semiconductor, many of the bonds will have deviations in length and orientation, which gives rise to a spread in energy levels. Near the edges of the conduction and valence bands, this spread causes the density of states to decay into the band gap, in an Urbach tail, instead of cutting off sharply. In addition, due to the lack of long-range order, the crystal momentum of electrons does not need to be conserved in amorphous semiconductors, effectively making the material be a direct band gap semiconductor, regardless of whether the crystalline material has a direct or indirect band gap.

In amorphous semiconductors, the lattice contains, in addition to deviations in bond length and orientation, so-called dangling bonds. These are valence orbitals which are not involved in bonding, due to spacing and orientation between atoms. The dangling bonds may be positively charged, neutral or negatively charged, and can contribute to energy levels deep in the band gap. [6]

2.2.4 Deep levels

This section will present the deep level states and their properties, which are an important influence in semiconductor devices and must be considered. These properties are measured using Deep Level Transient Spectroscopy (DLTS), described in chapter 5.2. This section is based on the work of Blood and Orton [10].

Defect states introduced by controlled doping of semiconductors are called shallow states. Their energy level is close to one of the band edges, and most of these states are activated at room temperature. Other defects in the material can have more tightly bound electrons which require more energy to activate. As a result, their energy levels are deeper into the band.
gap, and these are called deep level states. For silicon and other commonly used semiconductors, states are considered deep when the energy separation between the defect state and the closest band edge is larger than $\sim 0.05 eV$.

At high concentrations, deep states can influence both the position of the Fermi level and the free carrier concentration. However, at much lower concentrations, these centers can have a significant effect on the minority carrier lifetime, by acting as carrier traps and recombination centers. For a few devices, deep state centers are intentionally introduced to control lifetime, but for most devices they are an unwanted and disrupting presence, as reduced carrier lifetime is a problem in many applications, and especially in photovoltaics.

Carrier traps have two main properties: energy level and capture cross-section, collectively referred to as the trap signature.

**Carrier capture and emission**

A defect level at an energy level $E_t$ in the band gap with a trap concentration $N_t$ will have four different interactions with the charge carriers in the material, as seen in Figure 4. Two of the processes increase the electron occupancy of the defect level, namely electron capture and hole emission ($c_n$ and $e_p$). The other two processes, electron emission and hole capture ($e_n$ and $c_p$), decrease the occupancy. As electron emission and hole capture occurs at the occupied defect states, while electron capture and hole emission occurs at the unoccupied states, the net rate of change of electron occupancy of the defect level is given by

$$\frac{dn_t}{dt} = (c_n + e_p)(N_t - n_t) - (e_n + c_p)n_t$$  \hspace{1cm} (2.11)

where $n_t$ is the concentration of occupied defect states.

In thermal equilibrium, the capture and emission processes must be in detailed balance; i.e. the capture and emission rates must be equal for each carrier. This condition gives

$$e_n n_t = c_n (N_t - n_t)$$  \hspace{1cm} (2.12)
\[ e_p(N_t - n_t) = c_p n_t \]  
\[ \text{(2.13)} \]

At thermal equilibrium, the occupancy of the trap level is thus given by combining these two equations:

\[ \frac{\dot{n}_t}{N_t} = \frac{c_n}{c_n + e_n} = \frac{e_p}{e_p + c_p} \]  
\[ \text{(2.14)} \]

This can also be expressed by the Fermi-Dirac distribution function. If the degeneracy of a deep state is \( g_0 \) when unoccupied and \( g_1 \) when occupied by a single electron, the occupancy is

\[ \frac{\dot{n}_t}{N_t} = \left\{ 1 + \frac{g_0}{g_1} e^{\frac{E_t - E_F}{kT}} \right\}^{-1} \]  
\[ \text{(2.15)} \]

Combining equations (2.14) and (2.15) gives, for electrons and holes respectively,

\[ \frac{e_n}{c_n} = \frac{g_0}{g_1} e^{\frac{E_t - E_F}{kT}} \]  
\[ \text{(2.16)} \]

\[ \frac{e_p}{c_p} = \frac{g_0}{g_1} e^{\frac{E_F - E_t}{kT}} \]  
\[ \text{(2.17)} \]

As \( (g_1/g_0) \approx 1 \), this part of the expression can be ignored when considering the following. These equations imply that when the Fermi level is higher than the trap level, electron capture and hole emission dominate over their inverse processes, and the trap is occupied by electrons. Conversely, when the Fermi level is below the trap level, the opposite occurs, and the trap is empty. When the trap level is at the same energy as the Fermi level, \( e_n \approx c_n \) and the trap occupancy is 1/2, from equation (2.14). The variation in emission/capture ratios is closely related to the occupation probability given by the Fermi-Dirac distribution, seen in equations (2.3) and (2.15).

**Capture and emission rate**

In order to analyze the properties of deep level traps, the capture and emission rates must be properly defined. The capture process of electrons and holes is characterized by a capture cross section \( \sigma_{n,p} \). Over a short time interval \( \Delta t \) a change in trap occupancy \( \Delta n_t \) caused by electron capture at the \( (N_t - n_t) \) unoccupied states is given by

\[ \Delta n_t = \sigma_n \langle v_n \rangle n (N_t - n_t) \Delta t \]  
\[ \text{(2.18)} \]

The flux of electrons is the product of the concentration of free electrons \( n \) and the root mean square of the thermal velocity \( \langle v_n \rangle \) of said electrons. The electron capture rate per unoccupied state is given as following by rearrangement:

\[ c_n = \frac{\Delta n_t}{\Delta t} = \sigma_n \langle v_n \rangle n \]  
\[ \text{(2.19)} \]
For hole capture, a similar expression can be written.

While the carrier capture rate of traps is proportional with the concentration of free charge carriers $n$, which depends on the doping of the material, the emission rate is not. The emission rate for charge carriers depends on the energy difference between the trap level and the corresponding band ($E_C - E_t$ for electrons), and the capture cross-section $\sigma_n$, which both are intrinsic properties of the trap. This rate can be found for electrons by substituting $c_n$ in equation (2.16) with the expression found in (2.19) and replacing $n$ with equation (2.7):

$$e_n(T) = \sigma_n \langle v_n \rangle N_C e^{-\frac{E_C - E_t}{kT}}$$

The energy required for an electron to be excited into the conduction band from the trap can be related to Gibbs free energy as $\Delta G(T) = \Delta H - T \Delta S = E_C - E_t$:

$$e_n(T) = \sigma_n \langle v_n \rangle N_C e^{-\frac{\Delta G}{kT}}$$

$$e_n(T) = e^\frac{\Delta S}{kT} \sigma_n \langle v_n \rangle N_C e^{-\frac{\Delta H}{kT}}$$

By substituting $\langle v_n \rangle N_C = \gamma T^2$, where $\gamma$ contains all the temperature independent constants, and defining an apparent capture cross section $\sigma_{na} = e^{\Delta S/k} \sigma_n$, the emission rate can be expressed as

$$e_n(T) = \sigma_{na} \gamma T^2 e^{-\frac{\Delta H}{kT}}$$

where $\gamma$ is given by

$$\gamma T^2 = \langle v_n \rangle N_C = \sqrt{\frac{3kT}{m_n^*}} \left( \frac{2\pi m_n^* kT}{h^2} \right)^{\frac{3}{2}} = \sqrt{\frac{3k}{m_n^*}} \left( \frac{2\pi m_n^* k}{h^2} \right)^{\frac{3}{2}} T^2$$

Thus, $\gamma$ is a material- and charge carrier dependent property.

Dividing this by $T^2$ means that a plot of $\ln(e_n/T^2)$ vs $1/T$, called an Arrhenius plot, gives the straight line

$$\ln(e_n/T^2) = -\frac{\Delta H}{kT} + \ln(\sigma_{na} \gamma)$$

from which $\Delta H$ can be found from the slope of the line and $\sigma_{na}$ from where the extrapolated line intercepts the y axis.

### 2.3 pn-junctions

A pn-junction is a boundary between an n-type material and a p-type material, usually within the same semiconductor crystal. This can be manufactured either by implantation or diffusion of dopants into a semiconductor, or by production of a layered material structure. In the following section, a perfect metallurgical junction is assumed, with an uniform doping concentration in both the n-type and p-type material.
2.3.1 Equilibrium properties

A good junction gives a sharp step in charge carrier concentrations at the junction. This leads to diffusion of mobile charge carriers into the opposite material, and leaves ionized dopant atoms in a volume adjacent to the junction. The ionized dopants create an electric field $E$, which produces a drift current countering the diffusion current. At equilibrium, these currents balance each other. The volume near the junction becomes nearly devoid of mobile charge carriers, and is called the depletion region. In order to calculate the width of the depletion region, two assumptions are made. Firstly, the depletion of charge carriers in the depletion region is considered complete in the whole width $W$ around the junction, and the space charge in this region is provided solely by the ionized donors $N_D$ on the n-side and acceptors $N_A$ on the p-side. Secondly, the depletion region is considered to have a sharp cutoff, with no space charge and thus no electric field outside the depletion region.

Due to charge neutrality considerations, the total charge $Q$ from the dopant ions must be equal on both sides of the junction. This leads to the extent of the depletion region being dependent on the doping concentration in the n-type and in the p-type part of the semiconductor:

$$|Q| = qAN_dx_n = qAN_ax_p$$  \hspace{1cm} (2.24)

where $A$ is the area of the junction and $x_{n,p}$ are the penetration lengths of the space charge region into the n- or p-type material, respectively. In this case, the positive $x$-direction is considered to be pointing from the p-type material to the n-type, with $x = 0$ at the junction. The electric field $E$ points in the $-x$-direction, from the n-side to the p-side, and is thus negative throughout the depletion region. Its highest numerical value is found at the junction. The electric field distribution in the transition region can be calculated using Poisson’s equation:

$$\frac{d^2V(x)}{dx^2} = -\frac{dE(x)}{dx} = \frac{\rho}{\varepsilon}$$  \hspace{1cm} (2.25)

where $\rho$ is the charge density and $\varepsilon$ is the permittivity of the semiconductor. Through the depletion approximation, the charge carriers in the depletion region can be neglected, and the charge density is determined by the donor or acceptor density. This gives two regions of constant space charge:

$$\frac{dE(x)}{dx} = -\frac{q}{\varepsilon}N_a \quad \text{for} \quad -x_p < x < 0$$  \hspace{1cm} (2.26a)

$$\frac{dE(x)}{dx} = \frac{q}{\varepsilon}N_d \quad \text{for} \quad 0 < x < x_n$$  \hspace{1cm} (2.26b)

By integrating these two equations, the following expression is found for the maximum value for $E$, found at $x = 0$:

$$E_0 = -\frac{q}{\varepsilon}N_dx_n = -\frac{q}{\varepsilon}N_ax_p$$  \hspace{1cm} (2.27)
As the electric field is assumed to be zero in the neutral regions outside the depletion region, there is a constant potential $V_n$ in the n-side and a constant potential $V_p$ in the p-side. The difference $V_0 = V_n - V_p$ between these two is called the contact potential. The relation between the electric field and the contact potential is derived from equation (2.25):

$$E(x) = -\frac{dV(x)}{dx} \quad \text{or} \quad -V_0 = \int_{-x_p}^{x_n} E(x)dx \quad (2.28)$$

When the material is in equilibrium, the Fermi level must be constant throughout the whole material. As the junction is formed, the energy bands bend to allow this. The difference in energy between corresponding bands in the n- and p-type material resulting from band bending is $qV_0$ (see Figure 5a).

As the width $W$ of the depletion zone is given by $W = x_n + x_p$, and the negative of the contact potential is simply the area under the $E(x)$ triangle, the contact potential can be written as:

$$V_0 = -\frac{1}{2} \frac{\varepsilon}{\varepsilon} E_0 W = \frac{1}{2} \frac{q}{\varepsilon} N_d x_n W \quad (2.29)$$

Using $W = x_n + x_p$ and simplifying equation (2.24) gives:

$$x_n = \frac{WN_a}{N_d + N_a} \quad (2.30)$$

and then:

$$V_0 = \frac{1}{2} \frac{q}{\varepsilon} \left( \frac{N_d N_a}{N_d + N_a} \right) W^2 \quad (2.31)$$

Solving for $W$ gives an expression for the width of the depletion region in terms of the contact potential, the doping concentrations, and the known constants $q$ and $\varepsilon$:

$$W = \left[ \frac{2\varepsilon V_0}{q} \left( \frac{N_d + N_a}{N_d N_a} \right) \right]^{1/2} \quad (2.32)$$

### 2.3.2 The junction under a bias voltage

If a bias voltage $V_{ext}$ is applied to the junction, the equilibrium is disrupted, and the width $W$ of the depletion region changes, as seen in Figure 5. The voltage element $V_0$ in equation 2.32 is replaced with the total junction voltage $V_j = V_0 - V_{ext}$. In this case, $V_{ext}$ is considered positive when the external bias is positive on the p-side relative to the n-side. The Fermi level is no longer defined in this state, as the junction is no longer at equilibrium. To describe the electron populations in the junction in this situation, so-called quasi-Fermi levels, separated by energy $qV_e$, are defined.

Under forward bias, the width of the depletion region decreases, and the electrostatic potential barrier at the junction is lowered by the applied
At equilibrium \((V_{ext} = 0)\)

\[ I(V) = I_0 \left( e^{\frac{qV}{kT}} - 1 \right) \]  
(2.33)

This is known as the \textit{ideal diode equation}.

As the equation shows, current can flow relatively freely under forward bias, while almost no current flows under reverse bias of the junction. The pn-junction shows rectifying behavior, and such structures are known as diodes in electronics.

The above calculations assume that no carrier recombination or generation occurs in the depletion region. In real devices, this is not entirely true. For junctions with a wide depletion region or with defects in the band gap, recombination in the depletion region can be an influencing factor. By
including an ideality factor $n$ in the ideal diode equation, this is accounted for:

$$I(V) = I_0 \left( e^{\frac{qV}{nkT}} - 1 \right)$$  \hspace{1cm} (2.34)

Depending on the material and temperature, $\eta$ usually varies between 1 and 2. As a wide depletion region increases recombination, $\eta$ is also partially dependent on bias voltage. Other factors also cause deviation from the ideal diode equation. At high forward bias, ohmic resistance in the neutral regions becomes significant, limiting the current. The reverse saturation current may increase beyond ideal calculations, through thermal generation of electron-hole-pairs in the depletion region or breakdown effects.

### 2.3.3 Depletion capacitance

Capacitance is the ability of a device to store an electric charge $Q$, and is defined as

$$C = \left| \frac{dQ}{dV_j} \right|$$  \hspace{1cm} (2.35)

The charge on either side of the junction is given by equation (2.24), and inserting (2.30) returns

$$|Q| = qAN_d x_n = qAN_a x_p$$

$$|Q| = qA \frac{N_d N_a}{N_d + N_a} W = A \left( \frac{2q \varepsilon V_j}{N_d + N_a} \right)^{\frac{1}{2}}$$  \hspace{1cm} (2.36)

Using this expression for $|Q|$ gives the following for $C$:

$$C = \left| \frac{dQ}{dV_j} \right|$$

$$C = A \left( \frac{2q \varepsilon}{V_j} \frac{N_d N_a}{N_d + N_a} \right)^{\frac{1}{2}}$$

$$C = \varepsilon A \left( \frac{q}{2 \varepsilon V_j} \frac{N_d N_a}{N_d + N_a} \right)^{\frac{1}{2}} = \frac{\varepsilon A}{W}$$  \hspace{1cm} (2.37)

If one side of the junction is doped more heavily than the other side (e.g. $N_a \gg N_d$), the capacitance simplifies to

$$C = A \left( \frac{\varepsilon q N_d}{2V_j} \right)^{\frac{1}{2}},$$  \hspace{1cm} (2.38)

and is thus defined by the doping in the low doped side of the junction. This is referred to as an asymmetrical junction, and is denoted in this case as a p$^+$-n-junction. In addition, equation (2.24) shows that the depletion region will extend further into the lower doped side.
2.4 Schottky- and heterojunctions

In addition to pn-junctions within a single material, junctions with similar properties can be made between a semiconductor and a metal, or between two different semiconductors. Both types will be considered briefly in the following section.

2.4.1 Metal-semiconductor junctions

At the interface between a metal and a semiconductor the Fermi levels have to align, just as for pn-junctions. To better gain an understanding of this structure, it is useful to introduce the vacuum level, which refers to the energy of a free stationary electron in vacuum. The vacuum level is separated from the Fermi level of a metal by the energy $q\phi$, where $\phi$ is the work function of the metal. In semiconductors, the corresponding energy difference $q\chi$ is between the conduction band edge and the vacuum level, where $\chi$ is called the electron affinity of the semiconductor. When the Fermi levels align, a bending of the vacuum level occurs in the semiconductor, causing discontinuities in available states at the interface. According to the Schottky-Mott model, the vacuum level bending is equal to the valence and conduction band bending, and causes barriers for charge carriers, called Schottky barriers.

A large barrier height causes depletion of majority charge carriers in the semiconductor due to band bending, and the junction will show rectifying behavior. This type of junction is known as a Schottky diode. As the mobile charges of the metal can accumulate in a negligible depth at the interface, the Schottky diode is comparable to a strongly asymmetric pn-junction, and the depletion region and its capacitance will be governed by the doping of the semiconductor.

If the Fermi level alignment does not cause depletion, the junction will show ohmic behavior. Ohmic contacts are essential in order to connect semiconductor devices to external circuitry, and as such, it’s very important to be able to make good ohmic contacts. By using a highly doped semiconductor, the metal-semiconductor junction can show ohmic behavior as well, because the depletion region is thin enough for electrons to tunnel through.

According to the Schottky-Mott model, the Schottky barrier height is given by $\Phi_{bn} = \phi - \chi$ for electrons for a metal on an n-type semiconductor. Likewise, the barrier height for holes in a Schottky contact on a p-type semiconductor becomes $\Phi_{bp} = \chi - \phi + E_g$. However, this model ignores the appearance of electron states in the band gap at the interface [12], and deviations from the model often occurs in real systems. This effect is called Fermi level pinning, and reduces the effect of the metal’s work function on the barrier height. The severity of the effect depends on the interface between the metal and the semiconductor.

As in a pn-junction, when a bias voltage is applied across the Schottky
contact, the band structure and depletion zone shifts, and the current can be described with the diode equation (2.34). The difference between the junctions is in the reverse saturation current $I_0$. The Schottky diode is a majority carrier device, so there is no reverse drift of minority carriers across the junction. Instead, the reverse saturation current arises from thermally excitation of majority carriers across the barrier, described by the thermionic emission theory:

$$I_0 = A^* AT^2 e^{-\frac{\phi_b}{kT}}$$

(2.39)

Here, $A$ is the area of the diode, and $A^*$ is the Richardson constant, given by:

$$A^* = \frac{4\pi qm^*k^2}{\hbar^3} = \left(\frac{120}{cm^2K^2}\right) \frac{m^*}{m_0}$$

(2.40)

This constant is determined by the effective mass $m^*$ of the charge carriers in the semiconductor. Other conduction mechanisms may be present in real devices, and may accompany the thermionic emission. One of these is tunneling of electrons through the Schottky barrier, if the barrier is sufficiently narrow.

### 2.4.2 Heterojunctions

While the pn-junctions discussed in section 2.3 are made of a single material, it is also possible to make a junction where the n and p sides are made from different materials. This is done by epitaxially depositing one of the junction materials on top of the other, creating the junction at the interface between the two materials.

When creating the band diagram of the heterojunction, the difference in electron affinities and band gap sizes between the two semiconductors is considered. These two factors lead to different alignments of the conduction band ($\Delta E_C = q(\chi_1 - \chi_2)$) and the valence band ($\Delta E_V = \Delta E_g - \Delta E_C$). Depending on the band alignment, a heterojunction is classified as one of three types, as shown in Fig. 6: Type-I (straddling) heterojunction, Type-II (staggered) heterojunction or Type-III (broken-gap) heterojunction.

One important property of heterojunctions is the presence of one or two charge carrier barriers, caused by $\Delta E_C$ for electrons and $\Delta E_V$ for holes. This makes it possible for heterojunctions to have very different properties for electrons and holes. Utilizing this effect is called band gap engineering, and can be used to tailor the band structure to suit specific needs.

As for Schottky diodes, the actual band structure in a heterojunction is also determined by the electronic states at the interface. Fermi level pinning, which depends on both the interface quality and material properties, can cause significant deviations from the theoretical model.
2.5 The solar cell

Solar cells are often divided into three categories, called ‘generations’, based on which technology is utilized. [13] Nearly all are based on pn-junctions in semiconductors, where the junction acts as the active part of the solar cell. The ‘first generation’ solar cells are based on silicon wafers, split into mono-crystalline and multi-crystalline silicon. Together, these two account for about 90% of the total photovoltaic energy production. [2] The remaining share is dominated by the ‘second generation’ thin-film solar cells. By using thin films of various semiconductors, the material costs of the cells can be drastically reduced, with a smaller loss of efficiency.

Both the first and second generation of solar cells are based on a single junction, and their conversion efficiency has a theoretical upper limit of $\sim 30\%$, called the Shockley-Queisser limit. [14] The ‘third generation’ of photovoltaic technology, still mostly on the research state, attempts to increase performance beyond this limit. In addition, non-semiconductor photovoltaic devices are considered third generation solar cells. The Shockley-Queisser limit and third generation photovoltaic technology will be discussed in subsection 2.5.2.

This section is based on the works of Nelson [6] and Green [13] unless otherwise stated.

2.5.1 Operating principles

When photons with energy higher than the band gap hits the solar cell, electrons are excited from the valence band into the conduction band, creating electron-hole pairs. If these pairs reach the depletion region before they can recombine, they are swept across the junction by the electric field, generating a reverse current in the cell.

Two important parameters of the solar cell is identified from manipulating the external circuit to each extreme. With an open circuit, there is no possibility for a net current, and a potential called the open circuit voltage
V_{OC} builds up across the cell. By short-circuiting the solar cell, there is no voltage build-up, and the photo-generated current flows with no resistance. This is called the short-circuit current $I_{SC}$, and can be expressed as

$$I_{SC} = qA \int b_s(E)QE(E)dE$$

(2.41)

where $q$ is the elementary charge, $A$ is the surface area of the solar cell, and $b_s(E)$ is the flux density of photons with energy $E$. $QE(E)$ is the quantum efficiency of the solar cell, which is the ratio of incident photons that contribute an electron to the external circuit. This integral is made over the whole solar spectrum. As this integral is highly dependent on the characteristics of the illumination, standard solar test conditions have been defined in order to compare different solar cells. The standard conditions involve a solar cell at a temperature of 25°C, illuminated by incident light at an irradiation power of 1 kW/m$^2$ with a spectrum composition corresponding to AM 1.5. The air mass parameter describes the spectrum of the incident light after it has passed through the atmosphere. AM0 is the spectrum of the light exiting the sun. AM1 corresponds to the light hitting the earth when the sun is directly overhead, while AM1.5 corresponds to the sun at an elevation of 42°, increasing the travel length for the photons through the atmosphere to 1.5 times of the atmosphere thickness. In Figure 7 are shown the spectra for AM0 and AM1.5 compared to the radiation from a black body at the sun’s surface temperature, 6000K.

In order to harvest energy from the solar cell, an external load has to be applied. This creates a potential over the cell between the two extremes described above, and gives rise to a dark current given by the diode equation 2.34:

$$I_{dark}(V) = I_0 \left( e^{\frac{qV}{nkT}} - 1 \right)$$

(2.42)

The dark current flows in the opposite direction of the photo-generated current, and as such, the total current is given by the difference between them:

$$I_{total}(V) = I_{dark}(V) - I_{SC}$$

(2.43)

The relationship between the voltage and photocurrent from a solar cell is shown in Figure 8. The maximum power point $P_{max}$ is the highest effect the cell can deliver at the given illumination, and is given as:

$$P_{max} = I_{mp}V_{mp}$$

(2.44)

The fill factor $FF$ is defined as the relation between $P_{max}$ and the ultimate operation point ($V_{OC}$ and $I_{SC}$):

$$FF = \frac{I_{mp}V_{mp}}{I_{SC}V_{OC}}$$

(2.45)
and describes the ‘squareness’ of the IV curve. The fill factor should be as close to unity as possible. The efficiency $\eta$ of the solar cell encompasses all the loss mechanisms, and is defined as the ratio of the maximum power output to the power of the incident light $P_s$:

$$\eta = \frac{I_{mp}V_{mp}}{P_s}$$

(2.46)

2.5.2 Theoretical efficiency limit and third-generation solar cells

Conventional single-junction solar cells have several loss mechanisms, severely limiting their energy conversion efficiency.

The most limiting loss mechanism is the spectral losses. In order to create an electron-hole pair in a silicon based solar cell, a photon needs to have an energy at or above 1.11 eV, corresponding to a wavelength of 1100 nm or lower. This means that all photons with lower energy than 1.11 eV will pass through the cell without creating electron-hole pairs. The majority of the energy in the solar spectrum is located in the lower-wavelength section of the spectrum. These photons have enough energy to excite electrons into the conduction band; however, the excess energy from the photon will excite the electron further into the conduction band. From here, it will quickly relax to the band edge, heating the lattice in the process. Every photon will only contribute one electron-hole pair to the circuit at 1.11 eV. The
photons with a wavelength longer than 1100 nm don’t have enough energy to excite electrons across the band gap, and will either penetrate the cell, or contribute to heating.

Other loss mechanisms include shading from the front contacts, voltage losses in the junction and at the contacts, and recombination of electron-hole pairs in the semiconductor. These mechanisms are less limiting, however. The spectral losses alone, other loss mechanisms not accounted for, limit the conversion efficiency of a cell to about 44% [13], a little more than the Shockley-Queisser limit, as mentioned at the beginning of this section.

As the spectral losses present the most limiting loss mechanism in solar energy, this is what the third generation photovoltaics is trying to circumvent. There are several approaches to use more of the solar spectrum for energy conversion. These include among others creating multiple electron-hole pairs from high-energy photons, hot carrier harvesting and multiband cells through quantum wells. However, the simplest and as of now most used third generation type of solar cell is the tandem cell. In this concept, multiple cells with different band gaps are stacked on top of each other, with the band gap decreasing from the top. As the sunlight passes through the stack, each cell converts a range of photon energies close to its band gap.

While the highest theoretical output comes from independently operating each cell in the stack, this increases the complexity of the cell module significantly. Instead, the cells are usually designed connected in series. This simplification comes at the cost of efficiency. The current output from each cell has to match the others, or it will drain power from the circuit rather
than supplying it.

Another design consideration is lattice matching. A lattice mismatch between the layers in a tandem solar cell will increase the number of electronic states at the interface, which can reduce the conductivity in the structure.
Part II
Processing and characterization methods

3 Synthesis methods

3.1 Plasma in semiconductor synthesis

A plasma is a partially ionized gas, and is involved in several processes in the semiconductor field, as well as in regular fluorescent lightning. To understand two of the synthesis processes used here, magnetron sputtering and plasma-enhanced chemical vapour deposition (PECVD), an understanding of the properties of plasma is needed.

Plasma occurs naturally at very high temperatures. At lower temperatures, plasma can also be induced by an electric field. A plasma produced this way is called a glow discharge. This name originates from the characteristic glow of this plasma, caused by the relaxation of atoms that have been excited, but not completely ionized by inelastic collisions with free electrons.

A plasma is usually initiated by applying a high voltage over a low-pressure gas. This initiation voltage is several hundred volts, depending on gas pressure and spacing between the electrodes. When the electrical field in the plasma reactor exceeds the break-down field of the gas, a high voltage arc flashes between the electrodes, creating a large number of ions and free electrons. The ions and electrons are accelerated by the electric field towards the cathode and the anode, respectively. When the ions strike the cathode, secondary electrons are freed from the cathode material and accelerated towards the anode. Collisions between electrons and neutral atoms in the gas cause further ionization, freeing more electrons and creating more ions. If the electric field is strong enough, these two processes sustain the plasma.

Due to the movement of ions and electrons in the reactor, several different regions are formed between the electrodes. These are shown in Figure 9. Near the cathode, the electron density is much lower than the ion density, as the electrons are rapidly accelerated away from the cathode. In this area, very few atoms are excited by inelastic collisions, and a dark space is formed, called Crooke’s dark space. As the distance from the cathode increases, so do the kinetic energies of the accelerating free electrons, and excitation of atoms happen, creating a glowing space. Further in, another dark space is formed, as the majority of the electrons have been accelerated to very high energies (above 15 eV), causing ionization rather than excitation. This region is called the Faraday dark space. Finally, the region close to the anode has a very low electron density, as the anode acts as an electron sink. This region is called the anode dark space.
Figure 9: The different regions of an electrically induced plasma

In processing, wafers or other materials are placed on top of the cathode. In Crooke’s dark space, there is a large electric field, which accelerates ions in this region rapidly toward the cathode, bombarding the material. The energy provided by this ion bombardment is used to drive various processes.

3.2 Magnetron sputtering

Sputtering is a process where particles are fired at high energies toward a solid target material, causing atoms or molecules to be ejected from it. This effect has three main applications: etching, analysis and deposition.

The target material is etched away during sputtering. This is used in ion milling and reactive ion etching (RIE). These methods have two main advantages, anisotropy and low selectivity, as opposed to wet chemical etches, which often are very isotropic and have high selectivity. By analysing the ejected atoms from the target, a qualitative and quantitative depth profile can be achieved from the sample. This is done in secondary ion mass spectrometry (SIMS), which can measure chemical concentrations down to ppb (parts per billion) levels. The third main application of sputtering is sputter deposition, where the ejected material from the target is deposited as a thin film on a substrate. This application is one of the deposition techniques used in the sample synthesis in this work.

Sputter deposition offers several advantages, which makes it widely used in both the semiconductor industry, in optics, and production of digital storage media. Material can be deposited at a low substrate temperature, with good adhesion. Compared to e-beam evaporation (described in section 3.4), sputter deposition inflicts little radiation damage. In addition, it is better suited to deposit compounds and alloys. It consists of a parallel-plate plasma reactor in a vacuum chamber, arranged so that high energy ions strike the target material. To ensure that as many of the ejected atoms as possible are deposited on the substrate, the cathode and anode are closely spaced.

The plasma is usually initiated from an inert gas at low pressure, to avoid chemical reactions with the target material or the substrate. In some cases, reactive gases are used, where the deposited material is produced through
a reaction between the gas and the sputtered material. For inert gases, the most efficient sputter action occurs when the mass of the plasma ions match the mass of the atoms in the target material.

In the plasma, the ions have a wide range of energies, and depending on the energy of the ions striking the target, one of four things happen. Low-energy ions, typically with energies below 10 eV, will usually be reflected, or adsorb to the surface. At very high energies, usually above 10 keV, ions will penetrate into the bulk of the material, causing structural damage. Between these two extremes, the ions will penetrate a few atomic layers into the material. Most of the energy from the ions will be transferred to atoms at or near the surface, causing substrate atoms or clusters to be knocked out from the surface.

If the target and substrate materials are conducting, the plasma can be sustained with a simple direct current. However, if one or both of the materials are insulating, a charge will build up on that electrode and kill the plasma. In that case, the polarity of the potential is alternated in order to avoid the charge build-up. This is usually done with a radio frequency potential, earning the name RF sputtering. When the electric field is alternating at radio frequency, the light electrons respond instantaneously to the alternations, avoiding build-up of charge. The heavier ions’ paths do not change with the radio frequency. To direct the ions toward the target while maintaining an alternating field, a direct potential difference is superimposed between the two electrodes.

### 3.3 Plasma-enhanced chemical vapour deposition

Plasma-Enhanced Chemical Vapour Deposition (PECVD) is a method for thin film deposition used at low temperatures. It is a subclass of chemical vapour deposition (CVD), which utilizes chemical reactions of precursor gases to produce the film.

With some variations to the specific steps, all CVD processes share a core process. The reaction precursors are transported in a gas flow, often diluted in an inert gas, into the reaction chamber, where the substrate is located. Once inside, the precursors adsorb onto the substrate and diffuse along it to growth sites, where they react and form the film. By-products of the reaction desorb from the surface of the film and are swept away by the gas flow.

For most CVD processes, the energy required for the precursors to react is provided by heating the substrate. In PECVD, a plasma is used to provide this energy instead, making the process possible at lower temperatures. As many of the materials deposited by PECVD are insulating, RF plasmas are typically used for these depositions. The high kinetic energy of the free electrons in the plasma causes dissociation of precursor molecules into highly reactive radicals, which then can react with the substrate. In addition, the
films are exposed to ion bombardment from the plasma. These ions can provide energy to the adsorbed species, allowing them to diffuse further along the surface. Because of this, the PECVD process has a good step coverage and yields high density films.

3.4 Contact deposition

Deposition of contact metals on samples is often done by evaporation, and evaporation of atoms from a melted material gives an equilibrium vapour pressure which largely depends on the temperature.

There are three commonly used heating systems for evaporation, each with its own benefits and drawbacks. The simplest type is a resistively heated system, which usually consists of either a small coil of wire, or a dimpled boat made from a resistive material. The charge material is placed as a solid bar on the wire, or in the boat, which is then heated by a controlled current. A major drawback with resistive heating is the fact that the temperature of the filament must be at least as high as the material to be evaporated. This leads to evaporation and outgassing from the filament itself. For charge materials that require high temperatures to evaporate, such as refractory metals, the temperature becomes too high for most resistive heating elements.

In order to heat the charge to higher temperatures, it can instead be placed in a crucible wrapped in a metal coil. RF power is run through the coil, and this induces eddy currents in the charge, heating it. By water-cooling the coil, it can be kept at a low enough temperature to eliminate material loss from it. However, as the crucible is being heated as well, contamination of the charge from the crucible is still significant. The third evaporation system avoids contamination from the crucible by heating only
Figure 11: An e-beam evaporation setup. Note that the magnet has to be much larger than shown here. From Campbell [7].

the charge. In order to accomplish this, an electron beam (e-beam) is focused at the charge. The beam is provided by a simple electron gun, typically a tungsten wire, where electrons are boiled off the wire by high biasing and current flow. The electron gun is often located under the crucible to reduce contamination from the filament wire, and the electron beam is directed by a magnetic field. The electron beam is focused on a small area on the charge material, and the crucible itself can in some systems be cooled, to minimize contamination. Such a setup can be seen in Figure 11. The biggest issue with e-beam evaporation is radiation damage of the substrate. Highly excited electrons in the charge material emit high-energy radiation when they relax to core levels. Because of this, e-beam evaporators cannot be used on certain sensitive devices unless the damage can be removed later by thermal annealing.
4 Sample preparation

This chapter describes the production details and parameters for the samples made and used in this work. A schematic of the finished device structures is shown in Figure 12, and a block diagram providing the steps of the production process is shown in Figure 13. As the figures indicate, two different structure types were made, with buffer deposition on one or both sides. In both structures, both n-type and p-type silicon was used. The buffer layer of amorphous silicon was deposited at three different temperatures, and for each temperature, the buffer was deposited at three different thicknesses. In addition, samples without buffer layers were made, for a total of 40 different samples.

The substrates used in this work were double-side polished silicon wafers of n- and p-type with (1 0 0) orientation. Both wafer types had resistivity in the range 1-10Ωcm. The wafers were cut to appropriate sample sizes (10x10 mm² for single side structures and 30x30 mm² for double side structures) using a Rofin scribing laser. The Rofin uses a 532 nm wavelength laser, and for the wafer cutting, the power was set to 32A at 50kHz, the laser moved at 400 mm/sec, executing each movement 50 times. In addition, a shadow mask was cut for contact deposition, using other laser parameters: 26A at 15kHz, a movement speed of 100mm/sec and 500 executions. The shadow mask had a repeating pattern of four different circular holes, with designed diameters of 300, 450, 600 and 750 µm.

After cutting, the samples underwent a standard RCA clean. Prior to deposition, all samples were cleaned in a weak solution of hydrofluoric acid and rinsed in water in order to strip off any naturally grown oxide.

The buffer layer of amorphous silicon was deposited using a 310 MKII PECVD system from Advanced Vacuum. The amorphous layer was deposited at three different substrate temperatures, 180°C, 200°C and 220°C,
Table 1: Deposition times in minutes and seconds used for PECVD of amorphous silicon buffer layers at the given temperatures.

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Estimated film thickness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 nm</td>
</tr>
<tr>
<td>180°C</td>
<td>0:37</td>
</tr>
<tr>
<td>200°C</td>
<td>0:26</td>
</tr>
<tr>
<td>220°C</td>
<td>0:38</td>
</tr>
</tbody>
</table>

and for each temperature, it was deposited with thickness of approximately 5, 10 and 20 nm. One sample of each doping type and structure was left undeposited. The deposition was performed with an RF power of 50 W and a chamber pressure of 600 mtorr, using a precursor gas consisting of 2% silane in nitrogen with a flow rate of 2000 sccm (standard cubic centimeters per minute, cm³/min). In order to calculate the deposition rate for each temperature, samples were partially covered by smaller silicon samples, making a step between the bare wafer and the deposited amorphous silicon. Using a Veeco Dektak 8 stylus profilometer, the thickness of the deposited film could be determined by measuring the step heights on the samples for the various deposition times. The deposition times calculated from the determined rate and used for the depositions are shown in Table 1.

The 3x3 cm² samples for the double-sided structures were flipped after the first deposition, and got an identical buffer layer deposited on the other side. These samples were to be used for QSSPC, and an undeposited back side would have led to a high surface recombination and misleading results.

The 1x1 cm² samples were collected after all had completed PECVD process and loaded into a Semicore Triaxis magnetron sputter, along with samples without the buffer layer. The sputtering was performed with a chamber pressure of 1.2×10⁻⁶ torr. Pre-sputtering was performed with closed shutters in order to remove any contamination or unwanted oxide from the target surface. The sample holder was heated to 400°C and was rotated at about 10rpm. The aluminum doped zinc oxide (AZO) film was deposited by co-sputtering with one ZnO target RF operated at 50 W, and one aluminum target DC operated at 3W. With a deposition rate of ∼ 1.4nm/min, 70 minutes of deposition was used to produce a film approximately 100nm thick.

Ohmic front contacts were deposited onto the ZnO with e-beam evaporation, through the shadow mask mentioned in earlier in the chapter. A Leybold L560 setup was used, and the contacts were deposited at a pressure of 10⁻⁶ mbar. Due to observed contamination of the nickel source, a thicker layer of aluminum was deposited instead of the planned structure (10nm Al and 100nm Ni). The aluminum layer was deposited at a rate of 0.1nm/s to approximately 170nm, measured by the built-in thickness meter.
Figure 13: Block diagram showing the sample preparation process steps.
After contact deposition, the zinc oxide around the contacts was etched by a buffered oxide etch, consisting of a mixture of $\sim 5\%$ phosphoric acid and $\sim 5\%$ acetic acid in water. During visual inspection of the samples with a microscope, leftover oxide were found on several samples, and a new etch was performed on these. The lateral etching of the oxide covered by the contacts is considered negligible in this process.
5 Characterization methods

5.1 Profilometry

A profilometer is a measuring instrument used to measure the profile of a surface, in order to quantify its roughness and measure its topography. Both contact and non-contact profilometers exist, but in this work only a contact method was used.

In a contact profilometer, a stylus is lowered onto the surface of a sample, and then moved across the surface a specified distance, pushing down with a specified force. The vertical movement of the tip is tracked electrically by the profilometer, and a two-dimensional profile graph is acquired. This profile graph can be used to determine step heights, and both width and height/depth of peaks and craters in the structure. [16]

In order to get a more complete assessment of the surface, the profilometer can be used to do a three-dimensional mapping as well. This is achieved by a raster scan, where a high number of closely spaced parallel profile traces are combined to make a three-dimensional map of the surface. [17]

5.2 Deep-level transient spectroscopy (DLTS)

The deep-level transient spectroscopy (DLTS) technique was introduced in 1974 by David Vern Lang, and is a non-destructive technique used to study electrically active defects in semiconductors. These defects cause energetically "deep" energy states in the semiconductor band gap, and are known as charge carrier traps. [18] These deep level states are discussed in section 2.2.4.

5.2.1 Principle of operation

The variation in trap occupancy over time \( n_t(t) \) can be found by solving equation (2.22), with an initial concentration \( n_t(t = 0) = n_t(0) \):

\[
n_t(t) = \frac{a}{a + b} N_t - \left( \frac{a}{a + b} N_t - n_t(0) \right) e^{-(a+b)t} \tag{5.1}
\]

This can also be considered an exponential relaxation from the initial concentration. The boundary at infinite time is a steady state concentration of \( n_t(\infty) = \frac{a}{a + b} N_t \). Substituting in the above equation gives

\[
n_t(t) = n_t(\infty) - (n_t(\infty) - n_t(0)) e^{-t/\tau} \tag{5.2}
\]

Here, \( \tau \) is the time constant for the relaxation, and is defined by \( \tau = (a + b)^{-1} = (e_n + c_n + e_p + c_p)^{-1} \). The direction of the relaxation depends on whether the initial concentration \( n_t(0) \) is larger or smaller than the steady
state concentration $n_t(\infty)$. For the two extreme initial conditions, full traps and empty traps, simplifications to the above expression can be made:

\begin{align}
  n_t(0) &= 0 & n_t(t) &= n_t(\infty) \left(1 - e^{-\frac{t}{\tau}}\right) \\
  n_t(0) &= N_t & n_t(t) &= \frac{a}{a+b} N_t + \frac{b}{a+b} N_t e^{-\frac{t}{\tau}}
\end{align}

(5.3) (5.4)

These expressions assume sharp edges at the depletion region. In reality, there are transition regions of gradually changing occupancy at the depletion region edge.

As shown in section 2.2.4, the emission rate from a trap is a function of temperature. This function can be used to find the trap signature. It is typical to use an asymmetrical pn-junction or a Schottky diode when measuring. This is because the depletion region then mainly extends into one side of the junction, which simplifies calculations. In the entirety of this section, a $p^+n$-junction will be used as an example.

In a $p^+n$-junction, the studied states are electron traps in the lightly doped n-type material. As this is a majority carrier trap, it is reasonable to assume that its capture and emission rates are much higher than those of minority carriers, in this case $c_n \gg c_p$ and $e_n \gg e_p$. In addition, as there are no free charge carriers in the depletion region, the capture rate of majority carriers can be considered negligible, $c_n(N = 0) = 0$. Given this, the electron emission rate alone governs the rate of change of electron occupancy of the trap, and equation (2.11) is reduced to

\[ \frac{dn_t}{dt} = -e_n n_t \]

(5.5)

The electron traps are initially considered to be completely filled, so $n_t(0) = N_t$. This assumption is reasonable, as $c_n > e_n$ due to the positions of the trap level and the Fermi level, which implies filling of the electron traps when there is no bias sustaining the depletion region. Taking this into consideration, the above equation integrates to

\[ n_t(t) = N_t e^{-e_n(T) n_t} \]

(5.6)

The change in occupancy of the traps is measured by the depletion capacitance. As described in section 2.3.3, the capacitance of a $p^+n$-junction is given by equation (2.38). In this case, the charge of the carrier traps must also be included, and the charge density $\rho = q N_d$ must be replaced to reflect this. An acceptor-like deep level is neutral when unoccupied, and becomes negatively charged when occupied by an electron, so the total charge density can be expressed as $\rho(t) = q(N_d - n_t(t))$. Unoccupied donor-like levels, conversely, are positive, and become neutral when occupied, giving the total charge density $\rho(t) = q(N_d + N_t - n_t(t))$. Donor-like traps are assumed in
the following derivation. Replacing the charge density in equation (2.38) gives

$$C(t) = C(\infty) \sqrt{1 - \frac{n_t(t)}{N_d + N_t}},$$

(5.7)

where $C(\infty)$ is the capacitance at reverse bias after all traps are emptied:

$$C(\infty) = A \sqrt{\varepsilon_r \varepsilon_0 q (N_d + N_t)}$$

(5.8)

For trap concentrations much lower than the doping concentration ($n_t, N_t \ll N_d$), the change in capacity $\Delta C(t) = C(\infty) - C(t)$ can be expressed as:

$$\Delta C(t) = C(\infty) - C(\infty) \left(1 - \frac{n_t(t)}{N_d}\right)^{\frac{1}{2}}$$

(5.9)

By using the approximation $(1 - x)^a \approx (1 - ax)$, which is valid for very small values of $x$, the relative change in capacity becomes

$$\frac{\Delta C(t)}{C(\infty)} = \frac{n_t(t)}{2N_d},$$

(5.10)

and finally, by inserting equation (5.6):

$$\frac{\Delta C(t)}{C(\infty)} = \frac{N_t}{2N_d} e^{-e_n(T)t}$$

(5.11)

This relation is the basis of the DLTS method.

### 5.2.2 Signal generation

There are two cycled steps making up the DLTS method. First, the sample is kept at a constant reverse bias in order to keep trap states in the depletion zone empty. A majority carrier pulse reduces the bias to zero (or close), filling the depletion zone with free carriers. If the pulse is long enough, all trap levels will be filled, and the initial condition assumption $n_t(0) = N_t$ will be satisfied.

After the pulse, the sample returns to the previous reverse bias and the depletion zone is again emptied. The electron traps will begin emitting electrons, creating a variation in $n_t$ and thus $C$ with time. As the electrons are emitted, the net positive charge density increases in the depletion region. Because of this, the depletion region contracts, leading to an increase in the depletion capacitance. As the traps are emptied, the capacitance returns toward the reverse bias capacitance $C(\infty)$.

The depth of the investigated material can be controlled by varying the values of the reverse bias and the pulse voltages. In order to trap carriers in
interface defects, a forward bias pulse is used to reduce the depletion region from the built in voltage. The defects in the bulk of the material can be investigated by using a zero or small reverse bias during the filling sequence. By running multiple scans over various pulse bias voltages, it is possible to make a profile of the defect distribution versus the depth of the material.

The DLTS signal, in its simplest form, can be created from two capacitance measurements separated in time. By assigning a negative sign to one of the measurements, the sum becomes the change in capacitance in the sample between the two measurements. The essential feature of DLTS is the ability to set up a rate window, so that the measuring apparatus gives an output only when the thermal emission rate is within the window. If the sample temperature is increased or reduced at a constant rate, the emission rate of carriers from defect centers present in it will vary, and the measuring instrument will give a response peak whenever the emission rate from the defect center is within the rate window. An important property of the rate window is that the output is proportional to the amplitude of the transient. By combining transients at different temperatures, the DLTS signal is formed, as shown in figure 15.

From the peak position of the signal, the carrier emission rate at that specific temperature can be calculated. If a different rate window is used, the peak position will shift, and the emission rate at a different temperature is found. By using several different rate windows, the temperature dependence of the emission rate can be mapped. In this work, six different rate windows are used. The signal from the $i$-th window is defined as

$$S_i(T) = \frac{1}{n_i} \sum_{t_j=t_d}^{t_d+t_i} \Delta C(T, t_j) \omega(t_j)$$  \hspace{1cm} (5.12)$$

where $n_i$ is the number of measurements done in each rate window and $t_i$ is the duration of it. During the trap filling pulse, when the depletion
Figure 15: Capacitance transients produced at different temperatures. A peak in the signal occurs when the time window $t_2 - t_1$ matches the emission rate. For different time windows, the peaks occur at different temperatures. Adapted from Lang [18].

zone is quenched, the capacitance meter might overflow, resulting in errors. To avoid this, a delay time $t_d$ is added between the pulse bias and the first capacitance measurement. $\Delta C(T,t)$ is the capacitance change caused by emission from the traps as given by equation (5.11), and $\omega(t)$ is the incorporated weighting function.

The simplest weighting function is adding a negative sign to one of the measurements, as mentioned earlier. However, more sophisticated weighting functions are often used. The choice of weighting function affects the sensitivity, the signal-to-noise ratio, and the selectivity; that is, the ability to distinguish two different close peaks. By comparing more than two capacitance measurements, more of the transient is utilized, creating a more accurate signal. The weighting function used in this work is called lock-in weighting, and is defined as:

$$\omega(t) = \begin{cases} 1 & t_d + 2^{i-1}\tau < t \leq t_d + 2^i\tau \\ -1 & t_d < t \leq t_d + 2^{i-1}\tau \end{cases}$$

(5.13)

The number of measurements required for the lock-in weighting function is $n_i = 2^i$ for the $i$th rate window. $\tau$ represents here the time interval between each of the measurements. As can be seen, the lock-in weighting function simply assigns a negative sign to the first half of the capacitance measurements, and a positive sign to the second half. [19]

5.2.3 DLTS signal analysis

At the peak of the DLTS signal for each window, its derivative with respect to the temperature must be zero. This can also be seen from Figure 15. By
Table 2: Details for the six different rate windows used in a lock-in DLTS measurement

<table>
<thead>
<tr>
<th>Window number</th>
<th>Number of measurements $2^i$</th>
<th>Window length $t_i$ [ms]</th>
<th>$e_n t_i$</th>
<th>$F_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>20</td>
<td>1.448962</td>
<td>0.12488</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>40</td>
<td>1.81717</td>
<td>0.15490</td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>80</td>
<td>2.09799</td>
<td>0.17597</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>160</td>
<td>2.28229</td>
<td>0.18880</td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>320</td>
<td>2.39056</td>
<td>0.19594</td>
</tr>
<tr>
<td>6</td>
<td>64</td>
<td>640</td>
<td>2.44976</td>
<td>0.19971</td>
</tr>
</tbody>
</table>

splitting the variables, this gives the following expression:

$$\frac{dS(T)}{dT} = \frac{dS}{d(e_n t_i)} \frac{d(e_n t_i)}{dT} = 0$$  \hspace{2cm} (5.14)

The last factor in this expression is only zero at absolute zero temperature, and can thus be disregarded, leaving:

$$\frac{dS}{d(e_n t_i)} = 0, \quad T > 0$$  \hspace{2cm} (5.15)

From this, it can be seen that the zero-point of the derivative comes from a specific combination of the rate window and the emission rate. The value of $e_n t_i$ can be numerically calculated for each window length $t_i$ by solving equation (5.12) using the lock-in weighting function. Values for a 5 ms delay time and 20 ms minimum window length are given in Table 2. By reading out the temperature corresponding to a peak in the DLTS signal from each different rate window, the relation between temperature and emission rate can be calculated.

As shown in section 2.2.4, the emission rates for the temperatures corresponding to the peaks in different rate windows can be set up in an Arrhenius plot, as given by equation (2.23):

$$\ln(e_n/T^2) = -\frac{\Delta H}{kT} + \ln(\sigma_{nA}\gamma)$$

As described in section 2.2.4, $\gamma$ is a material- and charge carrier property, and for electrons and holes in silicon the values are $\gamma_n = 3.58 \times 10^{21} s^{-1} K^{-2} m^{-2}$ and $\gamma_p = 1.82 \times 10^{21} s^{-1} K^{-2} m^{-2}$, respectively. From a linear regression of the six points in the plot, the enthalpy of formation $\Delta H$ and the apparent cross section $\sigma_{nA}$ of the trap state can be found from the slope of the line and its intercept with the $e_n/T^2$-axis, respectively.

The trap concentration can also be determined from the DLTS signal. By inserting equation (5.11) into equation (5.12), the following expression
is acquired:

\[
S_i(T) = \frac{C(\infty)N_i}{2N_d} \left\{ \frac{1}{n_i} \sum_{t_j=t_0}^{t_0+t_i} e^{-e_n(T)t_0(t_j)} \right\} = \Delta C_0 F_i \tag{5.16}
\]

The expression in brackets is a numerical term \( F_i \) which is constant for a given time window. The values for the parameters used in this work are reported in Table 2. From this, the trap state concentration is given by

\[
N_t = \frac{2N_d S_{i,p}(T_p)}{C_p(\infty)F_i} \tag{5.17}
\]

The subscript \( p \) indicates values at the temperature corresponding to a signal peak.

### 5.3 Current-voltage measurement

One very useful characteristic for many semiconductor devices is the current through the device as a function of the applied voltage. Important figures of merit in this regard are ideality factor, barrier height and series resistance.

The current-voltage relationship is described by equations (2.33) and (2.38):

\[
I(V) = I_0 \left( e^{\frac{qV}{\eta kT}} - 1 \right) = A^* A T^2 e^{\frac{\phi_{IV,na}}{kT}} \left( e^{\frac{qV}{\eta kT}} - 1 \right) \tag{5.18}
\]

The superscript \( IV \) is here used to separate the IV-deduced barrier values from those deduced from CV measurements. The Schottky barrier height is related to the Fermi level of the sample for both n-type and p-type substrate, so the valence band of ZnO is neglected.

The ideality factor can be found from equation (5.18). For bias voltages \( V > \frac{2kT}{q} (\approx 0.08V \text{ at } 300K) \), this can be simplified to \( I(V) = I_0 e^{\frac{qV}{\eta kT}} \), as the exponential term will dominate. Taking the logarithm yields the linear function

\[
\ln I(V) = \ln I_0 + \frac{qV}{\eta kT} \tag{5.19}
\]

Rearranging this, the ideality factor can be found from the slope of the given plot:

\[
\frac{\partial}{\partial V} (\ln I(V)) = \frac{\partial}{\partial V} \left( \ln I_0 + \frac{qV}{\eta kT} \right)
\]

\[
\frac{1}{\eta} = \frac{kT}{q} \frac{\partial}{\partial V} (\ln I(V)) \tag{5.20}
\]

It is important to note that the linearity of the plot is lost at higher bias, as the series resistance becomes dominating, so the ideality factor must be found from the data in the region with linear behavior.
The barrier height can also be found from this plot. $I_0$ is found from extrapolating the linear fit to $V = 0$, and $\phi_{bn,bp}^{IV}$ is calculated from equation (2.39):

$$\phi_{bn,bp}^{IV} = kT \ln \frac{A^*AT^2}{I_0}$$  (5.21)

Rectification is a measure of the difference between the current in forward and reverse bias of a diode. In many applications, a high rectification is desired. As shown in this section, the current is limited by the series resistance in forward bias. During reverse bias, deviations from the ideal diode model arise from interface defects or from conduction mechanisms other than thermionic emission.

### 5.4 C-V measurement

In an asymmetrical pn-junction or a Schottky junction, the depletion capacitance is given by equation (2.38):

$$C = A \left( \frac{\varepsilon q N_{d,a}}{2V_j} \right)^{\frac{1}{2}} = A \left( \frac{\varepsilon q N_{d,a}}{2(V_0 - V_{ext})} \right)^{\frac{1}{2}},$$

where $N_{d,a}$ indicates the doping concentration in the lowest doped semiconductor. From measurements of the capacitance as a function of applied reverse bias $V_{ext}$, both the built-in potential $V_0$ and the doping concentration of the low doped side can be calculated. This is typically done by plotting $1/C^2$ versus $V$:

$$\frac{1}{C^2} = -\frac{2}{A^2\varepsilon q N_{d,a}} V_{ext} + \frac{2}{A^2\varepsilon q N_{d,a}} V_0 = a V_{ext} + b$$  (5.22)

The slope of the plot $a = \frac{2}{A^2\varepsilon q N_{d,a}}$ contains the doping concentration, and the built-in potential can be found by extrapolating the line to $1/C^2 = 0$, where $V_{ext} = V_0$. To account for the majority carrier (Debye) tail extending into the depletion region, the term $\frac{kT}{q}$ is added to the intercept value.

By treating the junction between Al-doped ZnO and silicon as a Schottky contact, the barrier height can be calculated from the CV data as well. The barrier height is given as the sum of the built-in potential, the separation between the Fermi level and the majority carrier band, and the Debye tail-term:

$$\phi_{bn,bp}^{CV} = V_0 + \xi + \frac{kT}{q}$$  (5.23)

For n-type substrates, $\xi = E_C - E_F$, and $\xi = E_F - E_V$ for p-type substrates. $\xi$ is calculated from equations (2.7) through (2.9), and the carrier concentrations are assumed equal to the doping concentrations found from the $1/C^2$ plot. This gives, for n-type substrates:

$$n_0 = N_d = N_C e^{-\frac{E_C - E_F}{kT}} = N_C e^{-\frac{\xi}{kT}}$$  (5.24)
\[ \xi = -kT \ln \frac{N_d}{N_C} \quad (5.25) \]

5.5 Comparison of IV- and CV-derived barrier heights

Calculating the barrier height from both CV and IV measurements will measure different mechanisms, due to the inherent differences in the derivation of these values.

For the IV data, the barrier height is calculated from forward and reverse bias measurements. The result depends on the carrier transport across the junction, which makes it sensitive to defects or mechanisms not described by thermionic emission or the diode equation. Two of these mechanisms are inhomogeneous barrier height and image force lowering, both of which result in a lower measured barrier height. An inhomogeneous barrier height results in higher current flow in the regions with lower barriers. Image force lowering is the effect that a charge carrier approaching the interface from the semiconductor side will induce an opposite mirror charge in the metal (or degenerate semiconductor), which will attract the carrier, effectively lowering the barrier.

The CV measurements depend on build-up of charges rather than the transport of these across the interface, so neither of the above mechanisms will affect it.

5.6 Quasi-steady-state photoconductance

Quasi-steady-state photo conductance (QSSPC) is a method used to measure the lifetime of minority carriers in silicon [20] and a few other materials. It was developed by combining elements from two previously used optical lifetime measurement methods, namely photoconductance decay and the steady-state photoconductance method.

Photoconductance decay is a commonly used method based on generating charge carriers by optical excitation using a very short light pulse, and then monitoring their decay as a function of time. The effective lifetime \( \tau_{\text{eff}} \) is then obtained from the slope of the curve:

\[ \tau_{\text{eff}} = \Delta n \left( \frac{d(\Delta n)}{dt} \right)^{-1} \quad (5.26) \]

where \( \Delta n \) is the number of photogenerated carriers. With this method, the light pulse has to be significantly shorter than the effective lifetime in the wafer in order to get accurate results.

Alternately, the wafer can be exposed to steady-state illumination. The increased carrier density generated by the illumination leads to an increase in wafer conductance which can be expressed by:

\[ \sigma_L = q(\Delta n \mu_n + \Delta p \mu_p)W = q\Delta n(\mu_n + \mu_p)W \quad (5.27) \]
where $W$ is the wafer width. In the steady state, the rate of electron-hole pair generation must be equal to the recombination rate. By approximating $\Delta n$ to be uniform across the wafer, the effective lifetime can be expressed as

$$\tau_{eff} = \frac{\sigma_L}{J_{ph}(\mu_n + \mu_p)} \quad (5.28)$$

With this method, the photogeneration current density $J_{ph}$ and the electron and hole mobilities $\mu_{n,p}$ have to be acquired separately.

QSSPC incorporates both of the above methods by using a light pulse that varies very slowly compared to the effective lifetime of the wafer. This makes it possible to measure the photoconductance over a large range of illumination intensities in a very short time, and avoids the sample heating from steady-state illumination. In addition, very low lifetimes can be measured without needing short light pulses. The range of measurable lifetimes is only limited by signal strength, i.e. illumination.
Part III
Results and discussion

6 Results and discussion

The experimental results are presented and discussed in this chapter. First, the results from measurements of current (IV) and capacitance (CV) are presented in section 6.1, the results from the DLTS measurements in section 6.2, and finally the lifetime measurements in section 6.3.

For simplicity, the samples will be named as the deposition temperature and the thickness of the buffer layer deposited. For example, the sample with a 10 nm buffer layer deposited at 220°C will be referred to as 220-10nm. Whether the sample is deposited on p- or n-type silicon will be specified unless clear from the context.

6.1 Current and capacitance measurements

IV and CV measurements were carried out in room temperature for all single-side polished samples. For each sample, several diodes were measured, and the IV and CV measurements for each diode were performed by rewiring externally on the sample holder, without moving the probe connecting the sample to the equipment.

Variations between different diodes on the same sample was observed in many cases. In order to choose which diode to use for CV and DLTS measurements, the IV data from the different diodes were compared. The criteria for choosing a diode was high rectification and a low ideality factor, unless the diode in question deviated drastically from the others.

The data from the IV measurements are presented in semi-logarithmic plots, and are sorted by substrate type (n- and p-Si) and buffer deposition temperature. All plots can be seen in Figure 16, for n- and p-type substrates. Capacitance measurements are shown in Figure 17 for n-type samples and Figure 18 for p-type samples.

Two different values of the reverse saturation current $I_0$ can be obtained from the IV plot, depending on the bias voltage. The first value, here referred to as $I_{0}^{fwd}$, is calculated from the linear region at forward bias, from the intercept with the y-axis. The other value, referred to as $I_{0}^{rev}$ can be found from a readout of the current at $V = -3V$.

The calculated values of the ideality factor $\eta$, the built-in voltage and the barrier heights calculated from $I_{0}^{fwd}$, $I_{0}^{rev}$ and the CV measurements for each sample are summarized in Table 3 for p-type samples and in Table 4 for n-type samples.
Figure 16: Current vs. voltage plots for samples on a) n-type and b) p-type silicon substrates, at three different deposition temperatures. Each plot compares the different buffer thicknesses at each specific deposition temperature, which are, from top to bottom, 180°C, 200°C and 220°C.
Samples on n-type substrates

The IV measurements of the structures deposited on n-type substrates are shown in Figure 16a. As can be seen in Table 4, the buffered samples show a very low increase in ideality compared to the sample without buffer. The 180-20nm sample has the highest calculated ideality factor at $\eta = 2.46$, which is less than 10% higher than the ideality of the sample without a buffer.

Aside from the fact that the presence of a buffer layer lowers both the forward and reverse current in the structure, there is no clear correlation between the buffer thickness/temperature of deposition and either the ideality or rectification of the samples. However, when observing the barrier height calculated from the linear region at forward bias, two trends seem to surface. Firstly, the barrier is higher for thicker buffers at the same temperature. Secondly, as the temperature increases, the barrier height lowers. These trends ring true for all but the 180-20nm sample. The barrier heights from the reverse readout seem to be virtually unaffected by the buffer layer. The values vary by less than 5% from the unbuffered sample. It must be noted that the reverse current does not seem to be saturated at $V = -3V$ for any of the n-type samples.

The CV measurements, plotted in Figure 17, show an increasing capacitance with increasing buffer thickness. For two of the samples, 200-20nm and 220-10nm, the intersection of the $1/C^2$ line with the x-axis implied a negative built-in voltage, which is non-physical. This can sometimes be associated with a high conductance; however, no such significant deviation was observed from the CV data. Given that the capacitance of a pn-junction is proportional to the reciprocal of the depletion layer thickness, the perceived negative built-in voltage may be due to an underdeveloped depletion region compared to the theoretical model, which can come from parasitic capacitance or voltage loss on the sample away from the junction.

Samples on p-type substrates

The IV measurements of the structures on p-type substrates are shown in Figure 16b. For the three samples with 20 nm buffer, there is a significant increase in both the ideality factor and the reverse current, leading to very low rectification in all three samples. Contrary to this, the two samples with a 5 nm buffer deposited at 200°C and 220°C show a large enough drop in observed reverse current to achieve higher rectification than the unbuffered sample.

There is no clear trend to find in the CV data for the p-type samples, shown in Figure 18. The samples with buffer deposited at 180° show an increasing built-in voltage with increasing buffer thickness, while the samples deposited at 220° show the opposite relation. There is also a large variation in the built-in voltages found, from 0.61V on sample 200-20nm to 1.25V on
Figure 17: a) Capacitance vs. voltage and b) $1/C^{-2}$ vs. voltage plots for samples on n-type silicon substrates, at three different deposition temperatures. Each plot compares the different buffer thicknesses at each specific deposition temperature, which are, from top to bottom, 180°C, 200°C and 220°C.
Figure 18: a) Capacitance vs. voltage and b) $1/C^2$ vs. voltage plots for samples on p-type silicon substrates, at three different deposition temperatures. Each plot compares the different buffer thicknesses at each specific deposition temperature, which are, from top to bottom, 180°C, 200°C and 220°C.
Figure 19: Barrier heights calculated from IV and CV data for a) p-type and b) n-type samples, as seen in Tables 3 and 4. Note that the barrier values calculated from CV does not exist for the samples 200-20nm and 220-10nm, as non-physical values were found for the built-in voltage for these.

6.2 DLTS measurements

The presence of electrically active defects at a junction interface can have a negative impact on the device performance. Deep levels were discussed in section 2.2.4, and the DLTS method was explained in section 5.2. In this section, the results from the DLTS measurements are presented and
Table 3: Summary of properties of all p-type samples, calculated from CV and IV data. For the IV barrier heights, $I_0^{\text{fwd}}$ refers to using $I_0$ calculated from the linear forward region, while $I_0^{\text{rev}}$ uses the reverse current at $V = -3V$.

<table>
<thead>
<tr>
<th>Buffer characteristics</th>
<th>IV data</th>
<th>CV data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\eta$</td>
<td>$\phi_{bp}^{\text{IV}}(I_0^{\text{fwd}})$</td>
</tr>
<tr>
<td>Temperature $^\circ\text{C}$</td>
<td>Thickness [nm]</td>
<td>[eV]</td>
</tr>
<tr>
<td>-</td>
<td>0</td>
<td>2.65</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.61</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2.60</td>
</tr>
<tr>
<td></td>
<td>180</td>
<td>2.30</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>3.83</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>3.14</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>4.41</td>
</tr>
<tr>
<td></td>
<td>220</td>
<td>4.03</td>
</tr>
</tbody>
</table>

Table 4: Summary of properties of all n-type samples, calculated from CV and IV data.

<table>
<thead>
<tr>
<th>Buffer characteristics</th>
<th>IV data</th>
<th>CV data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\eta$</td>
<td>$\phi_{bp}^{\text{IV}}(I_0^{\text{fwd}})$</td>
</tr>
<tr>
<td>Temperature $^\circ\text{C}$</td>
<td>Thickness [nm]</td>
<td>[eV]</td>
</tr>
<tr>
<td>-</td>
<td>0</td>
<td>2.27</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.34</td>
</tr>
<tr>
<td></td>
<td>180</td>
<td>2.29</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2.46</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>2.28</td>
</tr>
<tr>
<td></td>
<td>220</td>
<td>2.18</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>2.43</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>2.04</td>
</tr>
</tbody>
</table>
discussed.

In order to probe defects at the interface, the measurements were performed using a reverse bias of $V = -1 \text{V}$ with filling pulses of $V_p = 2 \text{V}$ (relative to the reverse bias). With these parameters, it is expected that all defects at or near the interface are filled by the pulse and contribute to the capacitance transients. The results from the DLTS measurements are presented in Figure 20, for each of the substrate types and buffer deposition temperatures.

For the n-type samples, DLTS peaks are observed mostly in the high temperature range, which correspond to energy levels away from the conduction band. Several samples measured peaks at temperatures higher than the temperature range of the DLTS equipment, making these impossible to determine their energy value. For the sample 200-5nm, a low-temperature peak is observed corresponding to an energy of 0.30eV from the conduction band edge. The rest of the measurable peaks have an energy separation from the conduction band edge in the range of 0.48eV to 0.92eV. The broad peaks observed in several of the plots indicate multiple overlapping levels, and prevent precise determination of the different energy positions. The amplitude of the peaks is proportional to the defect concentration, it is clear that the thicker buffers increase the defect concentration.

On the p-type substrate the defect levels in the lower half of the band gap are investigated. Here, a variety of peaks are observed, with energy levels in the range of 0.10eV to 0.27eV above the valence band edge. As can be seen from Figure 20b, several of the peaks can be observed in multiple samples. For example, the peak at ~250K in sample 180-5nm can be seen in 200-10nm as well. In addition, as with the n-type samples, the broad peaks indicate multiple energy levels. Several of these lower, semi-obscured peaks seem to be at the same temperatures as higher, sharper peaks in other samples. For the p-type sample without buffer, there is a significant concentration of defects, apparently surrounding two barely distinguishable peaks in the low-temperature region and reaching into the higher-temperature region.

### 6.3 Lifetime measurements (QSSPC)

For the lifetime measurements, each of the double-sided samples (see Figure 12b) was subjected to a flash of slow-decaying light of the same strength and duration in the QSSPC setup. The result was plotted as the minority carrier lifetime as a function of minority carrier density in Figure 21.

As can be seen from the plots is that the samples with a very short carrier lifetime do not reach high levels of minority carrier concentration. There is a simple reason for this: In these samples, the recombination rate is higher than the generation rate provided by the lamp. This is especially true for the samples without buffer, which have a high surface recombination.

On the samples on n-type substrate, there is an increase in hole lifetime of up to a factor of three for the two lowest temperatures. For the buffer
Figure 20: DLTS comparison plots for samples on a) n-type and b) p-type silicon substrates, at three different deposition temperatures. Each plot compares the different buffer thicknesses at each specific deposition temperature, which are, from top to bottom, 180°C, 200°C and 220°C.
Figure 21: Minority carrier lifetime vs minority carrier density plots for samples on a) n-type (holes) and b) p-type (electrons) silicon substrates, at three different deposition temperatures. The legend box in each plot shows the deposition temperature of the buffer layer. Note the different values of the Y-axes on the different plots.
deposited at 220°C, the lifetime is increased by a full order of magnitude, showing a significant reduction in recombination. However, no pattern is found in the effect of the buffer layer thickness, and for each temperature, one of the samples display a lifetime shorter than or equal to the unbuffered sample.

For the samples on p-type substrate, there is a clear correlation between the buffer layer thickness and the observed electron lifetime, and all samples show increased lifetime compared to the unbuffered sample. As with the samples on n-type substrate, a large jump in lifetime can be seen for the samples deposited at 220°C, increasing the lifetime from 4-5 times higher than the unbuffered sample to close to an order of magnitude higher. This can suggest a structural shift in the interface between the crystalline substrate and the deposited layer between 200°C and 220°C.
7 Summary and Conclusions

Heterostructures of ZnO/aSi/Si have been investigated, with aSi layer thicknesses in the $5-20\ nm$ range. The aSi buffer layer was deposited by PECVD, with substrate temperatures between 180 and 220$^\circ$C. After PECVD, a layer of aluminium-doped ZnO was deposited by magnetron sputtering, followed by characterization, primarily focused on IV, CV and DLTS.

The IV characteristics show an increased rectification for thin buffers on p-type substrate for 200$^\circ$C and 220$^\circ$C. Other samples, for both n- and p-type substrates, show either similar or lower rectification than the respective samples without buffer.

However, a high ideality factor is observed for all samples, which may indicate a detrimental mechanism not related to the buffer layer.

7.1 Suggestions for further work

Transmission Electron Microscopy (TEM) could be used in order to get information about the abruptness of the junction and the structure of the aSi film, to determine whether the aSi/Si interface is abrupt, or if epitaxial growth occurs during deposition. In addition, the thickness homogeneity of the buffer film could be measured by Atomic Force Microscopy (AFM).

In order to more accurately estimate the effect of the buffer layer thickness and the deposition temperature on the electrical characteristics, a larger series of temperatures and/or thicknesses could be evaluated.

Different techniques for deposition of the ZnO layer is another possibility, to assess the effect on the junction. Different parameters of the buffer layer, for example passivation with hydrogen, is another interesting possibility.
8 Attachments

References


