Numerical Simulations of the Band Bending Profile of an Adatom-covered Semiconductor

---

J.M. van Kerkhof
Technical University of Delft,
Lorentzweg 1, The Netherlands

ABSTRACT: Accurately characterizing the properties of semiconductors at atomic resolution is crucial for advancing semiconductor technology. One of the key challenges in quantitatively interpreting Scanning Tunneling Spectroscopy (STS) is the influence of tip-induced band bending (TIBB) during semiconductor measurements. This thesis presents a numerical method that self-consistently solves the one-dimensional Poisson equation to correct for TIBB in adatom-covered semiconductors.
The model uses the Block-SOR-Newton method to solve the nonlinear system of equations that arises from the discretization of the Poisson equation. Simulations demonstrate that the model can accurately describe the effect of the STM tip voltage on band bending. However, numerical instabilities were observed for high doping concentrations and surface state energies close to the Fermi level, attributed to overshooting of the Newton method. Potential solutions, such as using a Newton-Krylov method and adaptive grid refinement, were proposed to address these instabilities.
Future work includes extending the model to non-equilibrium situations by introducing the full set of semiconductor equations and expanding to three dimensions to account for the STM tip geometry. These advancements would provide a useful tool for correcting STS data, ultimately deepening our understanding of semiconductor physics at the atomic scale.

---

[1] Corresponding author

Preface

Dear reader,

This thesis is the result of an exciting journey through the world of semiconductor physics and numerical methods. It has been an incredibly rewarding experience to work on this project, and I am grateful for the opportunity to have done so. I would like to express my gratitude to the people who have supported me along the way.

First, I would like to thank my supervisors, Alexandre Artaud and Johan Dubbeldam, for their guidance and support throughout this project. Not only have they provided me with valuable feedback and advice, but they also gave me an insider's view of the world of research. I always looked forward to our weekly meetings, since they were both informative and enjoyable.

I would also like to thank Dr. Alexander Heinlein and Prof.dr. Sander Otte for taking the time to read and grade this thesis. I look forward to our discussions during the defense.

Furthermore, I would like to thank my friends and family for their support during my studies. I am especially grateful to my partner Wieke, who has always been there for me, and Alek, who I have shared many interesting discussions with.

Next, I would like to thank my father and mother for encouraging my curiosity from a young age. They have always supported my endeavors, and I am grateful for their unwavering belief in me.

Finally, I would like to thank Dr. Schnedler and Dr. Ebert from Forschungszentrum Jülich for the interesting discussions via video calls and email. They have provided me with valuable insights regarding the modeling of semiconductor devices.

Contents

1 Introduction

Semiconductor technology forms the backbone of modern electronics, enabling advancements in computing, communication, and energy. From microprocessors to solar cells, the unique electronic properties of semiconductors are crucial across various domains. The continuous demand for faster and more efficient devices drives the semiconductor industry to constantly improve its technology. This is achieved by scaling down semiconductor devices and altering their electronic properties at the atomic level. To aid these developments, it is crucial to understand and measure semiconductor properties at the smallest possible scales.

One of the most powerful tools for atomic-scale measurements is the Scanning Tunneling Microscope (STM). Developed by Gerd Binnig and Heinrich Rohrer in 1981, the STM allows for direct imaging of surfaces with atomic resolution [1, 2]. It revolutionized surface science by enabling the first real-space, atomic-resolution images of conducting and semiconducting surfaces. By positioning a sharp metal tip near a sample surface and applying a bias voltage, electrons can quantum-mechanically tunnel between the tip and the sample. The resulting tunneling current is highly sensitive to the tip-sample distance, which makes it possible to image the semiconductor surface with sub-nanometer resolution [3].

Beyond topographic imaging, the STM can be used to probe the electronic properties of semiconductors via Scanning Tunneling Spectroscopy (STS). By measuring the tunneling current as a function of applied bias voltage, STS can provide information about the energetic distribution of electronic (surface) states [4]. This can give detailed information on the effect of surface impurities on the electronic properties of the semiconductor [5, 6]. STS is particularly valuable for identifying band gaps and adatom-induced states [7, 8].

However, there is one problem when interpreting STS data: part of the applied voltage drops in the semiconductor itself [9]. This is referred to as tip-induced band bending (TIBB), and it significantly hinders the quantitative description of semiconductor properties [10]. Hence, developing correction methods for TIBB is essential for accurate quantitative measurements via STM/STS.

In the past decades, several methods have been developed to correct for TIBB. First, analytical models have been proposed to estimate band bending [11, 12]. However, these models made several assumptions that are not always valid in practice [11]. When numerical methods became more accessible, it became possible to determine tip-induced band bending numerically. One of the most renowned programs for this purpose is *SEMITIP*, developed by Feenstra *et al.* [13–16]. This program can accurately determine TIBB by solving the Poisson equation self-consistently [15]. It takes into account the exact geometry of the tip and the sample, and can be used to correct STS data for tip-induced band bending.

However, the *SEMITIP* program assumes that the semiconductor is in a thermal equilibrium. This means that the program is not applicable to non-equilibrium situations, such as when the semiconductor is illuminated by light. To circumvent this limitation, Schnedler *et al.* developed a method that accounts for the complete set of semiconductor equations [17]. By solving the complete set of semiconductor equations self-consistently, this method can be used to correct STS data for tip-induced band bending in non-equilibrium situations. This is a major advancement for scenarios such as photovoltaics and surface photo-voltage measurements [18].

This thesis aims to contribute in two different ways to the field of tip-induced band bending correction. First, the thesis will provide a comprehensive overview of the required theory to understand the work of Feenstra *et al.* and Schnedler *et al.* This will include theory on band bending, surface states, and the semiconductor equations. Second, the thesis will present a model based on the work of Schnedler *et al.* to determine band bending of adatom-covered semiconductor surfaces. Adatoms can significantly modify local electronic structure, introducing localized states that alter band bending. Hence, accurately modelling their effect on TIBB is important for STS measurements.

The thesis is structured as follows. Chapter 2 provides an overview of the relevant theory on semiconductors and band bending. Chapter 3 presents the numerical methods that are used in the model. Chapter 4 presents the description of the model that is used to determine band bending. Chapter 5 presents the results of the model, and Chapter 6 provides a conclusion and outlook for future work.

By bridging semiconductor theory with computational modeling, this thesis paves the way to a deeper understanding of how adatoms and STM interactions influence band bending at a microscopic level. Down the line, these insights could aid researchers in the interpretation of STS data.

2 Theory of Semiconductor Physics

Before discussing the model that determines the adatom-induced band bending in the STM setup, it is important to understand the basic principles of semiconductor physics. This chapter will cover the basic theory of semiconductors, band bending, and surface states. Moreover, it will introduce the necessary equations that are used to model the band bending in the STM setup.

2.1 Semiconductor Theory

To understand band bending in semiconductors, it is important to have a basic understanding of the physical structure and properties of semiconductors. This section will provide a brief overview of the crystal structure of semiconductors, how to visualize valuable information in band diagrams and how to alter their properties via doping.

2.1.1 The Physical Structure of Semiconductors

Most often, and in the entirety of this thesis, semiconductors are crystalline materials. This means that the atoms of the material are arranged in a repeating fashion. The specific ordering of atoms within a material is called the crystal structure and many different types of crystal structures exist. However, a significant fraction of semiconductor materials have either a zincblende (GaAs), diamond (Ge,Si), or wurtzite (GaN) crystal structure. As illustrated in Figure 1, the atoms in these structures are bonded to their four nearest neighbours.

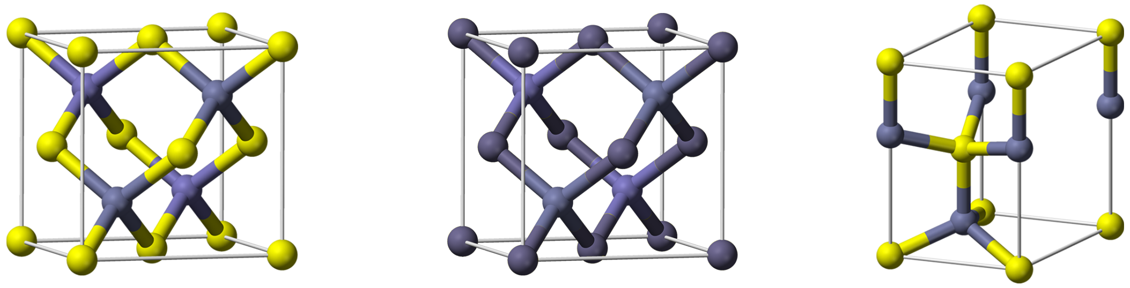

Figure 1. From left to right: zincblende, diamond and wurtzite crystal structures. Zincblende and wurtzite are two-element structures and diamond is a one-element structure. Therefore, compound semiconductors often have a zincblende or wurtzite structure, while elemental semiconductors have a diamond structure. In each structure, the atoms are bonded to their four nearest neighbours.

Only atoms with four valence electrons are able to form these crystal structures, since each atom must share one electron with each of its four nearest neighbours to form a bond. This is the reason that many semiconductors are elements from group IV of the periodic table, such as silicon and germanium, since these naturally have four valence electrons. However, semiconductors can also form from compounds of group III and group V elements, or group II and group VI elements, since these alloys also average to four valence electrons per atom.

These combinations are termed III-V and II-VI semiconductors respectively. Examples of III-V semiconductors are GaAs and InP, while examples of II-VI semiconductors are ZnSe and CdTe.

2.1.2 Band Diagrams

The band diagram is a useful tool for visualizing the energy bands of a semiconductor. It is often used in semiconductor physics to understand and predict the behaviour of semiconductors. To cover this topic, we first must review some electronic band theory.

Solid state physics tells us that the energy levels of electrons in crystalline materials are described by many bands, called the band structure. Each of these bands is described by a dispersion relation $E(\mathbf{k})$, where $\mathbf{k}$ is the crystal momentum [19]. A schematic of the band structure of GaAs, shown in Figure 2, illustrates this concept.

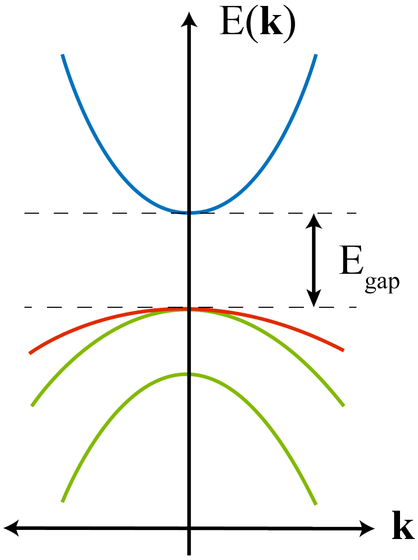

Figure 2. A schematic of the band structure of GaAs. The blue band is the conduction band and the red band is the valence band. The energy difference between the top of the valence band and the bottom of the conduction band is called the band gap.

Most often, we are interested in the valence and conduction bands. The valence band is the highest energy band that is fully occupied by electrons, while the conduction band is the lowest energy band that remains empty. The conduction band is appropriately named, since electrons that are excited to this band are free to move around in the semiconductor and therefore make conduction possible. The energy difference between the top of the valence and the bottom of the conduction band is called the band gap, and it signifies the minimal energy needed for an electron to excite from the valence to the conduction band.

The band gap is an important property, since it plays a prominent role in the material's conductivity. For example, materials with a large band gap are insulators, while metals have no band gap. Semiconductors have a band gap smaller than insulators, but larger than that of metals (typically between 0.2 and 3.5 eV at $T = 300$ K [19]). This allows for thermal excitations or photoexcitations to create free electrons in the conduction band and holes in the valence band, which are responsible for the conductivity of the semiconductor [19].

In semiconductor physics we are often interested in the spatial dependence of the semiconductor's electronic properties. For this reason, we plot the top of the valence band and the bottom of the conduction band as a function of position, creating the band diagram.

Since we are plotting the energies of the available electronic states, it is also convenient to include the Fermi level in the band diagram. In semiconductor physics, the Fermi level is defined as the chemical potential of the system at $T = 0K$ [19]. It plays an important role in the conductivity of the semiconductor, but also in determining other quantities such as the (surface) charge. In thermal equilibrium, the Fermi level is constant throughout the system. The Fermi level will be further explored in the subsequent sections.

In summary, there are four important quantities that can be directly taken from the band diagram. The position-dependent energies of the valence and conduction bands show how the semiconductor behaves at different locations. The band gap indicates how easily an electron can transition between the valence and conduction bands. Finally, the Fermi level is essential for determining many different properties of the semiconductor. An example of a band diagram of a *pn*-diode can be seen in Figure 3.

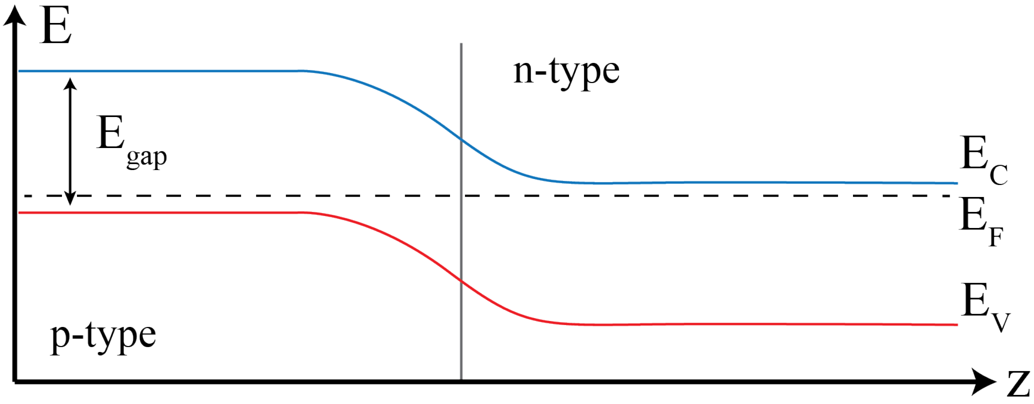

Figure 3. A schematic of a band diagram of a *pn*-diode. A *pn*-diode is a semiconductor device that consists of a *p*-doped and an *n*-doped semiconductor. The blue line represents the conduction band, the red line represents the valence band and black dotted line represents the Fermi level. This figure shows us that an *n*-type electron needs to overcome an energy barrier to go to the *p*-type region.

2.1.3 Charge Carriers and Doping

Precise tuning of the electronic properties of semiconductors is essential for their applicability. To do this, impurities are introduced in the crystal lattice. This process is termed doping.

To understand doping, it is important to cover how conduction occurs in semiconductors. Because of the relatively small band gap in semiconductors, thermal excitations can cause an electron to move from the valence to the conduction band. This process introduces a free electron in the conduction band and leaves a missing electron in the valence band, known as an (electronic) hole. Holes behave as free, positive quasiparticles in the valence band and act as charge carriers alongside the free electrons. The charge carriers are responsible for conductivity.

The concentrations of the charge carriers are determined by the Fermi level, E_F, which therefore plays an important role in the conductivity of the semiconductor. The free electron concentration n and hole concentration p relate to the Fermi level via the following approximations:

$$n = N_c \exp\left(\frac{E_F - E_c}{kT}\right) \tag{2.1}$$

$$p = N_v \exp\left(\frac{E_v - E_F}{kT}\right) \tag{2.2}$$

Where N_c and N_v are the effective densities of states in the conduction and valence bands respectively, E_c and E_v are the conduction and valence band energies, k is the Boltzmann constant and T is the temperature [20]. The equations show that n increases if E_F gets closer to the conduction energy, while p increases if E_F gets closer to the valence energy. Now, if we want to alter the conductivity of the semiconductor, we can choose to increase the availability of electrons or holes via doping. When we dope a semiconductor, some of the atoms in the crystal are substituted with different atoms known as dopants. If a dopant from a higher group is added, it will have a valence electron left, which can be easily excited to the conduction band. Such dopants are called donors, and the process is termed n-type doping. Conversely, a dopant from a lower group will accept electrons from the valence band, which would leave holes. These dopants are called acceptors, and the process is termed p-type doping. Figure 4 shows a simplified depiction of n-type and p-type doping of a silicon lattice.

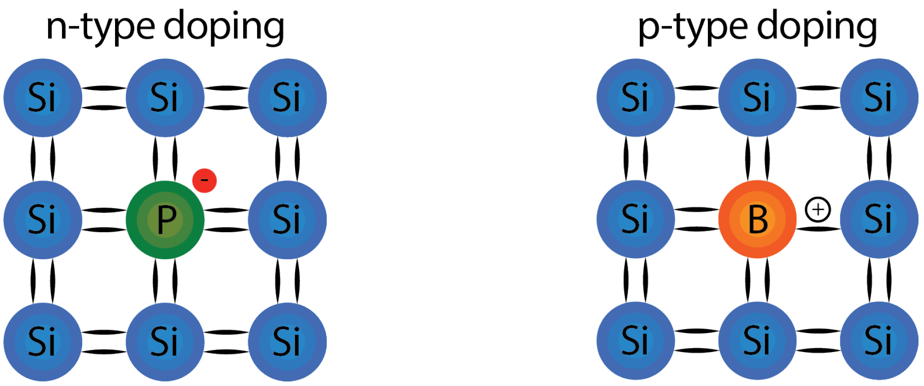

Figure 4. A schematic of n-type (left) and p-type (right) doping in a silicon (group IV) lattice. In the left figure, a phosphorus atom is added. Since it is from group V, it has an extra electron that can be excited to the conduction band. In the right figure, a boron atom is added. Since it is from group III, it can accept an electron, leaving a hole in the valence band.

Another way to understand doping is via the Fermi level. When we dope our material, each dopant creates a new electronic state. These states all have their own wavefunction which typically extends over multiple lattice constants. Therefore, if we add enough dopants, the electronic states will overlap and form donor or acceptor bands. These bands increase the availability of electrons or holes in the semiconductor system, which alters the chemical potential of the system, even at $T = 0K$. Therefore, the Fermi level shifts upward for n-type doping and downward for p-type doping. We can verify this with equations (2.1-2.2) by observing that n (p) can only increase if E_F goes up (down). This shift is illustrated in the band diagram shown in Figure 5.

Two additional types of doping are worth mentioning. The first is called amphoteric doping. Amphoteric doping occurs when a dopant can act both as a donor and an acceptor. An example would be doping GaAs (III-V) with silicon (IV). The silicon acts as a donor if it replaces a gallium atom, but as an acceptor if it replaces an arsenic atom. The eventual result depends on the manufacturing conditions [21].

The second type is called isoelectronic doping. This involves adding a dopant from the same group as the lattice atom. While the total number of valence electrons remains unchanged, the properties of the semiconductor, such as its bandgap, can still be altered. An example would be doping GaP (III-V) with nitrogen (V). Nitrogen has a different size and electronegativity than phosphorus, which can decrease the semiconductor band gap [22].

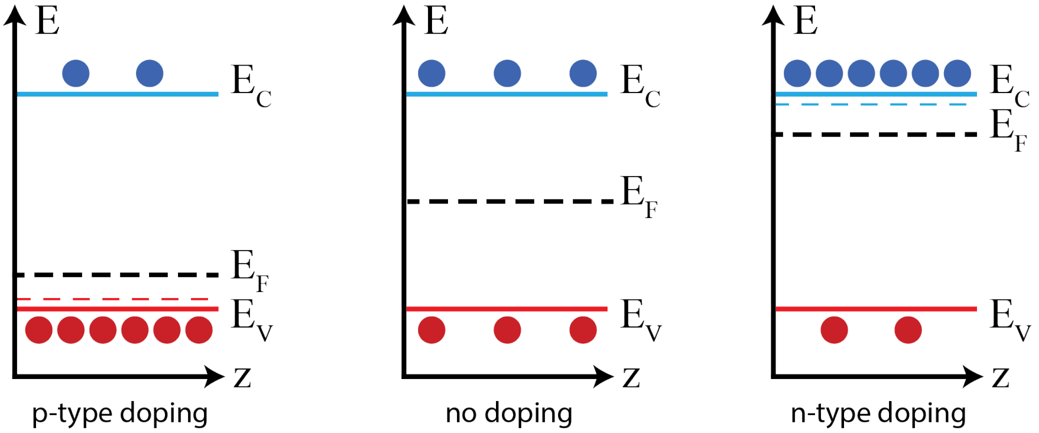

Figure 5. A schematic of the band diagram of a semiconductor with *p*-type doping (left), no doping (middle), and *n*-type doping (right). The black dotted line represents the Fermi level. The coloured dotted lines represent the dopant bands. The Fermi level shifts upwards for *n*-type doping and downwards for *p*-type doping. This affects the carrier concentrations in the semiconductor, as indicated by the coloured circles. The red circles represent holes and the blue circles represent electrons.

2.2 Band Bending

Band bending is a phenomenon that occurs at the interface between a semiconductor and another medium. It is characterized by a shift in the energy bands of the semiconductor near the interface, caused by differences in the Fermi level between the two media. When the media make contact, a new thermal equilibrium forms via charge redistribution. This process shifts the energy bands. Often, band bending plays an important role in the behaviour of semiconductor devices. This section explains what band bending is, why it occurs, and its implications.

2.2.1 Semiconductor-Metal Interface

Before we consider a mathematical treatment of band bending, it is useful to examine the semiconductor-metal interface qualitatively. Consider an *n*-doped semiconductor and a metal that are separated in space. Since they are not in contact, the Fermi levels of the two materials are independent. For this example, we assume that $E_{F,metal} < E_{F,sem}$. A new thermal equilibrium must be established when the two materials make contact. This equilibrium is reached when the Fermi levels of the two materials become equal. Therefore, electrons flow from the material with the higher Fermi level to the material with the lower Fermi level until equilibrium is reached. Thus, electrons move from the semiconductor to the metal.

However, due to charge conservation, electrons accumulating in the metal result in a negatively charged surface. The resulting charge generates an electric field that is directed towards the metal. Due to this electric field, electrons in the semiconductor have a higher electric potential energy near the interface than in the bulk of the semiconductor. Therefore, the energy bands bend upwards near the interface, as illustrated in Figure 6. Effectively, this pushes electrons away from the interface, depleting the mobile charge carriers and forming a so-called depletion layer. This layer is characterized by the immobile positive donor ions that remain after the electrons have been pushed away and is therefore positively charged. To adhere to charge conservation, the positive charge of the depletion layer must compensate the negative charge on the metal surface. The depletion layer is a specific case of a phenomenon known as the space charge layer, which is discussed in the next section.

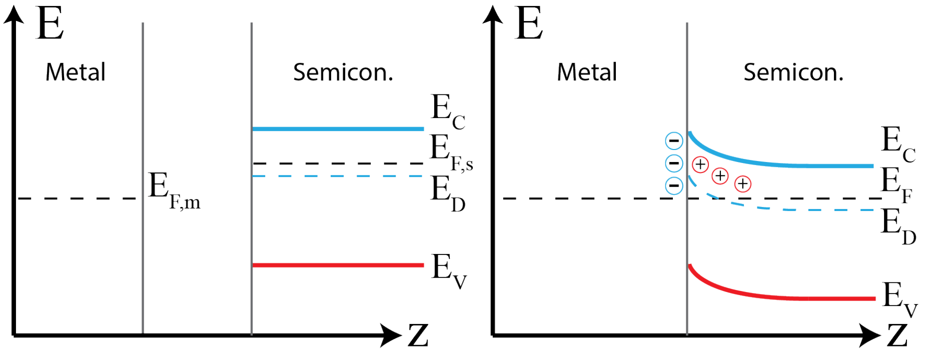

Figure 6. A schematic of band bending at a semiconductor-metal interface. The left figure shows the situation before contact. The right figure shows the situation after contact. The Fermi levels aligned and a negative charge accumulated on the metal surface. This is compensated by a positive depletion layer inside the semiconductor. The dotted blue line gives the donor band level E_D.

2.2.2 Origin of the Space Charge Layer

Having qualitatively discussed charge redistribution for a semiconductor-metal interface, we now give a detailed description of the space charge layer. This section will show why space charge layers form and how they can impact semiconductor device performance. To analyze band bending, we will consider a doped semiconductor with electronic surface states. These states are described by a density of states (DOS) function $N_{ss}(E)$ and can be classified as either donor or acceptor states. Surface states can occur for multiple reasons (e.g., surface dopants, dangling bonds) and will be discussed further in section 5.3.2. Figure

7 illustrates the situation. The following discussion closely follows Ref. [23].

A surface state can become charged by either accepting or donating an electron, where the occupation probability depends on the position of the Fermi level relative to the surface state. Since charge neutrality must be maintained, any charge Q_{ss} that is introduced by the surface states must be balanced by an equal and opposite charge Q_{sc} in the semiconductor, forming a space charge layer. This leads to the charge conservation equation:

$$Q_{sc} = -Q_{ss}. \tag{2.3}$$

Charge conservation (2.3) determines the Fermi level's position relative to the surface states. This can be seen as follows. Deep inside the semiconductor, the electronic effects of the surface are negligible. Therefore, when far from the surface, the position of the Fermi level relative to the band edges is determined by the bulk properties of the semiconductor (e.g., doping). When we look at the surface, it is important to note that the surface DOS $N_{ss}(E)$ is defined relative to the surface band edges [23]. Thus, if the band edges are shifted due to band bending, $N_{ss}(E)$ is also shifted. This shift in the surface DOS changes its relative position to the Fermi level, altering the charge Q_{ss} of the surface states. However, band bending also shifts the dopant bands relative to the Fermi level. As a result the number of occupied dopants changes, modifying the charge distribution Q_{sc} in the space charge layer. The system reaches thermal equilibrium when the Fermi level relative to the surface states aligns such that these charges balance, ensuring charge conservation. The resulting system is illustrated in Figure 7.

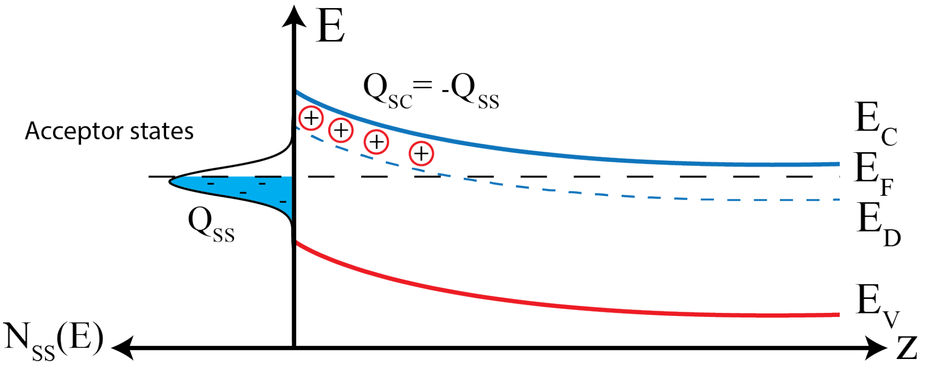

Figure 7. A schematic of band bending at an n-type semiconductor with acceptor surface states. The surface states under the Fermi level E_F are occupied, which creates a surface charge Q_{ss}. This charge is compensated by a positive (depletion) space charge layer in the semiconductor. This layer carries a charge $Q_{sc} = -Q_{ss}$. [23]

The space charge layer that forms due to this charge redistribution can be classified into three types: depletion, inversion, and accumulation layers. The differences are best illustrated in Figure 8a. To discuss the depletion and accumulation layers, consider an n-doped semiconductor with acceptor surface states. The surface states generate a negative charge, which causes upward band bending. Consequently, upward band bending lifts the donor band above the Fermi level, moving the electrons away from the surface. The positive donor ions are left behind while the electrons (which are the majority carriers) are depleted. This results in a depletion layer.

When strong upward band bending occurs, the space charge layer extends deeper into the semiconductor pushing even more electrons away. As a result the concentration of minority carriers (holes) surpasses the concentration of majority carriers (electrons) near the surface, leading to an inversion layer.

If instead the surface states are donor states, the surface generates a positive charge, causing downward band bending. This attracts electrons to the surface, forming an accumulation layer. Typically, accumulation layers are much narrower than the other two types of layers, since the mobile charge carriers can be tightly packed. This contrasts the depletion and inversion layers, where the dopant ions are immobile. For p-doped semiconductors the situations are reversed, with acceptor surface states leading to accumulation layers and donor surface states leading to depletion layers, where holes play the role of the electron.

The different types of space charge layers have different implications for semiconductor devices. This becomes clear if we consider the free charge carrier concentrations, which link directly to the conductivity. This is summarized in Figure 8b. Depletion layers have less free charge carriers, reducing the conductivity near the surface. Accumulation layers increase the conductivity near the surface by attracting free charge carriers. The conductivity in inversion layers is more complex. Near the surface it exhibits high conductivity due to an excess of minority carriers, similar to an accumulation layer. However, deeper into the semiconductor, majority carriers are depleted, forming a depletion-like layer. Many semiconductor devices rely on space charge layers to function properly. For example, the p-n junction diode depends on the depletion layer to create a barrier for charge carriers, while the MOSFET uses the inversion layer to control the conductivity of the semiconductor [20].

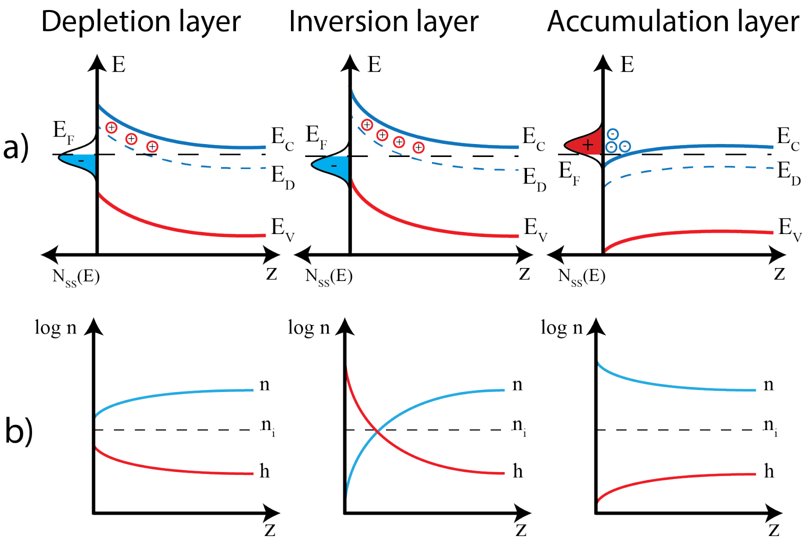

Figure 8. An illustration of the different space charge layers of an n-type semiconductor surface. (a) shows the band bending structure. We see that the inversion layer has more band bending than the depletion layer. The right figure has donor surface states, which creates a downward bending accumulation layer. (b) shows the free carrier densities of the different layers. n, h and n_i give the free electron, hole and intrinsic carrier concentrations respectively. The depletion layer has fewer majority carriers near the surface. For the inversion layer, the minority carriers eventually surpass the majority carriers. The accumulation layer has more majority carriers near the surface.

2.2.3 Poisson's Equation

After treating the formation of space charge layers, it is time to discuss the mathematical description of band bending. To describe band bending, we introduce the position-dependent electrostatic potential $\phi(\mathbf{x})$. We take the bulk of the semiconductor as our reference for $\phi = 0$. With the electrostatic potential, the amount of band bending at point $\mathbf{x}$ can be expressed as $\Delta E(\mathbf{x}) = -q\phi(\mathbf{x})$, where q is the unsigned elementary charge and ΔE denotes the change in energy. This allows us to use Poisson's equation to describe the band bending profile. Poisson's equation relates the electric potential $\phi(\mathbf{x})$ to the charge density $\rho(\mathbf{x})$ in the semiconductor. It is given by [23]:

$$\nabla^2\phi(\mathbf{x}) = -\frac{\rho(\mathbf{x})}{\varepsilon} \tag{2.4}$$

Where ε is the permittivity of the semiconductor. For most purposes it suffices to regard ε to be a scalar constant throughout the semiconductor, but to model anisotropic effects the tensorial permittivity must be used [24, 25].

The charge density in the semiconductor is completely determined by the concentration of charge carriers and ionized dopants. This allows us to write the charge density as [17]:

$$\rho(\mathbf{x}) = q\left(p(\mathbf{x}) - n(\mathbf{x}) + N_{\mathrm{D}}^{+}(\mathbf{x}) - N_{\mathrm{A}}^{-}(\mathbf{x})\right) \tag{2.5}$$

Where n and p are the electron and hole concentrations, and N_{D}^{+} and N_{A}^{-} are the ionized donor and acceptor concentrations. The concentration of the ionized dopants depends on the position of the dopant band levels relative to the Fermi level. Assuming Boltzmann statistics to be valid, this can be expressed as [20]:

$$N_{\mathrm{D}}^{+}(\mathbf{x}) = \frac{N_{\mathrm{D}}}{1 + g_D \exp\left(\frac{E_F - (E_D - q\phi(\mathbf{x}))}{kT}\right)} \tag{2.6}$$

$$N_{\mathrm{A}}^{-}(\mathbf{x}) = \frac{N_{\mathrm{A}}}{1 + g_A \exp\left(\frac{(E_A - q\phi(\mathbf{x})) - E_F}{kT}\right)} \tag{2.7}$$

Where N_{D} and N_{A} are the total donor and acceptor concentrations, E_D and E_A are the bulk donor and acceptor band levels, and g_D and g_A are the degeneracy factors. The degeneracy factors represent how many different configurations the ionized dopant can have. The degeneracy factor for a donor state is typically 2, because the donor can accept an electron with either spin [20]. For the acceptor state, the degeneracy factor is 4 for most cubic symmetric semiconductors (e.g., zincblende and diamond crystal structures) [26]. This is because they have a doubly degenerate valence band for $\mathbf{k} = 0$ and can accept a hole of either spin [20].

Equations (2.6 - 2.7) show that the ionized donor concentration increases if the donor band level shifts upward, while the ionized acceptor concentration increases if the acceptor band level shifts downward. This is in accordance with section 2.2.2, where we saw that upwards band bending in n-doped semiconductors creates a depletion layer, while for p-doped semiconductors this happens with downward band bending.

2.3 Surface State Theory

In section 2.2.2, we saw that surface states can cause band bending and the formation of space charge layers. In this section we will discuss the origin of these electronic surface states. Specifically, we will make a distinction between intrinsic surface states and adsorbate-induced surface states. We will also discuss the process and implications of Fermi level pinning.

2.3.1 Intrinsic Surface States

When discussing the solid-state physics of a semiconductor, it is often assumed that the material has an infinitely extending periodic lattice. Under this assumption, a periodic potential can be defined, which is a necessary condition for Bloch waves to exist as solutions to the Schrödinger equation. However, when we introduce a surface, this assumption is broken. This requires us to find a new solution to the Schrödinger equation which takes the existence of the surface into account. A thorough derivation of this solution can be found in chapter 6 of Ref. [23]. Here we sketch the main ideas behind the derivation and discuss the implications.

For an infinitely extending semiconductor, the wavefunction is given by a Bloch wave [23]:

$$\psi_{\mathbf{k}}(\mathbf{x}) = u_{\mathbf{k}}(\mathbf{x}) \exp(i\mathbf{k} \cdot \mathbf{x}) \tag{2.8}$$

Where $u_{\mathbf{k}}(\mathbf{x})$ is a periodic function with the same periodicity as the lattice. From a physcial standpoint, only real values for $\mathbf{k}$ are interesting, since we expect delocalized electronic states. However, when introducing a surface, we expect localized states to form. These states have imaginary values for $\mathbf{k}$, which corresponds to evanescent states. This also yields different energy eigenvalues, which are allowed to lie within the forbidden band gap of the semiconductor [23]. These states are called surface states. Specifically, in the nearly free electron model they are called Shockley states, while in the tight binding model they are known as Tamm states.

If we consider the tight binding model, the existence of surface states has an intuitive explanation. In the tight-binding model, the wavefunction is described as a linear combination of atomic orbitals (LCAO) [19]. If we truncate our semiconductor, the atoms at the surface have fewer neighbours than the atoms in the bulk. As a result, their atomic orbitals experience less hybridization, creating a state with a new energy level [23]. This effect is particularly strong for orbitals that normally would participate in bonding with atoms that are now missing. These states are called dangling bonds and their energy values typically deviate significantly from the bulk band edges [23].

In the LCAO framework, the surface states are a combination of conduction and valence band states. This means that the surface states can be either donor-like or acceptor-like states. Typically, the surface states close to the valence band are more donor-like, while they are more acceptor-like close to the conduction band [23]. This allows us to define a neutrality energy level E_N (also known as the branching point energy E_B) which separates the donor-like and acceptor-like states. If the Fermi level is equal to E_N, the surface will be neutral and no band bending will occur [23]. Figure 9 displays a situation with intrinsic surface states.

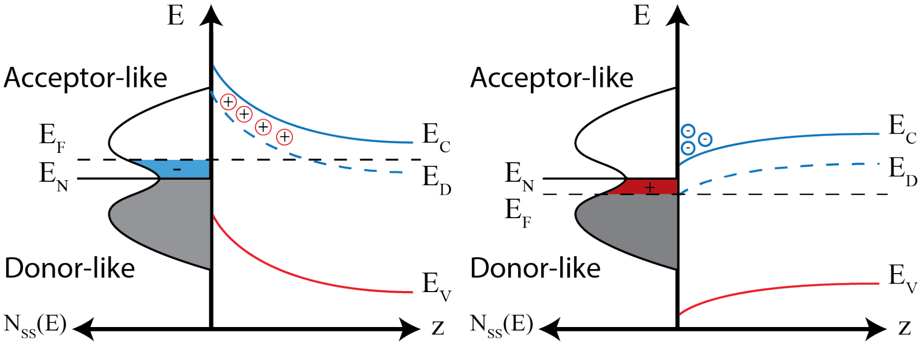

Figure 9. A schematic of band bending at a n-type semiconductor surface with intrinsic surface states. The neutrality level E_N separates the donor-like and acceptor-like surface states. The left figure shows a situation with a Fermi level above E_N, which creates a depletion layer. The right figure shows a situation with a Fermi level below E_N, which creates an accumulation layer.

2.3.2 Adatom-induced Surface States

When we adsorb atoms on top of the semiconductor surface they are known as adatoms. These adatoms will engage in chemical bonds with the semiconductor. The resulting bonding and antibonding orbitals will form additional surface states [23]. It can also affect the distribution of the intrinsic surface states [23, 27].

The question remains what type of surface states are formed by the adatoms. The answer to this question is not straightforward and can be tackled in many ways [28]. As with the intrinsic surface states, the adatom-induced surface states can be either donor or acceptor states. The type of state that is formed depends on the adatom and the semiconductor.

One way to determine the type of surface state is to consider the dipole moment formed by the adatom-semiconductor bond. This dipole moment is formed by the difference in the electronegativity of the adatom and the semiconductor, which causes the more electronegative atom to be slightly negatively charged, inducing a dipole moment [28]. If this dipole moment points towards the semiconductor, the induced surface state is acceptor-like. If it points away from the semiconductor, the state is donor-like [28].

Quantitatively determining the exact density of states of the adatom-induced surface states is a complex task. Firstly, it depends on the distribution of the adatoms on the surface. A low concentration will yield discrete energy levels, but monolayer coverage will result in a continuous band of states as the adatoms will interact with each other [28]. Secondly, the adatom material also plays a role. For example, the energy of the states induced by metal adatoms depend linearly on their ionization energy [29]. Some studies on adatom-induced surface states have used density functional theory (DFT) to calculate the density of states [30].

2.3.3 Fermi Level Pinning

In section 2.2.2, we saw that surface states can induce band bending. The charges of the surface states have to be compensated by a space charge layer, which happens via band bending. By that token, we would expect that the band bending will keep increasing as we increase the surface state density. However, this is not the case. Instead, the band bending saturates at a certain value. This phenomenon is known as Fermi level pinning.

As we have seen in section 2.3.1, a neutrality energy level E_N separates the donor and acceptor states. Typically, it is energetically favourable for the Fermi level to get close to E_N, since this will minimize the charge of the surface states [23]. So the band bending will adjust until the Fermi level is close to E_N. If this is the case, an increase in the density of surface states has a small effect on the amount of surface charge. Therefore, the band bending does not increase significantly anymore and it is said that the Fermi level is pinned to E_N [23].

Since Fermi level pinning fixes the position of the Fermi level at the interface, it can greatly impact the operation and creation of semiconductor devices. Solar cells and MOSFETs are examples of devices that are sensitive to Fermi level pinning [31, 32]

2.4 Band Bending Under Irradiation

In section 2.2, we saw that Poisson's equation can fully describe the band bending profile of a semiconductor when the system is in thermal equilibrium. However, when there is no thermal equilibrium, Poisson's equation alone will not suffice to describe the band bending properly. For example, when the semiconductor is illuminated with a laser. The carrier-photon interactions affect the charge distribution via generation and recombination mechanisms, which alter the band bending. This section aims to explain the relevant equations to describe irradiated semiconductors and will explain the different recombination mechanisms.

2.4.1 Continuity Equations

The photons will create additional charge carriers by excitation. To account for these additional carriers, we must introduce the continuity equation. Intuitively, the continuity equation states that charge is also conserved locally. In other words, when a net current flows out of a volume, the charge density of the volume must decrease. The continuity equation is stated as [24]:

$$\nabla \cdot \mathbf{J} + q\frac{\partial \rho}{\partial t} = 0 \tag{2.9}$$

Where $\mathbf{J}$ denotes the current density. We assume that there is a steady-state situation where the charge density is constant in time. We also observe that we can write our current density as the sum of the hole current density $\mathbf{J}_p$ and the electron current density $\mathbf{J}_n$. This allows us to rewrite equation (2.9) as [24]:

$$\nabla \cdot (\mathbf{J}_p + \mathbf{J}_n) = 0 \tag{2.10}$$

Which can be separated into two equations [24]:

$$\nabla \cdot \mathbf{J}_n = qR \tag{2.11}$$
$$\nabla \cdot \mathbf{J}_p = -qR \tag{2.12}$$

Where R denotes the recombination rate. A positive R means recombination, while a negative R means generation [24]. The mechanisms behind generation/recombination will be given in the subsequent section.

Generally, currents in semiconductors are composed of a diffusion term and a drift term. This allows us to write equations (2.11 - 2.12) as [17]:

$$\nabla \cdot (D_n \nabla n(\mathbf{x}) - \mu_n n(\mathbf{x})\, \nabla \phi) - R = 0 \tag{2.13}$$
$$\nabla \cdot (D_p \nabla p(\mathbf{x}) + \mu_p p(\mathbf{x})\, \nabla \phi) - R = 0 \tag{2.14}$$

Where D_i is the carrier-specific diffusion coefficient and μ_i is the carrier-specific mobility. D_i and μ_i are related via the Einstein-Smoluchowski relation [24]:

$$D_i = \mu_i \frac{kT}{q} \tag{2.15}$$

Which is valid for carriers described by Boltzmann statistics [24].
Equations (2.13 - 2.14) are the final form of the continuity equations. Chapter 2.3 in Ref. [24] provides an extensive derivation and an overview of all underlying assumptions for these equations.

2.4.2 Generation and Recombination Mechanisms

The recombination rate R in equations (2.13 - 2.14) is a crucial parameter in the continuity equations. It describes the rate at which charge carriers are generated or recombined. This section will discuss band-to-band radiative transitions. It will also briefly discuss two other important transition mechanisms: Shockley-Read-Hall recombination and Auger recombination.
Before we discuss the different mechanisms, it is important to distinguish between direct and indirect bandgap semiconductors. In direct bandgap semiconductors, the minimum of the conduction band and the maximum of the valence band occur at the same **k** in the electronic band structure. Most III-V semiconductors have a direct bandgap. In indirect bandgap semiconductors, the conduction band's minimum energy and the valence band's maximum energy occur at different **k** in the electronic band structure. Silicon is an example of an indirect bandgap semiconductor.
When radiative recombination occurs, an electron in the conduction band recombines with a hole in the valence band. The energy difference between the electron and the hole is emitted as a photon. This process is only possible in direct bandgap semiconductors. The process can also happen in reverse, where a photon excites an electron from the valence band to the conduction band. Therefore, radiative recombination plays an important role in describing irradiated semiconductors.
The recombination rate for radiative recombination in thermal equilibrium without band bending is given by [33]:

$$R_{\text{eq}} = b n_0 p_0 \tag{2.16}$$

Where b is the radiative recombination coefficient and n_0 and p_0 are the equilibrium carrier concentrations. However, when we introduce a generation term R_{gen} to account for the additional carriers created by the photons, a new equilibrium will be established. Now the new equilibrium carrier concentrations are $n_{\text{new}} = n_0 + \Delta n$ and $p_{\text{new}} = p_0 + \Delta p$. The recombination rate can now be written as [33]:

$$R_{\text{net}} = b(n_{\text{new}} p_{\text{new}} - n_0 p_0) \tag{2.17}$$

To take band bending into account, we must consider the position-dependent carrier concentrations. This yields [17]:

$$R(\mathbf{x}) = b(n(\mathbf{x})\, p(\mathbf{x}) - n_0 p_0) - R_{\text{net}} \tag{2.18}$$

Δn and Δp can be determined if the specifications of the source of illumination are known [17].
After discussing band-to-band radiative recombination, it is time to consider non-radiative

recombination briefly. The two most important types are Shockley-Read-Hall (SRH) recombination and Auger recombination. For SRH, a state in the bandgap of the semiconductor must exist, the so-called trap state. This state can trap electrons from the conduction band and then release them back to the valence band. This can also be seen as trapping an electron from the conduction band and trapping a hole from the valence band. The energy generated during combination is released via a phonon [33]. SRH recombination is often the dominant recombination process for indirect bandgap semiconductors [33].
Auger recombination is a recombination process that involves three particles. An electron collides with another electron in the conduction band, which then gets excited to a higher energy level within the conduction band. Then the other electron recombines with a hole in the valence band. Auger recombination can become the dominant recombination process with high carrier concentrations or narrow bandgaps [33].

2.4.3 The Semiconductor Equations

We can mathematically describe band bending for the illuminated semiconductor using the equations we introduced in the previous sections. This can be done by solving the following system of equations:

$$\nabla^2\phi - \frac{q}{\varepsilon}(n - p + N_{\mathrm{A}}^{-}(\phi) - N_{\mathrm{D}}^{+}(\phi)) = 0 \tag{2.19}$$

$$\nabla \cdot (D_n \nabla n - \mu_n n \nabla\phi) - R(n,p) = 0 \tag{2.20}$$

$$\nabla \cdot (D_p \nabla p + \mu_p p \nabla\phi) - R(n,p) = 0 \tag{2.21}$$

In this system ϕ, n and p are our unknowns. The ionized dopant concentrations are given by equations (2.6 - 2.7) and the recombination rate is given by equation (2.18). These equations are often referred to as the semiconductor equations. They can be compactly written as $F(\boldsymbol{\omega}) = \mathbf{0}$, where $\boldsymbol{\omega} = (\phi(\mathbf{x}), n(\mathbf{x}), p(\mathbf{x}))^{\intercal}$. The solution to this system of equations will give us the band bending profile of the semiconductor under irradiation.

3 Theory of Numerical Methods

The system of equations described in section 2.4.3 cannot be solved analytically [24]. Therefore, numerical methods are required to obtain a solution. In the model described in section 4, the Block-SOR-Newton method is deployed to solve a discretization of the system of equations. This section serves to provide the necessary mathematical background to understand the numerical methods used in the model. First, the Newton-Raphson method and the Successive Over-Relaxation (SOR) methods are discussed. Finally, they will be combined into the Block-SOR-Newton method, which is treated in more detail.

3.1 Newton-Raphson Method

For many nonlinear equations, it is often difficult, or even impossible, to find their roots. When this is the case, we must resort to numerical approximations of the roots. The Newton-Raphson method (often referred to as Newton's method) is arguably the most widely used and important technique for this purpose.

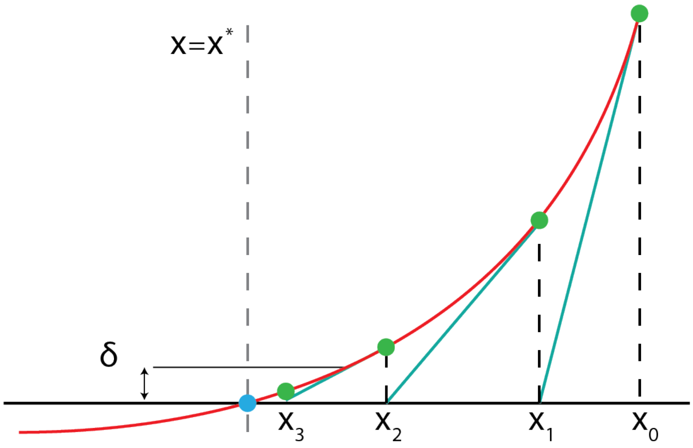

Figure 10. An example of Newton's method applied to a 1D function. The red curve represents the function $f(x)$, the blue lines represents the tangent line at x_k. For each iteration, the method comes closer to the root x^*. Since $f(x_3) < \delta$, the method terminates and accepts x_3 as the approximation.

Qualitatively, Newton's method works as follows. Imagine we have a sufficiently smooth function $f : \mathbb{R} \to \mathbb{R}$, and we want to find x^* such that $f(x^*) = 0$. We begin by inserting an initial guess x_0 of the root and we determine the tangent line at that point, this tangent line serves as a linear approximation of our function. Then we determine where the tangent line has its x-intercept and we name this point x_1. If we find $f(x_1)$ to be close enough to zero, we stop the method. Otherwise we repeat the process for point x_1 and continue until we find a satisfactory x_k such that $f(x_k) < \delta$, where δ is our desired accuracy. An example of applying Newton's method to a 1D function can be found in Figure 10. The k-th iteration of Newton's method for 1D functions can be summarized by the following equation:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \qquad \text{where } k \in \mathbb{N} \tag{3.1}$$

Newton's method also works for higher dimensional functions. Suppose we have a sufficiently smooth function $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$, then the mathematical formulation of equation (3.1) is slightly altered. Instead of using the first derivative, we must use the Jacobian matrix of $\mathbf{F}$, often denoted as J_F. Then the multidimensional formulation becomes

$$\mathbf{x_{k+1}} = \mathbf{x_k} - J_F^{-1}(\mathbf{x_k})\mathbf{F}(\mathbf{x_k}) \tag{3.2}$$

Where J_F^{-1} denotes the inverse of the Jacobian of $\mathbf{F}$. Since inverting a matrix is often computationally expensive, it is more practical to write this as

$$J_F(\mathbf{x_k})(\mathbf{x_{k+1}} - \mathbf{x_k}) = -\mathbf{F}(\mathbf{x_k}) \tag{3.3}$$

This is in the form of $A\mathbf{x} = \mathbf{b}$, so we can use numerical methods that can solve linear equations to solve for $\mathbf{x}$.

An advantage of Newton's method is that an approximation of the Jacobian J_F still allows a convergent method. This is useful if the Jacobian cannot be determined exactly. We can prove this by considering Newton's method as a fixed-point equation in the form of $\mathbf{x} = M(\mathbf{x})$, where $M(\mathbf{x}) = \mathbf{x} - J_F^{-1}(\mathbf{x})\mathbf{F}(\mathbf{x})$. The Ostrowski theorem states that if $\rho(M'(\mathbf{x}^*)) < 1$, there is an open neighbourhood S of $\mathbf{x}^*$ such that the sequence $\mathbf{x}_{k+1} = M(\mathbf{x}_k)$ converges to $\mathbf{x}^*$ for all $\mathbf{x_0} \in S$ [34]. Here, $\rho(M'(\mathbf{x}^*))$ denotes the spectral radius of the Fréchet derivative of M at $\mathbf{x}^*$. The Fréchet derivative of Newton's method is $M'(\mathbf{x}) = I - B^{-1}(\mathbf{x})J_F(\mathbf{x})$, where $B(\mathbf{x}) = \gamma J_F(\mathbf{x})$, with $\gamma \in \mathbb{R}$, is our approximation of the Jacobian. Thus, the Fréchet derivative of Newton's method is $M'(\mathbf{x}) = (1 - \frac{1}{\gamma})I$. We can see that the spectral radius of this matrix is $1 - \frac{1}{\gamma}$. Therefore, if $\gamma > 0.5$, the method is locally convergent [24].

Although Newton's method is a powerful tool, it is not without its drawbacks. A common problem is overshoot, where the method overestimates the correction needed to find the root. This can significantly slow convergence, or even cause the method to diverge. Another problem occurs if the initial guess is not in the region of convergence. In this case, the method can also diverge. Finally, determining the Jacobian can be too expensive to viably compute for each iteration. There are modifications to Newton's method that can help in specific cases, such as a damped Newton's method, Hirano's method or quasi-Newton methods like the BFGS method [35–37].

3.2 Successive Over-Relaxation Method

The Successive Over-Relaxation (SOR) method is an iterative method used to solve equations in the form of $A\mathbf{x} = \mathbf{b}$. It is a modification of the Gauss-Seidel method, which is a method that solves the equation by iterating over the elements of $\mathbf{x}$ and updating them one by one. First we will derive the SOR method and then we will discuss its properties. The derivation for the SOR method is based on Ref. [38].

3.2.1 Derivation of the SOR Method

Since we are concerned with iteratively solving $A\mathbf{x} = \mathbf{b}$ for $\mathbf{x}$, we are looking for a fixed-point equation in the form of $\mathbf{x} = M(\mathbf{x})$. To do this we begin by decomposing A into a lower-triangular matrix L, a diagonal matrix D and an upper-triangular matrix U. This decomposition yields $A\mathbf{x} = (L + D + U)\mathbf{x} = \mathbf{b}$. We can rewrite this as

$$(D + \omega L)\mathbf{x} = \omega\mathbf{b} - (\omega U + (\omega - 1)D)\mathbf{x} \tag{3.4}$$

where $\omega \in (0, 2)$ is the so-called relaxation parameter. This yields a fixed-point equation by inverting the matrix $D + \omega L$ and multiplying it with the right-hand side. This gives us the following fixed-point equation:

$$\mathbf{x}^{(k+1)} = (D + \omega L)^{-1}(\omega\mathbf{b} - (\omega U + (\omega - 1)D)\mathbf{x}^{(k)}) \tag{3.5}$$

where k denotes the iteration index.
Since inverting a matrix is computationally expensive, the structure of $D + \omega L$ as a lower-triangular matrix allows us to solve this equation using forward substitution. This yields the following algorithm for the SOR method:

$$x_i^{(k+1)} = (1 - \omega)x_i^{(k)} + \frac{\omega}{a_{ii}}\left(b_i - \sum_{j<i} a_{ij}x_j^{(k+1)} - \sum_{j>i} a_{ij}x_j^{(k)}\right) \tag{3.6}$$

where i is the index of the element of $\mathbf{x}$, a_{ij} is the element of the matrix A at row i and column j. This is the final formulation of the SOR method. Setting $\omega = 1$ reduces the SOR method to the Gauss-Seidel method [24].

3.2.2 Properties of the SOR Method

The performance of the SOR method heavily depends on the choice of the relaxation parameter ω and the properties of the coefficient matrix A. Below, we discuss the key properties and considerations for the SOR method.
Arguably the most crucial property is the convergence. Unfortunately, the convergence of the SOR method is not guaranteed for all matrices A. Even if a matrix is suitable for the SOR method, it might only converge for certain values of ω. There do exist conditions which guarantee convergence for some ω [39]. For example, a symmetric positive definite matrix A converges for all $0 < \omega < 2$. Also for strictly diagonally dominant matrices, the SOR method will converge for certain values of ω [40].
The choice of ω is also crucial for the rate of convergence of the SOR method. There are two types of ω that are used. When $0 < \omega < 1$, the method is called under-relaxation. While this can be beneficial when the convergence is unstable for $\omega \geq 1$, it typically results in slower convergence. For $1 < \omega < 2$, the method is called over-relaxation. A well-chosen ω in this range can significantly accelerate the convergence of the method [41]. The optimal value of ω depends on the matrix A, but is typically difficult to determine.

3.3 Block-SOR-Newton Method

The Block-SOR-Newton method is a combination of the Newton-Raphson method and the SOR method. It is used to solve a system of nonlinear equations in the form of $\mathbf{F}(\mathbf{x}) = 0$. The method is an iterative method that uses the Newton-Raphson method to linearize the system of equations and the SOR method to solve the linearized system. We will first derive the SOR-Newton method and then work towards the Block-SOR-Newton method. The derivations are done analogously to Ref. [24].

Imagine we have the following nonlinear system of equations:

$$\mathbf{F}(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_n(\mathbf{x}) \end{pmatrix} \tag{3.7}$$

The goal is to find $\mathbf{x}$ such that $\mathbf{F}(\mathbf{x}) = 0$. We begin by using the Newton-Raphson method to linearize the system of equations to obtain equation (3.3). This can be written as:

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}^{(k)} \cdot \begin{pmatrix} \delta x_1 \\ \delta x_2 \\ \vdots \\ \delta x_n \end{pmatrix} = - \begin{pmatrix} f_1(\mathbf{x_k}) \\ f_2(\mathbf{x_k}) \\ \vdots \\ f_n(\mathbf{x_k}) \end{pmatrix} \tag{3.8}$$

Where $\boldsymbol{\delta}\mathbf{x} = \mathbf{x_{k+1}} - \mathbf{x_k}$ denotes the correction vector, and k the iteration step. Now the SOR method must be used to solve this system for $\boldsymbol{\delta}\mathbf{x}$, for simplicity of notation we take $\omega = 1$ to simplify the equations, it will be reinserted later. After decomposing the Jacobian and rearranging the terms we obtain equation (3.4):

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & 0 & \cdots & 0 \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}^{(k)} \cdot \begin{pmatrix} \delta x_1 \\ \delta x_2 \\ \vdots \\ \delta x_n \end{pmatrix}^{(m+1)} = - \begin{pmatrix} f_1(\mathbf{x_k}) \\ f_2(\mathbf{x_k}) \\ \vdots \\ f_n(\mathbf{x_k}) \end{pmatrix} - \begin{bmatrix} 0 & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ 0 & 0 & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}^{(k)} \cdot \begin{pmatrix} \delta x_1 \\ \delta x_2 \\ \vdots \\ \delta x_n \end{pmatrix}^{(m)} \tag{3.9}$$

where m denotes the iteration index of the SOR method. This equation can be decoupled into n equations, each representing the correction of one element of $\mathbf{x}$. For the i-th component, we obtain:

$$\frac{\partial f_i(\mathbf{x_k})}{\partial x_i} \cdot \delta x_i^{(k,m+1)} = -f_i(\mathbf{x_k}) - \sum_{j=1}^{i-1} \frac{\partial f_i(\mathbf{x_k})}{\partial x_j} \cdot \delta x_j^{(k,m+1)} - \sum_{j=i+1}^{n} \frac{\partial f_i(\mathbf{x_k})}{\partial x_j} \cdot \delta x_j^{(k,m)} \tag{3.10}$$

To end up with the final formulation of the Block-SOR-Newton method, we remark that the right-hand side of equation (3.10) can be interpreted as a series expansion. This yields the following algorithm for the Block-SOR-Newton method:

$$\begin{aligned}\frac{\partial f_i(\mathbf{x_k})}{\partial x_i} \cdot \delta x_i^{(k,m+1)} = -\omega f_i(x_1^{(k)} + \delta x_1^{(k,m+1)}, x_2^{(k)} + \delta x_2^{(k,m+1)}, \dots, \\ x_i^{(k)}, x_{i+1}^{(k)} + \delta x_{i+1}^{(k,m)}, \dots, x_n^{(k)} + \delta x_n^{(k,m)})\end{aligned} \tag{3.11}$$

where ω is the relaxation parameter.

Intuitively, the Block-SOR-Newton method applies the SOR method to the linearization produced by the Newton-Raphson method. Thus, we have an outer loop that iterates over the Newton-Raphson method and an inner loop that iterates over the SOR method. The advantage of the Block-SOR-Newton method is that we only have to compute the derivatives on the diagonal of the Jacobian matrix (i.e., when $i = j$). This is an improvement over the SOR-Newton method (equation (3.10)), which needs all elements of the Jacobian matrix. Therefore, applying Block-SOR-Newton can significantly reduce the computational cost of the method.

When implementing the Block-SOR-Newton method, it is important to choose the relaxation parameter ω carefully, as it can significantly affect the convergence of the method. The choice of ω is typically determined via trial and error [24].

4 The Model Description

To model the band bending profile of an adatom-covered semiconductor in a scanning tunneling microscope (STM), we must solve the semiconductor equations $F(\boldsymbol{\omega}) = \mathbf{0}$, where $\boldsymbol{\omega} = (\phi(\mathbf{x}), n(\mathbf{x}), p(\mathbf{x}))^\intercal$, given in section 2.4.3. This system of equations can only be solved numerically, which will be done using the numerical methods given in section 3. In this chapter, a framework is presented that uses a finite difference scheme to give a numerical solution to the one-dimensional Poisson equation. First, the problem is introduced in more detail, covering the STM-semiconductor setup and the motivation for the model. Then, the mathematical model is presented, which includes a discussion of the boundary conditions and the discretization of the relevant equations. Finally, an expansion of the model to include the continuity equations and multiple dimensions is briefly discussed.

4.1 Overview of the Problem

Before we introduce the mathematical model, it is important to understand the problem at hand. The model has two main objectives. First, to determine the effect of adatom-induced surface states on the band bending profile of a semiconductor. Second, to determine how the band bending profile is affected by the electrostatic potential of an STM tip. To achieve these objectives, we must develop an understanding of the physical setup of the problem. This section will explain the relevant details and considerations of the STM-semiconductor setup, as well as the motivation for the model.

4.1.1 The STM-semiconductor Setup

During STM or STS measurements, a sharp metal tip is brought close to the surface of the sample (typically a tip-sample distance between 4-7 Å [3]). This tip is set to a constant electrostatic potential ϕ_{tip}, which generates an electric field between the tip and the sample. Then, if we use a semiconductor as our sample, the band bending profile of the semiconductor is affected by the electric field generated by the STM tip.

This interaction can be modelled by considering a metal-vacuum-semiconductor system, where the metal represents the STM tip. Figure 11 displays a schematic representation of the setup. In this system, we assume that the STM tip has a constant electric potential ϕ_{tip} (where the semiconductor bulk potential is defined as $\phi = 0$), and that there are no charge carriers present in the vacuum region. ϕ_{tip} is determined by the applied voltage between the tip and the semiconductor V and the so-called local contact potential difference of the tip $\Delta\phi$ [15]:

$$\phi_{tip} = V + \frac{\Delta\phi}{q} \tag{4.1}$$

Where $\Delta\phi$ is given by [15]:

$$\Delta\phi = W_m - \chi - (E_{C,\text{bulk}} - E_F) \tag{4.2}$$

Where E_F is the Fermi level of the semiconductor, $E_{C,\text{bulk}}$ is the bulk conduction band edge of the semiconductor, W_m is the work function of the STM tip and χ is the electron affinity of the semiconductor. The electron affinity is defined as the energy difference between the vacuum level and the conduction band energy E_C.

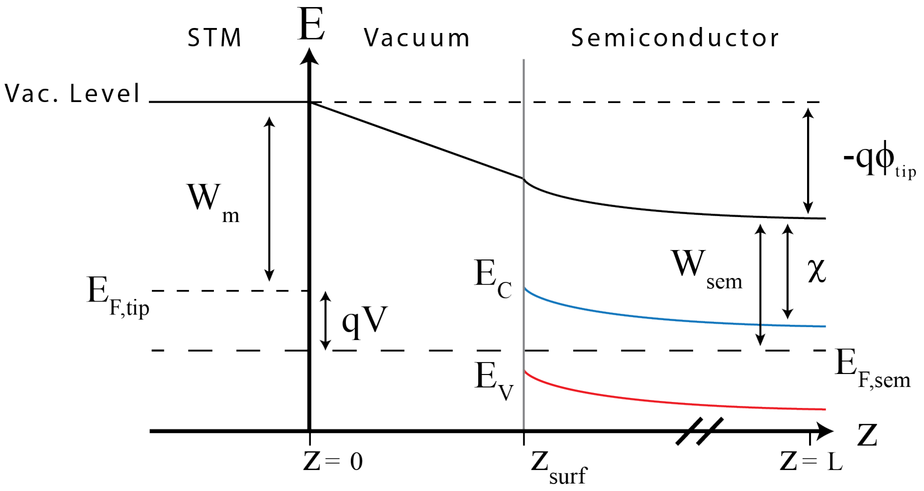

Figure 11. Schematic representation of the STM-semiconductor setup. The solid black line represents the vacuum level, which gives the energy of a stationary electron far away from the setup. The vacuum level moves with the electrostatic potential energy $-q\phi(z)$. The blue and the red lines represent the conduction and valence band edges of the semiconductor respectively. From this picture it is clear that $q\phi_{tip} = W_m - W_{sem}$, where W_m is the work function of the STM tip and $W_{sem} = \chi + (E_{C,\text{bulk}} - E_F)$ is the work function of the semiconductor, with χ being the electron affinity of the semiconductor. At the right boundary ($z = L$) the electrostatic potential ϕ is zero.

The semiconductor will screen the electric field generated by the STM tip. Compared to metals, semiconductors have a relatively small amount of charge carriers. Since charge carriers are required to screen an electric field, the electric field will extend further into the semiconductor. Therefore, unlike metals, the electric field penetrates the semiconductor and its band bending profile is affected. A reasonable measure that describes how far the electric field extends into the semiconductor is the Debye length λ_D. The Debye length is given by [20]:

$$\lambda_D = \sqrt{\frac{\varepsilon k T}{q^2 N}} \tag{4.3}$$

Where ε is the permittivity of the semiconductor, T is the temperature of the semiconductor and N is the doping concentration of the semiconductor. After a distance of λ_D from the surface, the electric field will decrease to $1/e$ of its original value. Therefore, it is critical that the thickness of the semiconductor is at least multiple Debye lengths to ensure that it is reasonable to assume that $\phi = 0$ in the bulk of the semiconductor.

For this model, we will use a GaAs(110) surface as our sample. This is mainly because of the intrinsic surface state properties. In a clean, well-cleaved GaAs(110) there is no significant band bending [23]. This means the band bending caused by the STM tip or adatom-induced surface states will be easier to observe. Moreover, GaAs is a much-studied semiconductor, which makes it easier to validate the results of the model with experimental data [16].

Understanding the STM-semiconductor setup is crucial for the model, as it provides the context in which the band bending profile of the semiconductor is determined. The band bending profile is important for accurate interpretation of STM/STS measurements, which leads us to the motivation of the model.

4.1.2 Motivation for the Model

An STM can be used to perform scanning tunneling spectroscopy (STS). During STS, the tip is fixed above a certain point on the sample and the tunneling current I is measured as a function of the applied tip-sample bias voltage V. Via this I-V relation the local density of states (LDOS) of the sample can be determined [4]. The LDOS is a measure of the number of states available to electrons at a certain energy.

However, a part of the applied voltage V is dropped within the semiconductor, which creates a tip-induced band bending profile [9]. Effectively, this changes the potential difference between the STM tip and the surface of the semiconductor. This means that the measured I-V relation is influenced by the band bending profile of the semiconductor. Therefore, to correctly determine the LDOS of the sample, a realistic estimate of the band bending profile is required.

This model aims to calculate the band bending profile to account for tip-induced voltage drop, enabling more accurate STS measurements. To address this, the model will provide a numerical solution to the relevant Poisson's equation, which can give the band bending profile of a semiconductor in thermal equilibrium and without illumination. By considering both the (adatom-induced) surface states and the presence of the STM tip, STS measurements can be modelled. This makes the model a potentially valuable tool for the interpretation of STS measurements.

4.2 Mathematical Model

This section presents the mathematical model that will be used to solve the band bending profile of a semiconductor in the presence of an STM tip. The model is based on a finite difference scheme, which will be used to solve the one-dimensional Poisson's equation. By only solving Poisson's equation, the model is only valid for semiconductors in thermal equilibrium and without illumination. First the relevant equations and the discretization of Poisson's equation are discussed, followed by the boundary conditions that are used. Then it is explained how the surface states are modelled. Finally, it is discussed how the model can be extended to solve all semiconductor equations in multiple dimensions. This extension can be used for systems where the semiconductor surface is illuminated.

4.2.1 Set of Equations and Discretization

We consider a 1D variant of the STM-semiconductor setup as described in section 4.1.1. This system has three distinct regions. First we have the STM tip, which is modelled as a point at $z = 0$ with a fixed electrostatic potential ϕ_{tip}. Then we have the vacuum region, where we assume that there are no charge carriers. Finally, we have the semiconductor region which starts with its surface at point z_{surf} and extends to $z = L$, where $L - z_{\text{surf}}$ is multiple Debye lengths.

To determine the band bending profile of the semiconductor, we need to find $\phi(z)$ by solving Poisson's equation, incorporating charge carrier densities, ionized dopants, and surface states. This is done by applying a discretization scheme to Poisson's equation. First a 1D grid from $z = 0$ to $z = L$ is defined consisting of l elements. The discretization of Poisson's equation follows the approach of Ref. [24] and Ref. [17] and is given by:

$$\frac{\dfrac{\phi_{i+1} - \phi_i}{k_i} - \dfrac{\phi_i - \phi_{i-1}}{k_{i-1}}}{\dfrac{k_i + k_{i-1}}{2}} + \frac{q}{\varepsilon}(n_i - p_i - N^+_{\text{D,i}} + N^-_{\text{A,i}}) = 0 \tag{4.4}$$

Here i is the index of the grid element, $k_i = z_{i+1} - z_i$ is the length of the grid element and ϕ_i is shorthand for $\phi(z_i)$ (idem for n_i, p_i, $N^+_{D,i}$ and $N^-_{A,i}$). By assuming parabolic bands, the charge carrier densities n_i and p_i are given by the following integrals [20]:

$$n_i = \int_{E_C - q\phi_i}^{\infty} g_c(E) f(E) dE \tag{4.5}$$

$$p_i = \int_{-\infty}^{E_V - q\phi_i} g_v(E)[1 - f(E)] dE \tag{4.6}$$

Where E_C and E_V are the bulk conduction and valence band edges of the semiconductor, $g_c(E)$ and $g_v(E)$ are the density of states in the conduction and valence band and $f(E)$ is the Fermi-Dirac distribution. These results can be written in the more convenient form [17]:

$$n_i = N_C \frac{2}{\sqrt{\pi}} F_{1/2}\left(\frac{E_F - (E_C - q\phi_i)}{kT}\right) \tag{4.7}$$

$$p_i = N_V \frac{2}{\sqrt{\pi}} F_{1/2}\left(\frac{(E_V - q\phi_i) - E_F}{kT}\right) \tag{4.8}$$

Where N_C and N_V are the effective density of states in the conduction and valence band and $F_{1/2}$ is the Fermi-Dirac integral of order $1/2$ given by [42]:

$$F_{1/2}(E_0) = \int_0^{\infty} \frac{\sqrt{E}}{1 + e^{E - E_0}} dE \tag{4.9}$$

The Fermi-Dirac integral provides an accurate way to calculate the charge carrier densities. However, for non-degenerate semiconductors, where E_F is not close to the band edges, the Fermi-Dirac integral can be approximated by the Maxwell-Boltzmann distribution (see equations (2.1-2.2)).

The concentrations of ionized donors and acceptors $N^+_{\mathrm{D,i}}$ and $N^-_{\mathrm{A,i}}$ are given by (see section 2.2):

$$N^+_{\mathrm{D,i}} = \frac{N_{\mathrm{D}}}{1 + 2\exp\left(\frac{E_F - (E_D - q\phi_i)}{kT}\right)} \tag{4.10}$$

$$N^-_{\mathrm{A,i}} = \frac{N_{\mathrm{A}}}{1 + 4\exp\left(\frac{(E_A - q\phi_i) - E_F}{kT}\right)} \tag{4.11}$$

Where N_{D} and N_{A} are the total donor and acceptor concentrations and E_D and E_A are the donor and acceptor ionization energies.

Inserting equations (4.7), (4.8), (4.10) and (4.11) into equation (4.4) gives an equation that is only dependent on the discretization of variable $\phi(z)$. This equation is valid for all grid elements i except for the element containing the semiconductor surface z_{surf}, the left boundary element $z = 0$ and the right boundary element $z = L$. The boundary conditions for these elements are discussed in the next section.

One important point that remains, is the determination of the Fermi level E_F. To ensure charge neutrality of the semiconductor, we know that the total charge deep in the semiconductor bulk must be zero. Therefore, we can find E_F by solving the following equation for E_F:

$$p_{\mathrm{bulk}}(E_F) - n_{\mathrm{bulk}}(E_F) + N^+_{\mathrm{D,bulk}}(E_F) - N^-_{\mathrm{A,bulk}}(E_F) = 0 \tag{4.12}$$

Where p_{bulk}, n_{bulk}, $N^+_{\mathrm{D,bulk}}$ and $N^-_{\mathrm{A,bulk}}$ are the charge carrier densities and ionized donor and acceptor concentrations deep in the semiconductor bulk. These can be found by setting $\phi_i = 0$ in equations (4.7), (4.8), (4.10) and (4.11), which give the charge carrier and ionized dopant densities. This equation can be solved using a root-finding algorithm.

4.2.2 Boundary Conditions

In our 1D grid, there are three points that need special attention: the left boundary $z = 0$, the right boundary $z = L$ and the point containing the semiconductor surface z_{surf}. Here the boundary conditions are presented for these points.

The left boundary models the STM tip, which has a fixed electrostatic potential ϕ_{tip}. Therefore, we introduce the Dirichlet boundary condition $\phi(z = 0) = \phi_0 = \phi_{\mathrm{tip}}$. The right boundary should be located in the bulk of the semiconductor, where we assume that the surface states and the STM tip have no influence. Therefore, the electric field should be zero at the right boundary. This is reflected by the Neumann boundary condition $\frac{d\phi}{dz}(z = L) = 0$.

Treating the semiconductor surface requires more effort, a complete derivation of the discretization can be found in the Appendix A. This is because the surface charge density σ presents a discontinuity in the electric field, therefore our discretization scheme of Poisson's equation is not directly applicable [24]. Instead, we use Gauss' law to find the boundary condition at the semiconductor surface. This gives the following discretized boundary condition at z_{surf}:

$$\varepsilon\frac{\phi_{s+1} - \phi_s}{k_s} - \varepsilon_0\frac{\phi_s - \phi_{s-1}}{k_{s-1}} - \frac{qk_s}{2}(n_s - p_s - N^-_{\mathrm{D,s}} + N^-_{\mathrm{A,s}}) + \sigma = 0 \tag{4.13}$$

Where s denotes the grid element containing the semiconductor surface. σ depends on the occupancy of the surface DOS, which is discussed in the next section.

Combining equations (4.4), (4.13) and the boundary conditions yields $\mathbf{F}(\boldsymbol{\phi}) = 0$, a set of l equations that must be solved for $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_L)^\intercal$. This is done using the Block-SOR-Newton method described in section 3.3 and will be briefly clarified in section 4.2.4.

The question remains how the surface charge density σ is determined. This is done by modelling the surface states of the semiconductor, which is discussed in the next section.

4.2.3 Surface State Model

The surface state modelling is based on the theory described in section 5.3.2. Since we use GaAs as our sample, the intrinsic surface states due to the surface dangling bonds are outside the band gap [23]. Therefore, as long as the band bending is not too large, we do not have to consider the intrinsic surface states of GaAs. This allows us to focus on the modelling of adatom-induced surface states. To model the adatom-induced surface states, a surface density of states $N_{\text{ss}}(E)$ is introduced. This DOS is assumed to be a Gaussian with a variance σ_{ss}^2 and a peak located at an energy E_{ss} with respect to the valence band maximum of the semiconductor in the bulk. The peak energy is shifted by $-q\phi_s$ in the presence of band bending. When band bending occurs, we write this as the Gaussian $E_{ss}(E + q\phi_s)$ with its peak located at E_{ss}.

The Gaussian DOS is an unrealistic assumption, as the shape of the surface DOS can be quite complex due to factors like the interaction between adatoms. For better estimates of the surface DOS, methods such as the density functional theory (DFT) can be employed [30]. However, other research indicates that sparsely distributed adatoms create sharp DOS peaks in the band gap [43–45]. Therefore, we assume that a sharp DOS peak can be adequately modelled by a Gaussian with a small variance.

The charge of the adatom-induced surface states depends on the charge neutrality level E_N, which depends on band bending. States above E_N are acceptor-like while states below E_N are donor-like, which is well illustrated in Figure 9 from section 2.3.1. This allows the surface charge density σ to be calculated by integrating the product of N_{ss} and the occupation probabilities for both the donor-like and acceptor-like states. The surface charge density is then given by [28]:

$$\sigma = q\left(\int_{E_N - q\phi_s}^{\infty} N_{\text{ss}}(E + q\phi_s) f(E - E_F) dE - \int_{-\infty}^{E_N - q\phi_s} N_{\text{ss}}(E + q\phi_s)[1 - f(E - E_F)] dE\right) \tag{4.14}$$

However, to make the model more computationally efficient, we approximate the Fermi-Dirac distribution by a step function. For a temperature of $T = 300\text{K}$ this approximation is good, except for values within 0.1 eV of the Fermi level, where the finer features become relevant. By taking ϕ_s out from the integrand and using the step function approximation we can simplify the surface charge density to:

$$\sigma = \begin{cases} q\displaystyle\int_{E_N}^{E_F + q\phi} N_{\text{ss}}(E) dE & \text{if } E_F > E_N - q\phi \\ -q\displaystyle\int_{E_F + q\phi}^{E_N} N_{\text{ss}}(E) dE & \text{if } E_F \leq E_N - q\phi \end{cases} \tag{4.15}$$

This is the surface charge density that will be used in equation (4.13). Now that all the relevant equations are presented, we can discuss how the model is solved.

4.2.4 Solving the Model

The model is based on solving $\mathbf{F}(\boldsymbol{\phi}) = 0$, a system of l equations formed by equations (4.4), (4.13) and the boundary conditions. The solution to this system is the solution vector $\boldsymbol{\phi}^*$ which gives the discretized values of $\phi(z)$. In this section we will provide an overview of the steps that the model takes to find $\boldsymbol{\phi}^*$. A flowchart that summarizes the steps can be found in Figure 12.

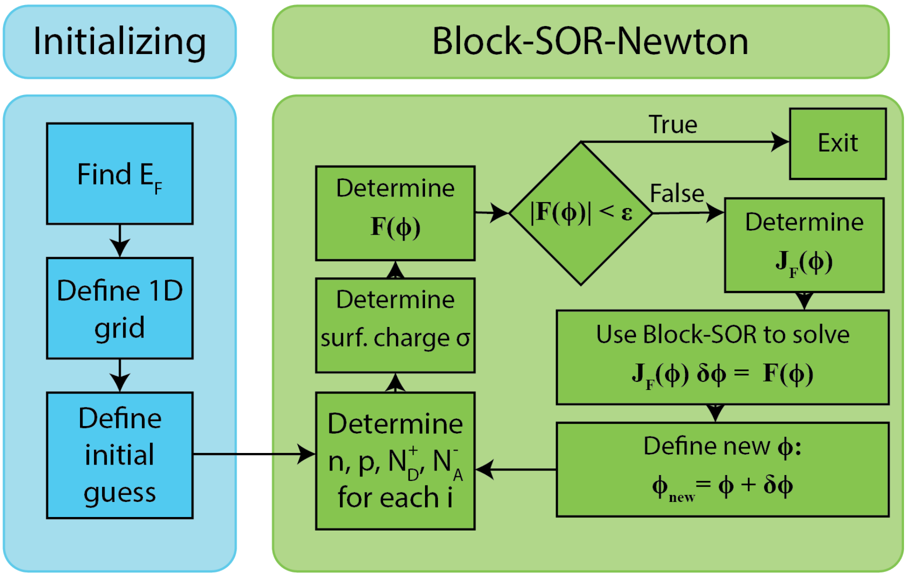

Figure 12. Flowchart that summarizes the steps the model takes to find the solution vector $\boldsymbol{\phi}^*$. The model starts with finding E_F, initializing the 1D grid and defining an initial guess. Then the model enters the Block-SOR-Newton loop, where it calculates the ϕ-dependent charge carrier densities and ionized dopant concentrations. Then the surface charge density is calculated and the residual vector $\mathbf{F}(\boldsymbol{\phi})$ is determined. If the norm of $\mathbf{F}(\boldsymbol{\phi})$ is smaller than a predefined tolerance ε, the model has found the solution and the loop is exited. Otherwise, the model determines the Jacobian matrix J_F and performs the Block-SOR algorithm to find the correction vector $\delta\boldsymbol{\phi}$. This correction is added to $\boldsymbol{\phi}$ and concludes an iteration. This repeats until the model has reached its desired accuracy.

First, the model determines the Fermi level E_F by solving equation (4.12) using the bisection method. The integrals in equations (4.7-4.8) are determined numerically. Then, the 1D grid is defined and the initial guess for $\boldsymbol{\phi}$ is set. Typically, this guess is $\boldsymbol{\phi} = \mathbf{0}$, which corresponds to a flat band bending profile. However, the model generally performs better if you can provide a guess close to the solution. Therefore, solutions to similar situations can be used as a better guess.

Then the model enters the Block-SOR-Newton method, described in section 3.3. The model first calculates the charge carrier densities n_i and p_i (equations (4.7-4.8)) and the ionized donor and acceptor concentrations $N^+_{\mathrm{D,i}}$ and $N^-_{\mathrm{A,i}}$ (equations (4.10-4.11)) for all grid elements i. Then, the surface charge density σ is calculated using equation (4.15). Next, the $\mathbf{F}(\boldsymbol{\phi})$ is calculated using equations (4.4), (4.13) and the boundary conditions. If the norm of $\mathbf{F}(\boldsymbol{\phi})$ is smaller than a predefined tolerance ε, the model has found the solution and the loop is exited. Otherwise, the model will determine a correction vector $\boldsymbol{\delta\phi}$ using the Block-SOR method.

To use the Block-SOR method, the model first determines the Jacobian matrix J_F. This is done using auto-differentiation, which is a technique that can calculate the derivative of elementary functions up to floating point accuracy [46]. However, the derivatives for equations (4.7-4.8) and (4.15) cannot be determined via auto-differentiation because of the integrals involved. The Fermi-dirac integrals in equations (4.7-4.8) can be numerically differentiated with adequate accuracy and speed [47], but this is not yet implemented in this model. Therefore, the derivatives of the carrier concentrations are neglected, which is acceptable when E_F is not near the band edges. The derivative for equation (4.15) can be determined analytically.

After the J_F is determined, the Block-SOR algorithm is used to solve $J_F(\boldsymbol{\phi})\boldsymbol{\delta\phi} = -\mathbf{F}(\boldsymbol{\phi})$ for $\boldsymbol{\delta\phi}$. This correction vector is added to $\boldsymbol{\phi}$ and concludes an iteration. This process repeats until the model converges. The model has converged when the norm of $\mathbf{F}(\boldsymbol{\phi})$ is smaller than a predefined tolerance ε. The value of ε is set to 1 in this model, which gives accurate results while maintaining feasible computational times.

The choice for the relaxation parameter ω in the Block-SOR method is important. It determines the numerical stability, as well as the speed of convergence. The optimal value for ω depends on the system and can be found by trial and error. During testing, it was found that $\omega = 0.3$ is a good starting point for this model. If one finds that the model converges too slowly, ω can be increased. However, if the model becomes unstable, ω should be decreased.

The Block-SOR-Newton method has linear convergence [48]. However, the convergence properties of the Block-SOR-Newton method applied to the semiconductor equations are not well understood [24].

4.2.5 Modelling the Semiconductor Equations in Multiple Dimensions

The model presented in this chapter is based on a 1D grid and can only handle semiconductors in thermal equilibrium. Although this can be a useful tool, it is not sufficient for most STM measurements. This is for two reasons. First, the 3D shape of the STM tip is not taken into account. This is important because the electric field generated by the STM tip is not constant over the surface of the semiconductor. Second, many STS measurements use illumination of the semiconductor, which disrupts the thermal equilibrium.

To make the model more applicable, it can be expanded to solve the full set of semiconductor equations in three dimensions. This way, also non-equilibrium situations can be modelled. In essence, the approach is the same as the 1D model. First, the semiconductor equations must be discretized for a 3D grid. An extensive derivation can be found in chapter 6 of Ref. [24]. This will yield a new set of equations $\mathbf{F}(\boldsymbol{\phi}, \mathbf{n}, \mathbf{p}) = 0$ that must be solved. The boundary conditions for the 3D grid can be found in Ref. [24]. It is recommended to first solve only Poisson's equation in 3D and use the results as an initial guess for the full

set of equations.
Furthermore, the STM tip geometry must be taken into account. This is most often done by considering the case of a hyperbolic tip to ensure cylindrical symmetry in the system [13]. Since the spatial variation of the electric field generated by the STM tip is significant, the grid must be fine enough to capture these variations. Therefore, it is important to define a non-uniform grid that is fine near the STM tip and coarser further away [17].
New challenges arise when solving for three dimensions. The number of grid elements increases cubically, which can lead to longer computational times. The Block-SOR-Newton method can still be used effectively in this situation [17]. However, it is worthwhile to investigate other methods that are more efficient for large systems of equations. Many modern methods for solving a large system defined by the discretization of partial differential equations are based on Krylov subspace methods [49]. Generally, these methods outperform classical iteration methods like the Block-SOR-Newton method [49].

5 Results and Discussion

This section explores the numerical solutions obtained by the model for band bending in adatom-covered semiconductors, introduced in Chapter 4. The primary objective is to assess the accuracy and stability of the model across different parameters, including doping concentrations, tip-sample voltages and the energetic position of the adatom states.

While the model successfully converges for certain parameter sets, it is found that numerical instabilities can occur for others. These instabilities are analyzed to identify the underlying causes and to propose potential solutions. First, an overview of the input parameters and their influences on the simulations is provided. Then, the results from successfully converged simulations are presented, followed by an analysis of the numerical instabilities. Finally, recommendations for future work are provided.

5.1 Simulation Parameters and Their Influence

The model's ability to find accurate solutions depends strongly on the choice of input parameters. Some inputs have a significant effect on the band bending, but also on the convergence properties of the model. This section will discuss the most important parameters and their influence on the simulations. A table of all constant simulation parameters is provided in Appendix B.

5.1.1 Discussion of the Simulation Grid

The accuracy and efficiency of the model depend significantly on the choice of the simulation grid. It must be fine enough to handle the large variations in potential in the space charge layer, while maintaining computational efficiency. This requires a balance between grid resolution and computational cost.

The grid can be separated into two distinct regions: the vacuum and the semiconductor. The vacuum region corresponds to the tip-sample distance, which is set to 0.925 nm, or 9.25 Å. Since the solution to the one-dimensional Poisson equation in a vacuum is a linear function, this region only requires two grid points to be fully described.

The semiconductor region should be large enough to capture the entire band bending profile. Thus, the size is set to four Debye lengths (λ_D) (Equation (4.3)). It is important to ensure sufficient resolution near the surface to accurately model the potential variations. Therefore, the first Debye length is divided into 60 grid points. The following three Debye lengths are divided into 20 grid points each, as potential variations are less significant further from the surface. This results in a total of 120 grid points for the semiconductor region. The non-uniform grid spacing allows for an accurate representation of the potential profile near the surface, while maintaining acceptable computation times. This grid setup is used for the simulations, unless stated otherwise.

A number of other grid spacings were tested to investigate their effect on the convergence properties of the model. The results show that for a converging solution, the grid spacing does not influence on the convergence rate for the simulations. This is displayed in Figure 13, where we can clearly observe that the different grid spacings all display linear convergence.

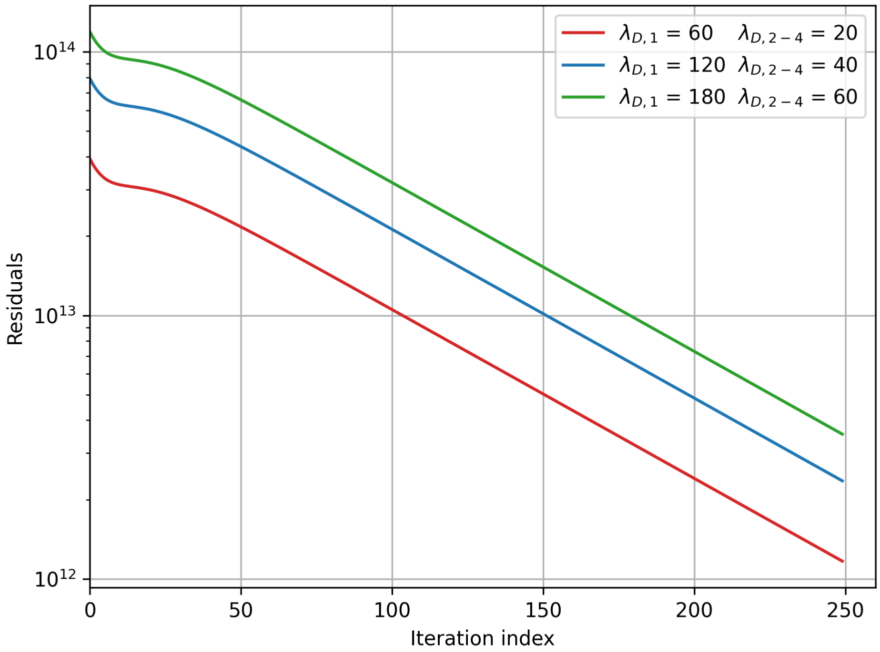

Figure 13. The residuals of the model for the first 250 iterations with different grid spacings. The simulations were conducted for an n-type GaAs(110) sample with a doping concentration of 10^8 cm^{-3}, a tip voltage of 1.5 V and no surface states. $\lambda_{D,i} = N$ corresponds to the number of grid points used to discretize the i-th Debye length. The figure shows that the grid spacing does not have an influence on the convergence rate of the simulation, it remains linear for all grid spacings. This result was reproduced for multiple parameter sets.

However, the grid spacing does seem to affect the robustness of the model. For situations with steep potential gradients, such as high doping concentrations, finer grid spacings were more likely to converge. This is likely due to the fact that the model can more accurately capture the potential variations near the surface. This phenomenon must be investigated more thoroughly before any conclusions can be drawn.

5.1.2 Key Input Parameters

This section outlines the most important input parameters in the numerical simulations. The parameters can be classified into four categories: Semiconductor, surface, STM and the numerical method parameters.

A GaAs(110) surface is used as the basis for all simulations. The material properties of GaAs can be found in Appendix B and are taken from the literature [23, 50]. The doping concentration is an important parameter, as it influences the Fermi level and the ionized dopant concentration in space charge layers. Therefore, it is varied during the simulations. The surface states also play an important role in the model. During the simulations, a Gaussian function with a variance of $\sigma_{ss}^2 = 0.01$ eV^2 is used to model sharp surface DOS

peaks. The energetic position of the surface states peak is varied to investigate its effect on the band bending profile. The charge neutrality level E_N, which determines the electronic properties of the surface states, is set to 0.7 eV above the valence band maximum.
For the STM tip, the tip voltage ϕ_{tip} is varied to investigate its effect on the band bending profile. Since we use a one-dimensional model, we do not consider the geometry of the tip. Finally, the numerical method parameters are important for the stability and accuracy of the simulations. Especially the choice of the relaxation parameter ω is crucial for the convergence of the Block-SOR-Newton method. The default value of ω is set to 0.3, as this provided relatively fast convergence for most simulations. However, it was found that for situations with high surface charges or high doping concentrations, it is necessary to decrease the value of ω to 0.1-0.2 to ensure convergence. Values of ω below 0.1 often displayed a significantly slower convergence rate, which makes the simulations computationally infeasible. Figure 14 shows the effect of the relaxation parameter on the convergence rate of the model. From the figure, it becomes clear that a good choice for ω can impact the rate of convergence greatly.

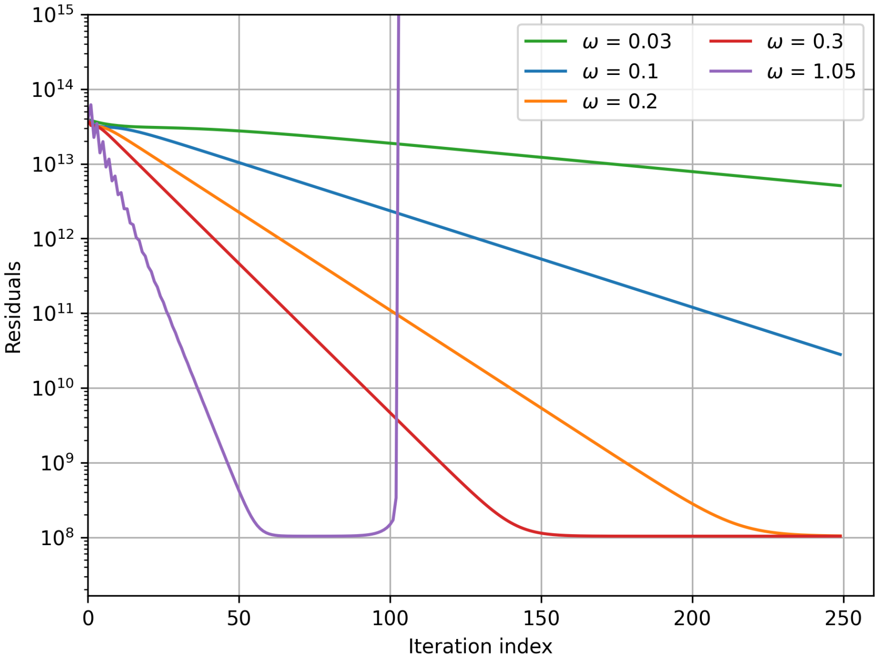

Figure 14. The residuals of the model for the first 250 iterations with different relaxation parameters ω. The simulations were conducted for an n-type GaAs(110) sample with a doping concentration of 10^8 cm^{-3}, a tip voltage of 1.5 V and no surface states. The figure shows that the convergence rate of the model is sensitive to the relaxation parameter ω. A value of $\omega = 1.05$ shows divergence, while $\omega = 0.1$ and $\omega = 0.03$ show a slow convergence rate. The figure also displays that the convergence stagnates for a residual of 10^8.

The convergence criterion is set to $\varepsilon = 1$, such that the method terminates if $|\mathbf{F}(\boldsymbol{\phi})| < \varepsilon$. This value yields accurate results, while maintaining feasible computational times. A problem that can occur is that the model stops converging and never reaches the convergence criterion, this is also displayed in Figure 14.

Sometimes, it was possible to circumvent this issue by providing a better initial guess for the potential profile. However, this was not always successful. The cause of this issue is not yet fully understood, but it seems to be related to the determined correction vector $\boldsymbol{\delta\phi}$ in the Block-SOR-Newton method. Whenever the convergence stagnated, it could be seen that for each iteration, the calculated correction vector almost exactly compensated for the previous correction vector. This led to a slow net change in the potential profile, which prevented the simulation from converging.

To further investigate the convergence properties of the model, a range of simulations were conducted for different parameter sets. The converging results are discussed in the following section, while the non-converging results are discussed in Section 5.3.

5.2 Overview of Converged Simulations

After discussing the input parameters and their influence on the simulations, it is time to evaluate the performance of the model. Simulations were conducted for a range of different tip voltages, doping concentrations and surface state peaks. This provides an indication on what parameters affect the convergence properties the most, which can be crucial for pinpointing possible improvements for the model.

When using higher doping concentrations or surface states close to the Fermi level, the model showed poor convergence properties. However, for some inputs the model successfully and consistently converged to a solution, allowing for the calculation of band bending in the semiconductor. This section provides the results obtained from these converged simulations and discusses their validity.

5.2.1 Adjusting the Tip Potential

The tip potential ϕ_{tip} is a key parameter in the model, as it is essential for determining tip-induced band bending. Therefore, the effect of the tip voltage on the band bending profile was investigated.

In this set of simulations, the tip voltage ϕ_{tip} was varied between -1.8 and 1.9 V. The GaAs sample was lightly n-doped with a doping concentration of 10^8 cm^{-3} and had a surface state density of $4.4 \cdot 10^{14}$ cm^{-2}. The energetic position of the surface state peak was set to 1.6 eV above the valence band maximum, which emulates the intrinsic surface states of GaAs(110). The results are displayed in Figure 15.

Figure 15 shows that a negative tip voltage leads to upward band bending, while a positive voltage results in downward band bending. This is in line with band bending theory, since a negative tip voltage creates an electric field that points towards the STM tip. Therefore, upward band bending occurs to form a positive depletion region that screens the electric field generated by the STM tip.

The results also show that the effect of the tip voltage on the band bending profile is nonlinear. For tip potentials between -0.7 and 0.3 V, the amount of band bending is sensitive to small changes in the tip voltage. However, for tip potentials outside this range, the effect of the tip voltage on the band bending profile seems to saturate. This is in

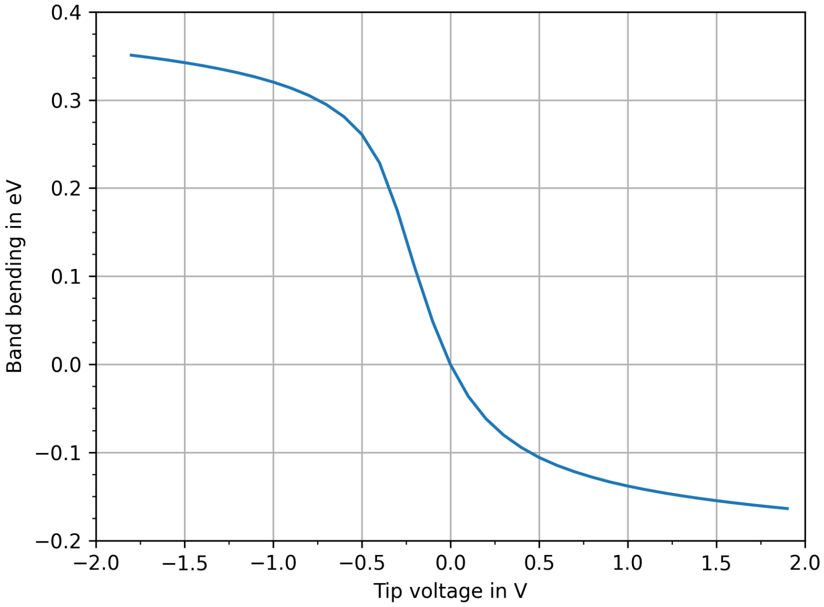

Figure 15. Total amount of band bending in doped GaAs(110) for tip voltages between -1.8 and 1.9 V. The GaAs sample is lightly n-doped with a doping concentration of 10^8 cm^{-3}. The figure displays that a negative tip voltage leads to upward band bending, while a positive voltage results in downward band bending. Moreover, the total amount of band bending is sensitive to small changes in the tip voltage for tip potentials between -0.7 and 0.3 V, and saturates for tip potentials outside this range.

accordance with other studies that determined the relation between tip voltage and band bending in n-doped GaAs(110) [51]. This saturation happens because beyond a certain point, most mobile carriers near the surface have been rearranged. Therefore, supplying a larger bias voltage has diminishing impact on the band bending.

However, it seems that the model might underestimate the amount of band bending if we compare it to other studies. de Raad *et al.* [51] have found band bending values of up to 1.5 eV for a tip-sample bias of -2 V with a tip work function of 4.6 eV using n-type GaAs with a doping concentration of 10^{18} cm^{-3}. The origin of this discrepancy is not yet clear, one explanation is that de Raad *et al.* also accounted for the geometry of the STM tip, while this model only works with one dimension. However, this would actually likely lead to an overestimation of the band bending by the 1D model according to literature [13, 52]. Another explanation is that the current grid setup might not be fine enough to capture the potential variations near the surface. Especially for space charge layers that do not extend far into the bulk, it is important to have a fine grid near the surface.

To investigate this further, a simulation was conducted with a finer grid near the surface for a tip voltage of -2.8 V. The first Debye length was divided into 540 grid points, while the following three Debye lengths were divided into 100 grid points each. While computation time was too long to conduct a full simulation, the results showed that the band bending profile was significantly steeper near the surface compared to the original grid setup. Specifically, the fine grid seemed to settle on a total band bending of around 0.5 eV, while the original grid setup showed a total band bending of 0.38 eV. So, the change in grid size increased the total band bending by 0.12 eV. This does not fully explain the discrepancy with literature, but it does suggest that the grid setup has a significant influence on the numerically obtained band bending profile.

5.2.2 Influence of Small Doping Concentrations

Another important parameter in the model is the doping concentration, as it determines the Fermi level and the ionized dopant concentration in the space charge layer. To investigate the effect of doping concentration on the band bending profile, simulations were conducted for doping concentrations between 10^8 and 10^{11} cm^{-3}. Doping concentrations beyond 10^{11} cm^{-3} are not included, as the model diverged or displayed slow convergence for these values. Concentrations below 10^8 cm^{-3} are not considered since the doping concentration for most applications is typically much higher. The GaAs sample had a surface state density of $4.4 \cdot 10^{14}$ cm^{-2} with the peak at 1.6 eV above the valence band minimum. The simulations were conducted for a tip voltage of -0.5 and 1.5 V. The results are shown in Figure 16.

The results in Figure 16 show that the amount of band bending decreases with increasing

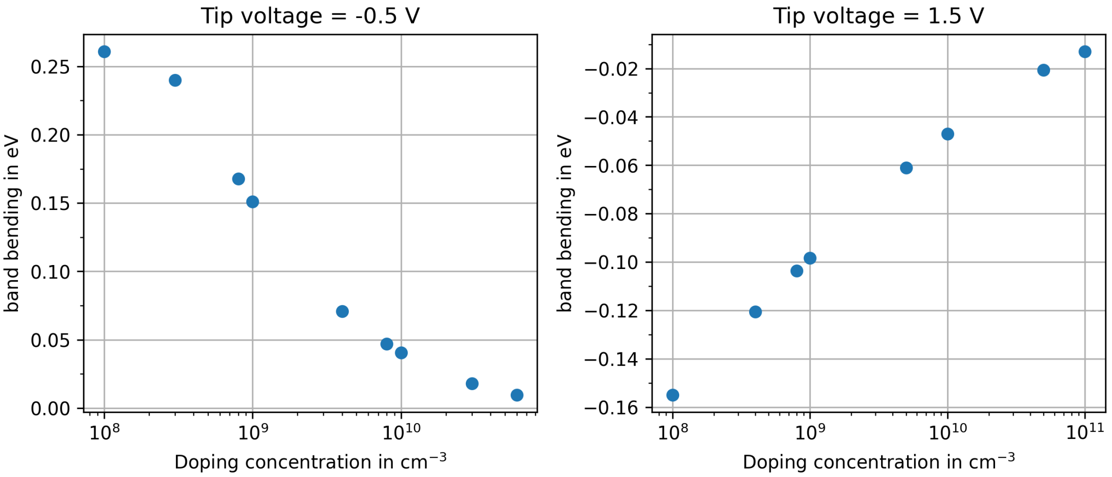

Figure 16. Total amount of band bending in n-doped GaAs(110) for doping concentrations between 10^8 and 10^{11} cm^{-3}. In the left figure there is a tip voltage of -0.5 V and in the right figure there is a tip voltage of 1.5 V. The figure shows that the amount of band bending decreases with increasing doping concentration. This is in line with band bending theory

doping concentration. This is in line with band bending theory, as higher doping concentrations lead to a larger number of ionized dopants in a depletion layer, as well as a larger number of electrons in an accumulation layer. This means that the space charge layer can more effectively screen the electric field generated by the STM tip, resulting in less band

bending.
For doping concentrations above 10^{10} cm^{-3}, we see that the absolute amount of band bending is below 0.05 V. This is lower than the band bending values found in the literature [51, 52]. This discrepancy could be due to the same reasons as discussed in the previous section. The grid setup might not be fine enough to capture the potential variations near the surface, especially for space charge layers that do not extend far into the bulk. This phenomenon should be investigated further to determine the exact cause of the discrepancy.
For doping concentrations above 10^{11} cm^{-3}, the model showed poor convergence properties. Either the convergence was slow or the model did not converge at all. This will be discussed in more detail in the following section.

5.3 Convergence Issues

While the model successfully converged for several parameter sets, certain simulations faced significant convergence issues. In particular, high doping concentrations and surface states close to the Fermi level caused the model to diverge or display slow convergence. This section will discuss the potential causes of these convergence issues and propose possible solutions.

5.3.1 High Doping Concentrations

As discussed in section 5.2.2, for concentrations above 10^{11} cm^{-3}, the model had difficulties with converging. When we introduce higher doping concentrations, the Fermi level moves closer to one of the band edges. This has two important implications for the model. Firstly, this can lead to higher charge carrier concentrations in accumulation layers, but also a higher concentration of ionized dopants in depletion layers. This leads to regions in the semiconductor with a considerable amount of net charge. Second, the screening of the electric field by the space charge layer becomes more effective due to the greater charge concentration, narrowing the space charge layer. This leads to a steeper potential gradient near the surface.
The first implication can be problematic for the model, this can be seen by considering the discretized Poisson's equation (Equation (4.4)). If we insert a high amount of charge in a grid point i, the residual at that point $F_i(\boldsymbol{\phi})$ can be large. This may lead to a large correction vector $\boldsymbol{\delta\phi}$, which can cause the method to overshoot. The next iteration will then try to compensate for this overshoot, leading to an overshoot in the opposite direction. This creates diverging oscillatory behavior in the potential profile, which prevents the model from converging. This phenomenon is illustrated in Figure 17.

One possible solution to this issue is to decrease the relaxation parameter ω, as this will lead to smaller correction vectors. However, this also leads to significantly slower convergence rates as illustrated in Figure 14. For example, for a doping concentration of 10^{12} cm^{-3}, the model only started to converge for $\omega = 0.005$. This is computationally infeasible, as the model would require thousands of iterations to converge.
One work-around that might be promising is to slowly increase the relaxation parameter. This was done by starting with a low value of $\omega = 0.005$ and allowing the model to converge

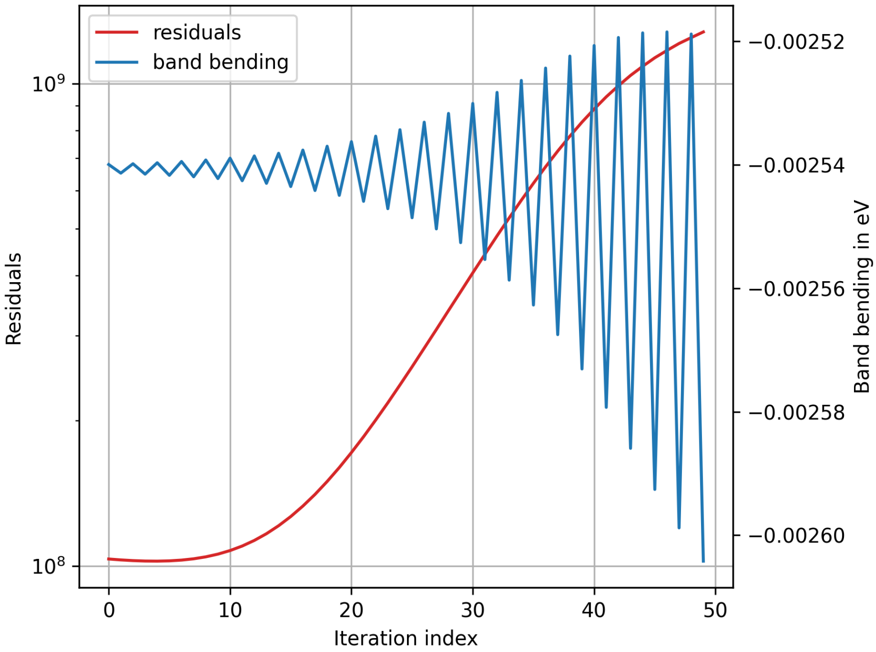

Figure 17. The residuals of 50 iterations. The simulations were conducted for an n-type GaAs(110) sample with a tip voltage of -1.5 V, a doping concentration of 10^{12} cm^{-3} and no surface states with a relaxation parameter of $\omega = 0.03$. The simulation had an initial guess provided by the same situation, but with a relaxation parameter of $\omega = 0.005$. The figure displays the oscillatory behaviour of the band bending profile. The oscillations start small, but quickly grow in amplitude. This is accompanied by an increase in the residuals, which indicates that the model is diverging.

for a few hundred iterations. Then, the relaxation parameter was increased slightly and the model was allowed to converge again. Using this method, it was possible to increase the relaxation parameter to a value of 0.07 before the model started to diverge. This is a significant improvement compared to the initial situation, but it is still computationally infeasible. Therefore, it is important to also investigate other potential solutions to this issue.

The second implication of high doping concentrations is that the space charge layer becomes narrower, leading to a steeper potential gradient near the surface. This can be problematic for the model, as the grid spacing might not be fine enough to capture the potential variations near the surface. According to Feenstra *et al.*, it is recommended to start with a coarse grid to get the global shape of the band bending profile, and then use this solution as an initial guess on a finer grid [53]. This will result in an accurate description of the band bending profile, without compromising on convergence properties [53]. Unfortunately, this method could not be tested for the current model, as the model did not converge for high doping concentrations. However, this could be promising for future work, as it accurately captures the potential variations near the surface.

5.3.2 Surface States Close to the Fermi Level

The energetic position of the surface DOS plays a crucial role in the band bending profile. If the surface states become charged, a significant amount of charge can accumulate in the surface states. This induces band bending, since the surface charge must be compensated by the space charge layer.

We will analyze the model's response to a situation with a DOS peak at 1.0 eV. There is no tip voltage and the GaAs sample is lightly n-doped with a doping concentration of 10^8 cm^{-3}. This simulation diverges at iteration step 52 when using $\omega = 0.03$. Figure 18 displays the total amount of band bending and the surface charge. The residual and Jacobian associated with the surface grid point are displayed in Figure 19.

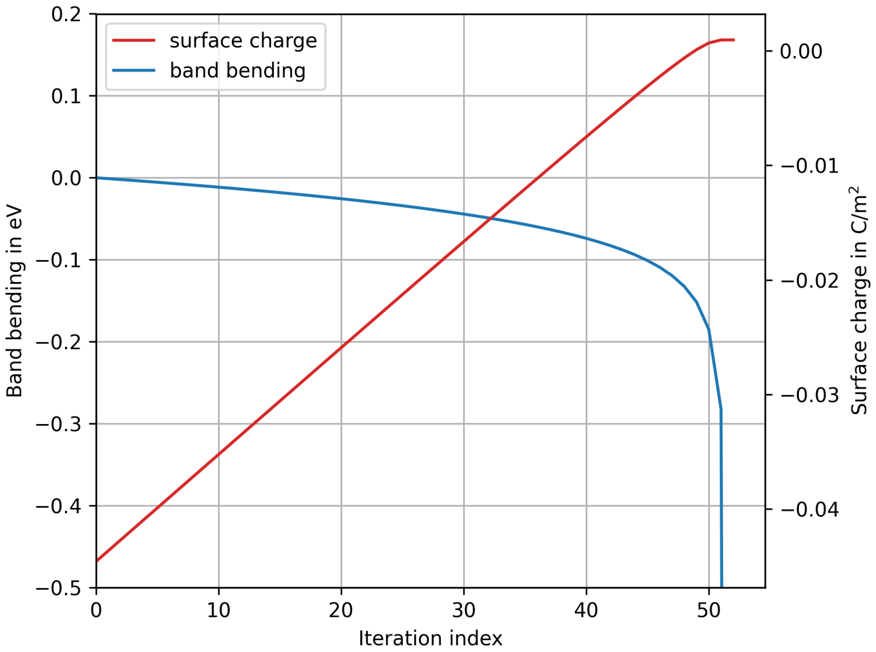

Figure 18. The total amount of band bending and the surface charge for a GaAs(110) sample with a surface state peak at 1.0 eV with a surface state density of $4.4 \cdot 10^{14}$ cm^{-2}. The GaAs sample is lightly n-doped with a doping concentration of 10^8 cm^{-3}. The figure shows that the band bending starts to increase rapidly around iteration step 52. This overshoot caused the simulation to diverge.

Figure 19 explains why our results starts to diverge. To understand this, it is important to recap equation (3.11). It tells us that the size of the correction at grid point i is proportional to the residual at that point and inversely proportional to the i^{th} diagonal element of the Jacobian matrix. Therefore, if the i^{th} diagonal element decreases, the correction at grid point i will increase.

This is exactly what happens in Figure 19 for the surface grid point. In iteration step 51 the associated Jacobian suddenly dropped from $5 \cdot 10^{12}$ to 10^{11}, while the residual remained almost constant. This caused the correction to increase with a factor close to 50 in the

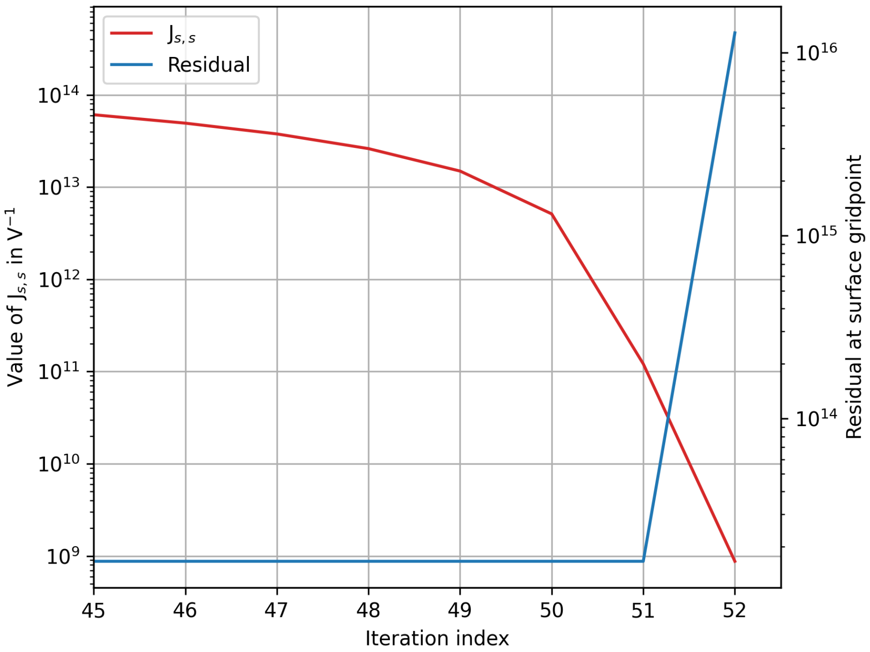

Figure 19. The residuals of the surface grid point and the entry $J_{s,s}$ of the Jacobian matrix associated with the surface grid point. The GaAs sample is lightly n-doped with a doping concentration of 10^8 cm^{-3} and has a surface state peak at 1.0 eV with a surface state density of $4.4 \cdot 10^{14}$ cm^{-2}. The figure shows that the residuals start to increase rapidly at iteration step 52. It also shows that $J_{s,s}$ decreases rapidly from $5 \cdot 10^{12}$ to 10^{11} from iteration step 50 to 51. This resulted in a large correction at the surface grid point, which resulted in an overestimation of the band bending. This caused the simulation to diverge.

next iteration. This led to an overshoot, which is clearly visible in Figure 18. This results in the same oscillatory behavior as discussed in Section 5.3.1.

The reason for this sudden drop in the Jacobian is rooted in the fact that a sharp Gaussian peak describes the surface DOS. To obtain the surface charge, the DOS is integrated. Therefore, the derivative of the surface charge with respect to the surface potential has the same function as the surface DOS (i.e., a sharp Gaussian). This means that for a small change in the surface potential, the Jacobian can make a large jump. In the case of Figure 19, this resulted in a sudden drop in the Jacobian.

To test this hypothesis, the simulation was conducted again, but the variance of the Gaussian peak was increased to $\sigma_{ss}^2 = 0.3$ eV^2. In this scenario, the model did not diverge, but the convergence was sub-linear and eventually stagnated. This is displayed in Figure 20. We see that the amount of band bending seems to settle after 300 iterations, while the residual settles after 500 iterations. This suggests that the sharp Gaussian is problematic for the model.

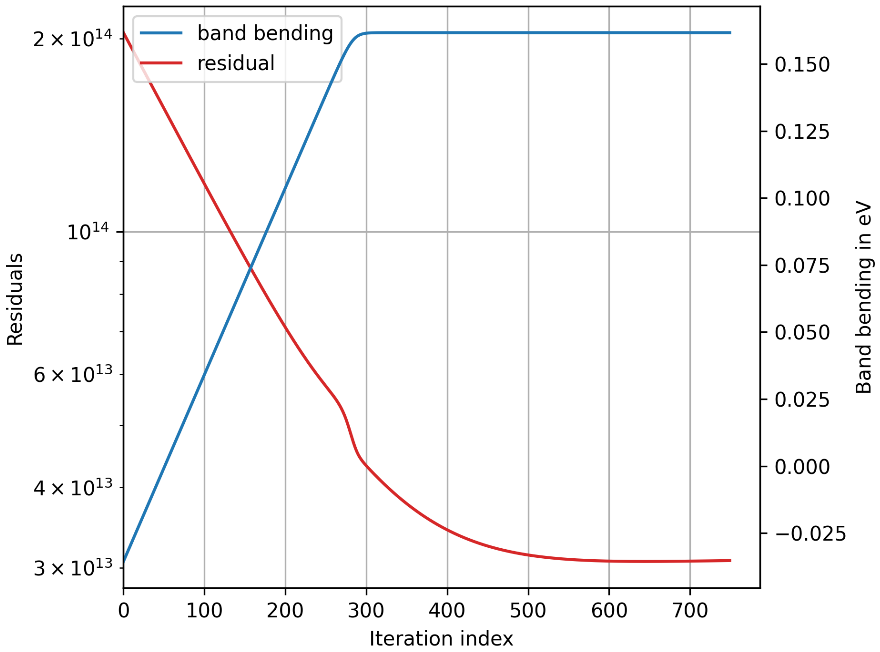

Figure 20. The residuals for 750 iterations. The GaAs sample is lightly n-doped with a doping concentration of 10^8 cm^{-3} and a surface state peak at 1.0 eV, a variance of $\sigma_{ss}^2 = 0.3$ eV^2 and a surface state density of $4.4 \cdot 10^{14}$ cm^{-2}. The figure shows that the simulation first converges linearly, but slows down after 300 iterations. The residuals settle after 500 iterations, which indicates that the model stagnates. The band bending profile seems to settle after 300 iterations.

5.4 Recommendations for Future Work

We have discussed the results obtained from the model and the convergence issues that were encountered. For some parameter sets, the model successfully converged to a solution. However, for other parameter sets, the model had difficulties with converging. This section provides recommendations for future work to improve the convergence properties of the model.

5.4.1 Improvements in Numerical Methods

In our model, the Block-SOR-Newton method was used to solve the Poisson equation. This is in accordance with the state-of-the-art band bending simulations developed by Feenstra *et al.* and Schnedler *et al.* [17, 53]. However, most modern solvers for other applications utilize more advanced methods to solve systems of nonlinear equations. This section will discuss potential improvements to the numerical methods used in the model.

First we will start with considering improvements for the outer Newton iterations. In the current implementation, the standard Newton method is used for each step. However, Newton's method can be sensitive to the initial guess, which can lead to slow convergence, or even divergence. In modern literature, Newton's method is often adjusted to improve the convergence properties of the method. Two important modifications are the damped Newton's method and quasi-Newton methods [35, 37]. This might significantly improve the convergence properties of the model, while also being a novel addition to the current state-of-the-art band bending simulations.

Next, we will consider improvements for the inner SOR iterations. In modern literature, the SOR method is often replaced by more advanced iterative methods, many based on Krylov subspace methods [49]. These methods often outperform classical iterative methods like SOR, especially for large systems of equations [49]. Therefore, it could be beneficial to investigate the use of Krylov subspace methods in the model. This would also be a novel addition to the current state-of-the-art band bending simulations.

Combining Newton's method with Krylov subspace methods is a powerful combination for solving systems of nonlinear equations. This is known as a Newton-Krylov method, and it is often used in modern scientific computing [54]. It would be interesting to investigate whether implementing a Newton-Krylov solver can improve the convergence properties of tip-induced band bending simulations. This could be done by comparing the convergence properties of current programs like *SEMITIP* with a model that uses a Newton-Krylov solver. This would provide valuable insights into the effectiveness of these methods for band bending simulations.

5.4.2 Improvements in the Model Description

The results showed that the model had difficulties with converging for certain parameter sets. This was especially the case for high doping concentrations and surface states close to the Fermi level. This section will discuss potential improvements to the model description to improve the convergence properties of the model.

First, we will consider improvements for the description of high doping concentrations. As discussed in Section 5.3.1, the model had difficulties with converging for doping concentrations above 10^{11} cm^{-3}. This was mainly due to overshoot of the correction vector in the Block-SOR-Newton method. One possible solution to this issue is to decrease the relaxation parameter ω, as this will lead to smaller correction vectors. However, this also leads to significantly slower convergence rates.

Another possible solution is to adjust the Jacobian matrix. In the current implementation of the model, the charge carriers are determined via Fermi-Dirac integrals. This makes it difficult to determine the derivative of the charge carriers with respect to the potential. Therefore, these derivatives are neglected in this implementation, which is reasonable when E_F is not close to the band edges. However, for high doping concentrations, this assumption does not hold. Therefore, it might be beneficial to investigate whether it is possible to determine an approximation of the derivatives of the charge carriers with respect to the potential. This would allow for a more accurate description of the Jacobian matrix, which could improve the convergence properties of the model.

Next, we will consider improvements for the description of surface states close to the Fermi level. As discussed in Section 5.3.2, the model had difficulties with converging for surface states close to the Fermi level. This was hypothesized to be due to the sharp Gaussian peak that describes the surface DOS. For wider Gaussians, the model did not diverge anymore, but it did stagnate. This suggests that the sharp Gaussian peak might be problematic for the model. One possible solution could be to provide a better initial guess for the potential profile. This would decrease the step size of the correction vector, which could prevent the model from diverging.

Lastly, we will consider improvements for the grid setup. As discussed in Section 5.2.1, the grid setup might not be fine enough to capture the potential variations near the surface. Therefore, it is suggested to further develop the non-uniform grid spacing near the surface. This would allow for an accurate representation of the potential profile near the surface, while maintaining acceptable computation times. One important aspect to consider is that sudden changes in grid spacing can lead to numerical instabilities. Therefore, it is important to ensure a smooth transition between different grid spacings. This could be achieved by using a smooth transition function that gradually changes the grid spacing.

5.4.3 Validating the Model

If the model is improved according to the recommendations provided in this section and it converges for a wide range of parameter sets, it is important to validate the model. This can be done in two ways. First, the model can be validated against other band bending models like *SEMITIP*. This would provide insights about the accuracy of the model in handling the tip-induced band bending. However, to also validate the adatom-induced band bending, it is important to compare the model to experimental data. This might also give insights in the limitations of the model description and provide directions for further improvements.

6 Conclusion

Accurate characterization of semiconductor electronic properties at atomic resolution is essential for advancing device technology. This thesis focused on developing a model to correct for tip-induced band bending (TIBB) in adatom-covered semiconductors, since TIBB significantly hinders the interpretation of Scanning Tunneling Spectroscopy (STS) data. By developing a model based on the work of Schnedler *et al.* and Feenstra *et al.*, we set out to create a numerical tool that is capable of describing the effects of adatom-induced surface states, doping, and STM tip potentials on the band bending in semiconductors.

A review of the relevant theory was provided, covering the role of doping, charge carriers, and surface states in band bending. It was explained how solving the Poisson equation can determine the band bending profile for semiconductors in thermal equilibrium. Particular attention was given to the role of surface states in modifying the band bending profile, and how this can be modelled quantitatively.

Using this theoretical framework, a numerical model was developed that solves the Poisson equation self-consistently in a one-dimensional STM-semiconductor system. The Block-SOR-Newton method was used to solve the nonlinear system of equations arising from the discretization of the Poisson equation. The adatom-induced surface states were modelled as a narrow Gaussian Density of States (DOS). We showed that the model converges under certain parameter sets, particularly for low doping concentrations and surface state energies far from the Fermi level. The model was able to accurately describe the effect of the STM tip voltage on the band bending. However, the model showed numerical instabilities for other parameter sets, which were analyzed to identify the underlying causes.

High doping concentrations and surface state energies close to the Fermi level were identified as important factors that can lead to numerical instabilities. These instabilities were often attributed to overshooting of the Newton method, insufficient grid resolution, or stagnation of the convergence process. To address these issues, we proposed several potential solutions, such as using a Newton-Krylov method, implementing adaptive grid refinement, and altering the derivatives related to the charge carriers.

Looking forward, the model can increase its applicability by introducing the full set of semiconductor equations, which allows for the study of non-equilibrium situations. This is particularly relevant for investigating illuminated semiconductors, where the photo-generated carriers alter the band bending profile. Furthermore, the model can be extended to include the effects of the STM tip geometry by expanding the model to three dimensions.

Ultimately, the model developed in this thesis provides a solid foundation, but further work is needed to address the numerical instabilities and increase the model's applicability. By introducing modern numerical techniques and expanding the model to include more physical effects, the model can become a powerful tool for refining STS data interpretation, ultimately deepening our understanding of semiconductor devices at the atomic scale.

A Derivation of the Surface Boundary Condition

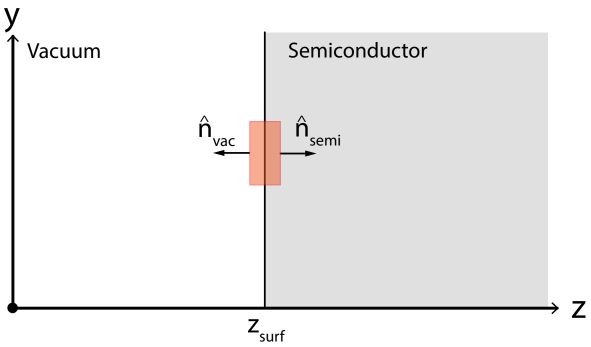

Figure 21. Sketch of the vacuum-semiconductor interface. The orange rectangle represents the pillbox used in the derivation of the surface boundary condition.

First, let us find an analytical expression for the surface boundary condition. We assume a non-negligible surface charge density σ on z_{surf}. We also assume a perfect vacuum with electric permittivity ϵ_0 and a semiconductor with electric permittivity ε_{semi}. Since we are interested in the potential near z_{surf}, we will tackle the problem by using Gauss' law in integral form. To do this, we create a pillbox with surface area A and thickness $\varepsilon \rightarrow 0$ (see Fig. 21). We then apply Gauss' law to this volume. Gauss' law states that

$$\oiint_S \mathbf{D} \cdot d\mathbf{a} - Q_{free} = 0 \tag{A.1}$$

With $\mathbf{D}$ being the electric displacement field and Q_{free} being the enclosed free charge. On our pillbox we define two unit normal vectors - $\mathbf{\hat{n}_{vac}}$ and $\mathbf{\hat{n}_{semi}}$ - that point into the direction of the vacuum and semiconductor respectively. Also note that $\mathbf{\hat{n}_{vac}} = -\mathbf{\hat{n}_{semi}}$. It can be easily shown that $Q_{free} = A\sigma$ in our case. Now we can simplify eq. (A.1) to

$$\begin{gathered} A(\mathbf{D_{vac}} \cdot \mathbf{\hat{n}_{vac}} + \mathbf{D_{semi}} \cdot \mathbf{\hat{n}_{semi}}) - A\sigma = \\ (\mathbf{D_{vac}} - \mathbf{D_{semi}}) \cdot \mathbf{\hat{n}_{vac}} - \sigma = (\mathbf{D_{semi}} - \mathbf{D_{vac}}) \cdot \mathbf{\hat{n}_{semi}} - \sigma = 0 \end{gathered} \tag{A.2}$$

For isotropic, homogeneous, linear media, the electric displacement can be written as:

$$\mathbf{D} = -\varepsilon \nabla \phi \tag{A.3}$$

where ε is the electric permittivity of the medium in which the field exists and ϕ is the electric potential. Finally, by taking $\mathbf{n_{\hat{semi}}}$ as our normal vector and inserting eq. (A.3) into eq. (A.2) we obtain:

$$\varepsilon_0 \frac{\partial \phi}{\partial z}\Big|_{vac} - \epsilon_{semi} \frac{\partial \phi}{\partial z}\Big|_{semi} - \sigma = 0 \tag{A.4}$$

Now we will continue by deriving the discretization of equation (A.4). We assume that our interface is located at $z = z_s$, such that $z < z_s$ denotes the insulator and $z > z_s$ denotes the semiconductor. Now we will replace the partial derivatives with their difference expressions by doing a Taylor expansion. We start out with the following expression:

$$\phi(z) = \phi(z_i) + (z - z_i)\frac{\partial \phi}{\partial z} + \frac{(z - z_i)^2}{2}\frac{\partial^2 \phi}{\partial z^2} + \mathcal{O}((z - z_i)^3) \tag{A.5}$$

Now we introduce the shorthand notation $k_i = z_{i+1} - z_i$, (such that $z_{i-1} - z_i = -k_{i-1}$). We insert $z = z_{s-1}$, which lies in the vacuum, into equation (A.5) and rearrange the equation to obtain:

$$\frac{\partial \phi}{\partial z}\Big|_{vac} = \frac{\phi(z_s) - \phi(z_{i-s})}{k_{s-1}} + \frac{k_{s-1}}{2}\frac{\partial^2 \phi}{\partial z^2}\Big|_{vac} + \mathcal{O}(k_{s-1}^2) \tag{A.6}$$

in completely similar fashion, but instead inserting $z = z_{s+1}$, which lies in the semiconductor, we obtain:

$$\frac{\partial \phi}{\partial z}\Big|_{semi} = \frac{\phi(z_{s+1}) - \phi(z_s)}{k_s} - \frac{k_s}{2}\frac{\partial^2 \phi}{\partial z^2}\Big|_{semi} + \mathcal{O}(k_s^2) \tag{A.7}$$

To eliminate the second order derivatives, we substitute the derivatives with the Poisson equation:

$$\frac{\partial^2 \phi}{\partial z^2}\Big|_{vac} = 0 \tag{A.8}$$

$$\frac{\partial^2 \phi}{\partial z^2}\Big|_{semi} = \frac{q}{\epsilon_{semi}}(n_s - p_s - N_{D,s}^- + N_{A,s}^-) \tag{A.9}$$

substituting eqs. (A.8) and (A.9) respectively into (A.6) and (A.7), and re-injecting them into eq. (A.4) we obtain:

$$\varepsilon\frac{\phi(z_{s+1}) - \phi(z_s)}{k_s} - \varepsilon_0\frac{\phi(z_s) - \phi(z_{s-1})}{k_{s-1}} - \frac{qk_s}{2}(n_s - p_s - N_{\mathrm{D,s}}^- + N_{\mathrm{A,s}}^-) + \sigma = 0 \tag{A.10}$$

Which is our final expression for the surface boundary condition.

B Simulation Parameters

In this section, we provide an overview of the input parameters that were held constant throughout the simulations. An overview of all the parameters that remained constant throughout the simulations is provided in tables below.

Temperature (K)	300
Energy band gap (eV)	1.424
Effective conduction band DOS (cm^{-3})	$4.7 \cdot 10^{17}$
Effective valence band DOS (cm^{-3})	$9.0 \cdot 10^{18}$
Relative permittivity ε_r	12.9
Intrinsic Debye length (μm)	2250
Donor band energy (eV)	1.394
Acceptor band energy (eV)	0.025
Charge neutrality level (eV)	0.7

Table 1. Overview of the material properties of the GaAs sample used in the simulations [23, 50].

Variance (eV^2)	0.01
Density of surface states (m^{-2})	$4.4 \cdot 10^{18}$

Table 2. Overview of the surface Density Of State properties. The DOS was modelled as a Gaussian.

References

[1] G. Binnig, H. Rohrer, C. Gerber and E. Weibel, *Tunneling through a controllable vacuum gap*, Applied Physics Letters **40** (1982) 178.

[2] G. Binnig, H. Rohrer, C. Gerber and E. Weibel, *Surface studies by scanning tunneling microscopy*, Phys. Rev. Lett. **49** (1982) 57.

[3] C.J. Chen, *Introduction to Scanning Tunneling Microscopy*, Oxford University Press, New York, NY, 2nd ed. (2008).

[4] B. Voigtländer, *Scanning tunneling spectroscopy (STS)*, NanoScience and Technology **69** (2015) 309.

[5] D. Kitchen, A. Richardella, J.M. Tang, M.E. Flatté and A. Yazdani, *Atom-by-atom substitution of Mn in GaAs and visualization of their hole-mediated interactions*, Nature **442** (2006) 436.

[6] F. Marczinowski, J. Wiebe, F. Meier, K. Hashimoto and R. Wiesendanger, *Effect of charge manipulation on scanning tunneling spectra of single Mn acceptors in InAs*, Phys. Rev. B **77** (2008) 115318.

[7] K. Schouteden, J. Debehets, D. Muzychenko, Z. Li, J.W. Seo and C. Haesendonck, *Adsorption of te atoms on Au(111) and the emergence of an adatom-induced bound state*, Journal of Physics Condensed Matter **29** (2017) .

[8] M. Gastaldo, J. Varillas, A. Rodríguez, M. Velický, O. Frank and M. Kalbáč, *Tunable strain and bandgap in subcritical-sized MoS_2 nanobubbles*, npj 2D Materials and Applications **7** (2023) 71.

[9] R. Feenstra, *A prospective: Quantitative scanning tunneling spectroscopy of semiconductor surfaces*, Surface Science **603** (2009) 2841.

[10] R.M. Feenstra and J.A. Stroscio, *Tunneling spectroscopy of the GaAs(110) surface*, *Journal of Vacuum Science & Technology B: Microelectronics Processing and Phenomena* **5** (1987) 923.

[11] R. Feenstra, J.A. Stroscio and A. Fein, *Tunneling spectroscopy of the Si(111)2 × 1 surface*, *Surface Science* **181** (1987) 295.

[12] R. Dombrowski, C. Steinebach, C. Wittneven, M. Morgenstern and R. Wiesendanger, *Tip-induced band bending by scanning tunneling spectroscopy of the states of the tip-induced quantum dot on InAs(110)*, *Phys. Rev. B* **59** (1999) 8043.

[13] R.M. Feenstra, *Electrostatic potential for a hyperbolic probe tip near a semiconductor*, *Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena* **21** (2003) 2080.

[14] R.M. Feenstra, S. Gaan, G. Meyer and K.H. Rieder, *Low-temperature tunneling spectroscopy of Ge(111)c(2x8) surfaces*, *Phys. Rev. B* **71** (2005) 125316.

[15] R.M. Feenstra, Y. Dong, M.P. Semtsiv and W.T. Masselink, *Influence of tip-induced band bending on tunnelling spectra of semiconductor surfaces*, *Nanotechnology* **18** (2006) 044015.

[16] N. Ishida, K. Sueoka and R.M. Feenstra, *Influence of surface states on tunneling spectra of n-type GaAs(110) surfaces*, *Phys. Rev. B* **80** (2009) 075320.

[17] M. Schnedler, V. Portz, P.H. Weidlich, R.E. Dunin-Borkowski and P. Ebert, *Quantitative description of photoexcited scanning tunneling spectroscopy and its application to the GaAs(110) surface*, *Phys. Rev. B* **91** (2015) 235305.

[18] M. Schnedler, *Quantitative scanning tunneling spectroscopy of non-polar III-V compound semiconductor surfaces* (07, 2015).

[19] C. Kittel, *Introduction to Solid State Physics*, Wiley, New York, NY, 8th ed. (2005).

[20] S. Sze and K. Ng, *Physics of Semiconductor Devices*, Wiley, New York, NY, 2nd ed. (1981).

[21] S.D. Parker, R.L. Williams, R. Droopad, R.A. Stradling, K.W.J. Barnham, S.N. Holmes et al., *Observation and control of the amphoteric behaviour of Si-doped InSb grown on GaAs by MBE*, *Semiconductor Science and Technology* **4** (1989) 663.

[22] D. Tjeertes, *Doping control at the atomic scale in III-V semiconductors*, doctoral dissertation, Eindhoven University of Technology, 2023.

[23] H. Luth, *Solid Surfaces, Interfaces and Thin Films*, Springer, Berlin, 5th ed. (2001).

[24] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*, Springer, Vienna, 1st ed. (1984).

[25] G. Donnarumma, J. Woźny and Z. Lisik, *Numerical solution of the anisotropic poisson equation for SiC semiconductors device simulation*, *Materials Science and Engineering: B* **176** (2011) 293.

[26] H. Haug and S.W. Koch, *Quantum Theory of the Optical and Electronic Properties of Semiconductors*, World Scientific, Singapore, 5th ed. (2009).

[27] J.T. Asubar, Z. Yatabe, D. Gregusova and T. Hashizume, *Controlling surface/interface states in GaN-based transistors: Surface model, insulated gate, and surface passivation*, *Journal of Applied Physics* **129** (2021) 121102.

[28] W. Mönch, *Semiconductor Surfaces and Interfaces*, Springer, Berlin, 2nd ed. (2001).

[29] W. Mönch, *Adsorbate-induced surface states and fermi-level pinning at semiconductor surfaces*, *Journal of Vacuum Science & Technology B: Microelectronics Processing and Phenomena* **7** (1989) 1216.

[30] E.H. Do, S.G. Kwon, M.H. Kang and H.W. Yeom, *Structural and electronic effects of adatoms on metallic atomic chains in Si(111)5×2-Au*, *Scientific Reports* **8** (2018) 15537.

[31] I.M. Dharmadasa, *Fermi level pinning and effects on CuInGaSe2-based thin-film solar cells*, *Semiconductor Science and Technology* **24** (2009) 055016.

[32] C. Hobbs, L. Fonseca, A. Knizhnik, V. Dhandapani, S. Samavedam, W. Taylor et al., *Fermi-level pinning at the polysilicon/metal oxide interface—part I*, *Electron Devices, IEEE Transactions on* **51** (2004) 971 .

[33] S. Li, *Semiconductor Physical Electronics*, Springer, New York, NY, 2nd ed. (2006).

[34] M.-H. Shih, *A further generalization of the Ostrowski theorem in banach spaces*, *Proceedings of The Japan Academy, Series A, Mathematical Sciences* **57** (1981) 168.

[35] S. Hanzely, D. Kamzolov, D. Pasechnyuk, A. Gasnikov, P. Richtárik and M. Takáč, *A damped newton method achieves global $\mathcal{O}\left(\frac{1}{k^2}\right)$ and local quadratic convergence rate*, in *Advances in Neural Information Processing Systems*, A.H. Oh, A. Agarwal, D. Belgrave and K. Cho, eds., 2022, https://openreview.net/forum?id=rjDziEPQLQs.

[36] K. Murota, *Global convergence of a modified newton iteration for algebraic equations*, *SIAM Journal on Numerical Analysis* **19** (1982) 793.

[37] C.G. Broyden, *The convergence of a class of double-rank minimization algorithms 1. general considerations*, *IMA Journal of Applied Mathematics* **6** (1970) 76.

[38] D. Young, *Iterative methods for solving partial difference equations of elliptic type*, *Transactions of the American Mathematical Society* **76** (1954) 92.

[39] C.G. Broyden, *On convergence criteria for the method of successive over-relaxation*, *Mathematics of Computation* **18** (1964) 136.

[40] K.R. James, *Convergence of matrix iterations subject to diagonal dominance*, *SIAM Journal on Numerical Analysis* **10** (1973) 478.

[41] G. Strang, *Linear Algebra and Its Applications*, Thomson Brooks/Cole, Belmont, CA, 3rd ed. (2006).

[42] R.B. Dingle, *The fermi-dirac integrals*, *Applied Scientific Research, Section B* **6** (1957) 225.

[43] V.D. Pham, Y. Pan, S.C. Erwin and S. Fölsch, *Quantum dots on the InAs(110) cleavage surface created by atom manipulation*, *Phys. Rev. Res.* **6** (2024) 013269.

[44] H. Guhl, W. Miller and K. Reuter, *Oxygen adatoms at $SrTiO_3$(001): A density-functional theory study*, *Surface Science* **604** (2010) 372.

[45] Z. Huang, G. Hao, C. He, H. Yang, L. Xue, X. Qi et al., *Density functional theory study of Fe adatoms adsorbed monolayer and bilayer MoS_2 sheets*, *Journal of Applied Physics* **114** (2013) .

[46] A.G. Baydin, B.A. Pearlmutter, A.A. Radul and J.M. Siskind, *Automatic differentiation in machine learning: a survey*, *Journal of Machine Learning Research* **18** (2018) 1.

[47] T. Fukushima, *Precise and fast computation of fermi–dirac integral of integer and half integer order by piecewise minimax rational approximation*, *Applied Mathematics and Computation* **259** (2015) 708.

[48] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Society for Industrial and Applied Mathematics, Philadelphia, PA (2000), 10.1137/1.9780898719468.

[49] M.H. Gutknecht, *A brief introduction to krylov space methods for solving linear systems*, in *Frontiers of Computational Science*, Y. Kaneda, H. Kawamura and M. Sasai, eds., (Berlin, Heidelberg), pp. 53–62, Springer Berlin Heidelberg, 2007.

[50] M. Brozel, *Gallium arsenide*, in *Springer Handbook of Electronic and Photonic Materials*, S. Kasap and P. Capper, eds., (Boston, MA), pp. 499–536, Springer US (2007).

[51] G.J. de Raad, D.M. Bruls, P.M. Koenraad and J.H. Wolter, *Interplay between tip-induced band bending and voltage-dependent surface corrugation on GaAs(110) surfaces*, *Phys. Rev. B* **66** (2002) 195306.

[52] G. Münnich, *Cross-sectional scanning probe microscopy on GaAs: Tip-induced band bending, buried acceptors and adsorbed molecules*, doctoral dissertation, üniversität Regensburg, June, 2014.

[53] R.M. Feenstra, "SEMITIP V6, Technical Manual." `https://www.andrew.cmu.edu/user/feenstra/semitip_v6/TechManual.html`, 2011.

[54] D. Knoll and D. Keyes, *Jacobian-free newton–krylov methods: a survey of approaches and applications*, *Journal of Computational Physics* **193** (2004) 357.