Numerical Modelling of Terahertz Response from a HEMT Structure

Masters Thesis
Physics
Aalborg University
10th of June 2019

AALBORG UNIVERSITY
STUDENT REPORT

Department of Materials and Production
Physics
Skjernvej 4a
9220 Aalborg

Title:
Numerical Modelling of Terahertz Response from a HEMT Structure

Project:
Master Thesis

Project period:
February 1st, 2019 - June 10th, 2019

Project group:
5.336c

Student:
Kristian Nedergaard Jakobsen
Niklas Linaa Larsen

Supervisor:
Thomas Søndergaard

Pages: 94
Appendix: 14
Submission date 10-06-2019

Abstract:

Dette specialeprojekt er udarbejdet under titlen *Numerical Modelling of Terahertz Response from a HEMT Structure*. Formålet med rapporten er at forstå dannelsen af plasma bølger i high electron mobility transistors (HEMTs) two-dimensional electron gas (2DEG) lag og være i stand til at numerisk at kunne bestemme resonans frekvenser i terahertz området, hvordan det påvirker excitation af disse bølger. Desuden vil sammenhænge mellem disse resonans frekvenser, længden af gaten i transistoren og afstand til 2DEGen undersøges. Til dette formål er der blevet opstillet en metode til at udregne strømmen i HEMT gaten, som er baseret på Maxwell's ligninger og tilhørende elektromagnetiske grænsebetingelser. Der blev som udgangspunkt brugt point matching for at beskrive strømmen, hvilket resulterede i inkonsistent konvergens. For at forbedre dette blev anden ordens basis funktioner anvendt, hvilket resulterede i en forbedret konvergens. Desuden blev strukturen analyseret for guidede modes, hvilket under antagelsen af at 2DEG'en er uendelig tynd gav resonans frekvenser som stemte overens med den numeriske model. Generelt var det muligt at bestemme resonans frekvenser i terahertz området for en given struktur, der overvejende var i overensstemmelse med hvad var forventet.

Preface

This master thesis is written by Kristian Nedergaard Jakobsen and Niklas Linaa Larsen 4th semester master's students of physics at the Department of Materials and Production, Aalborg University. The subject of the project is *Numerical Modelling of Terahertz Response from a HEMT Structure* and studies plasma waves in a HEMT structure and attempts to use numerical modelling to describe the phenomena.
The report is divided into five parts; a theoretical introduction which introduces the formulae and concepts which are used to describe and discuss the results; a section which describes the methods used in the modelling of the HEMT structure; an analysis in which specific structures are modelled and analysed; a discussion of the models where the theoretical knowledge is used to discuss the results; and the conclusion where the theoretical knowledge is utilised to conclude upon the models. In addition an appendix with long derivations, complementary theory, and additional figures is included.

Reading guide

In this report, references are a part of the text and are collected in the bibliography in the back of the report.
References are indicated with [*number*]. These numbers refer to the bibliography where books are stated with author, title, edition, and publisher while web-pages are stated with author, title and last date accessed.
In the text, vectors and matrices are shown as **A**.
Figures are numbered according to the chapter they appear in, i.e. the first figure in chapter 2 is numbered 2.1, the second figure has the number 2.2, etc. Figures have an explaining caption placed below.

Kristian Nedergaard Jakobsen   Niklas Linaa Larsen

Contents

1 Introduction **1**

2 Electromagnetic Theory **3**
2.1 Maxwell's Equations . . . 3
2.2 Optical Cross Sections . . . 3
2.2.1 Scattering Cross Section . . . 4
2.2.2 Extinction Cross Section . . . 5

3 Field-Effect Transistors **9**
3.1 Heterojunctions . . . 10
3.1.1 Energy-Band Diagrams . . . 10
3.1.2 Two-Dimensional Electron Gas . . . 11
3.1.3 AlGaN/GaN Field Effect Transistor . . . 12
3.2 Plasma Waves . . . 15
3.2.1 Density Response Theory . . . 17
3.3 Plasma Wave Instability . . . 22
3.3.1 Instability Origin . . . 26
3.3.2 Instability Conditions . . . 27

4 Method **29**
4.1 Guided Modes . . . 29
4.1.1 Gated Structure . . . 29
4.2 Integral Equation Method for 2DEG . . . 33
4.3 Reference Field . . . 37

5 Analysis and Discussion **43**
5.1 Mode Indices . . . 43
5.2 The PPS Method . . . 48
5.2.1 Calculating the current . . . 52
5.3 The RF Method . . . 58
5.4 Second Order Basis functions . . . 60
5.4.1 Current Analysis . . . 63
5.5 Radiation Pattern Analysis . . . 71
5.5.1 InGaAs HEMT . . . 71
5.5.2 AlGaN/GaN HEMT . . . 77

6 Final Words **87**
6.1 Discussion . . . 87
6.2 Conclusion . . . 89

Bibliography **91**

A Appendix

A.1 Energy Levels of 2DEG . . . 95
A.2 Linear Response Theory . . . 97
A.3 Fourier Transform on Coulomb Potential . . . 100
A.4 Fluid Dynamics . . . 101
A.5 Mode Index for an Ungated Structure . . . 103
A.6 Drude Model . . . 106
A.7 Fourier Transform of Second Order Basis Functions . . . 107

Introduction 1

Frequencies in the range of 0.1 THz to 3 THz, are often called the terahertz gap, as this range lies between the two well known domains of electronic and optical devices. Technology operating in this range is difficult to fabricate, as the technologies which work for higher or lower frequencies are not directly applicable to terahertz technology [1]. Many common materials, and even living tissue, is semi-transparent in the terahertz domain and because it is non-ionising, terahertz technology can be used for screening purposes. This can be utilised in more commercial applications such as monitoring of compounding processes or quality inspection of food products [2]. It has also gotten the attention in military uses or security for airports for being able to detect weapons and drugs [3]. The health sector has also gotten a keen interest as terahertz imaging has a potential for detection of cancer [4].

Everything with a temperature of over 10 K, emits terahertz radiation as a part of blackbody radiation, thus astronomers are also interested in terahertz technology to analyse the cosmic background radiation [1]. A limiting factor for technologies utilising terahertz radiation is that it is not capable of penetrating through water and therefore has limited range in the atmosphere. It is, however, reasonable for uses within ≈ 10 m, thus might have uses for short range technologies, i.e. high bandwith wifi systems [3]. There are several methods which are capable of producing terahertz radiation, however most methods are either expensive or inefficient. Terahertz technology is still in the developing stages and cheap and effective terahertz sources and receivers are sought [3]. A new method for creating terahertz radiation was postulated by Michael I. Dyakonov and Michael S. Shur in 1993 [5], where the theory behind using the propagation of plasma waves in a High Electron Mobility Transistor (HEMT) to create and measure terahertz radiation was established. Since then a lot of effort has went into the construction of HEMT structures which are capable of interacting with terahertz radiation. This could help revolutionise the terahertz technology, as HEMT structures are well known and thus the existing knowledge in the field can help create more efficient terahertz sources.

This project aims to understand the generation of terahertz radiation from the propagation of plasma waves in a HEMTs two-dimensional electron gas (2DEG). Furthermore, it seeks to develop a model for the response of a HEMT, to an incident electromagnetic wave, in the terahertz range, in order to determine resonance frequencies for a HEMT structure. To this end classical electromagnetism is used in combination with Fourier analysis relevant structures. The project has put special focus on the AlGaN/GaN HEMT structure, as this is a newer more prominent candidate for efficient terahertz sources [6][7].

Electromagnetic Theory 2

2.1 Maxwell's Equations

This section is based on references [8] and [9]

In order to establish the propagation of plasma waves, basic electromagnetic theory will be established. Propagation of electromagnetic waves are governed by Maxwell's equations. In a homogeneous, linear and isotropic material they are given as

$$\nabla \cdot \mathbf{D}(\mathbf{r}, t) = \rho, \tag{2.1}$$

$$\nabla \cdot \mathbf{B}(\mathbf{r}, t) = 0, \tag{2.2}$$

$$\nabla \times \mathbf{E}(\mathbf{r}, t) = -\frac{\partial \mathbf{B}(\mathbf{r}, t)}{\partial t}, \tag{2.3}$$

$$\nabla \times \mathbf{H}(\mathbf{r}, t) = \mathbf{J}_f(\mathbf{r}, t) + \frac{\partial \mathbf{D}(\mathbf{r}, t)}{\partial t}, \tag{2.4}$$

where ρ is the free charges, $\mathbf{E}$ is the electric field, $\mathbf{B}$ is the magnetic induction field, $\mathbf{J}_f$ is the free currents, $\mathbf{D} = \varepsilon\mathbf{E}$ is the displacement field and $\mathbf{H} = \frac{\mathbf{B}}{\mu}$ is the magnetic field. The electric permittivity,ε, of the material and is given as $\varepsilon = \varepsilon_r\varepsilon_0$, where ε_r is the relative permittivity and ε_0 is the vacuum permittivity, and μ is the magnetic permeability and is given as $\mu = \mu_r\mu_0$, where μ_r is the relative permeability and μ_0 is the vacuum permeability. The time dependence is assumed to be $e^{-i\omega t}$ unless otherwise specified.

Taking the curl of Eq. (2.3), taking the time derivative, and using the defined time dependence, gives

$$\nabla \times \nabla \times \mathbf{E} = i\omega\mu\nabla \times \mathbf{H} = i\omega\mu \left(\mathbf{J}_f - i\omega\varepsilon\mathbf{E}\right), \tag{2.5}$$

where the last equality comes from inserting Eq. (2.4). Eq. (2.5) is known as the vector wave equation. The vector identity $\nabla \times \nabla \times \mathbf{f} = \nabla(\nabla \cdot \mathbf{f}) - \nabla^2\mathbf{f}$ is used to obtain

$$\nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E} = i\omega\mu \left(\mathbf{J}_f - i\omega\varepsilon\mathbf{E}\right). \tag{2.6}$$

Inserting Eq. (2.1) and defining $\omega^2\varepsilon\mu = k^2$ gives

$$\left(\nabla^2 + k^2\right)\mathbf{E} = \frac{\nabla\rho}{\varepsilon} - i\omega\mu\mathbf{J}_f \tag{2.7}$$

For a situation where there are no free currents or charges, $\rho = \mathbf{J}_f = 0$, which is known as the Helmholtz equation.

2.2 Optical Cross Sections

This section is based on reference [10].

In this section expressions for far field approximations for the magnetic field from a scatterer will be derived, in order to calculate the radiation pattern emitted from a structure and the related cross sections.

2.2.1 Scattering Cross Section

For a scatterer on a layered structure, which will be considered here, the expression of the far field profile depends on which half-plane is considered. However since the derivation is similar for both half-planes, only the upper half-plane, corresponding to $\theta \in]0, \pi[$ will be derived thoroughly. Writing the field in the upper half-plane in its angular spectrum representation yields

$$H_{z,a}(x, y) = \int_{-\infty}^{\infty} \tilde{H}_a(k_x; y = 0) e^{ik_x x} e^{ik_{y,a} y} \mathrm{d}k_x, \tag{2.8}$$

where the subscript a denotes the upper half-plane (ambient) and $k_{y,a} = \sqrt{k_0^2 \varepsilon_a - k_x^2}$. For far field calculations, the evanescent waves can be ignored, thus the integral can be limited to values for which $\mathrm{Im}(k_{y,a}) = 0$. Furthermore, by writing the spacial coordinates in polar form:

$$x = r\cos(\theta) \quad y = r\sin(\theta), \tag{2.9}$$

and performing a similar change of variable for the wavevector components

$$k_x = k_0 n_a \cos(\theta_k) \quad k_y = k_0 n_a \sin(\theta_k), \tag{2.10}$$

with $\mathrm{d}k_x = -k_0 n_a \sin(\theta_k)\mathrm{d}\theta_k$, Eq. (2.8) becomes

$$H_{z,a}^{(ff)}(r, \theta) \simeq \int_0^{\pi} \tilde{H}_a(k_x) e^{ik_0 n_a r \cos(\theta - \theta_k)} k_0 n_a \sin(\theta_k) \mathrm{d}\theta_k, \tag{2.11}$$

where $\cos(\theta_k)\cos(\theta) + \sin(\theta_k)\sin(\theta) = \cos(\theta - \theta_k)$ has been used and the integration direction has been reversed. For large r the exponential function $e^{ik_0 n_a r \cos(\theta - \theta_k)}$ will vary very rapidly comparably to the rest of the terms of the integrand, except for when $\theta \simeq \theta_k$. Therefore for very large r, only values of $\theta_k \approx \theta$ will contribute to the integral. For these values the cosine term can be approximated by a Taylor expansion:

$$\cos(\theta - \theta_k) \simeq 1 - \frac{1}{2}(\theta - \theta_k)^2. \tag{2.12}$$

The integration limits can be extended to $-\infty$ and ∞ due to the rapid oscillation of the exponential function, as when values of θ_k are not near θ the extended limits will not contribute to the integral. Furthermore, the slowly varying terms can be placed outside the integral evaluated in $\theta_k = \theta$ as these will only contribute to the integral when $\theta \simeq \theta_k$. This leads to

$$H_{z,a}^{(ff)}(r, \theta) \simeq k_0 n_a e^{ik_0 n_a r} \tilde{H}_a(k_x(\theta)) \sin(\theta) \int_{-\infty}^{\infty} e^{-ik_0 n_a r \frac{1}{2}(\theta - \theta_k)^2} \mathrm{d}\theta_k. \tag{2.13}$$

The integral of Eq. (2.13) is a Gaussian integral with an imaginary argument. This has the known solution [10]

$$\int_{-\infty}^{\infty} e^{-az^2 + bz} \mathrm{d}z = e^{b^2/4a} \sqrt{\frac{\pi}{a}}, \tag{2.14}$$

which reduces Eq. (2.13) to

$$H^{(ff)}_{z,a}(r,\theta) \simeq k_0 n_a e^{ik_0 n_a r} \tilde{H}_a(k_x(\theta)) \sin(\theta) \sqrt{\frac{2\pi}{k_0 n_a r}} e^{-i\frac{\pi}{4}}, \quad 0 < \theta < \pi, \tag{2.15}$$

where it is used that $1/\sqrt{i} = e^{-i\pi/4}$. The far field expression governing the field in the lower half-plane (denoted by s) can be obtained from a similar derivation by changing the refractive index to that of the substrate, n_s, and remembering that the field is now examined in the interval $\theta \in]-\pi, 0[$, which leads to

$$H^{(ff)}_{z,s}(r,\theta) \simeq k_0 n_s e^{-ik_0 n_s r} \tilde{H}_s(k_x(\theta)) \sin(\theta) \sqrt{\frac{2\pi}{k_0 n_s r}} e^{i\frac{\pi}{4}}, \quad -\pi < \theta < 0. \tag{2.16}$$

In order to find the scattering cross sections, the scattered power is divided by the incident beam, given as

$$I_i = \frac{1}{2n_a}\sqrt{\frac{\mu_0}{\varepsilon_0}}|H_0|^2. \tag{2.17}$$

The scattered power is found by integrating over the time-averaged Poynting vector flux, giving

$$P_{sc,up} = \frac{1}{2}\sqrt{\frac{\mu_0}{\varepsilon_0}}\frac{1}{n_a}\int_0^{\pi} |H^{(ff)}_{sc}(r,\theta)|^2 r\mathrm{d}\theta. \tag{2.18}$$

The out-of-plane scattering cross sections is then given by normalising Eq. (2.18) with the incident power, Eq. (2.28), and likewise for transmitted, by substituting n_a with n_s, yielding

$$\sigma_{sc,up} = \frac{1}{|H_0|^2}\int_0^{\pi} |H^{(ff)}_{z,r}(r,\theta)|^2 r\mathrm{d}\theta, \tag{2.19}$$

$$\sigma_{sc,down} = \frac{1}{|H_0|^2}\frac{n_a}{n_s}\int_{-\pi}^{0} |H^{(ff)}_{z,t}(r,\theta)|^2 r\mathrm{d}\theta. \tag{2.20}$$

In order to plot the radiation pattern, the scattered power per unit angle is needed, this is termed the differential cross section and is given by

$$\frac{\partial \sigma_{sc}}{\partial \theta} = \begin{cases} \frac{1}{|H_0|^2}|H^{(ff)}_{z,r}(r,\theta)|^2 r & 0 < \theta < \pi, \\ \frac{1}{|H_0|^2}|H^{(ff)}_{z,t}(r,\theta)|^2 \frac{n_a}{n_s} r & -\pi < \theta < 0. \end{cases} \tag{2.21}$$

2.2.2 Extinction Cross Section

The extinction power represents the power lost due to scattering by the scatterer and absorption by the scatterer. For a layered structure an extinction cross section for both the reflected field due to scatterer and the transmitted field can be found. In order to derive these expressions the beam power is examined in the far field over a small angle $2\theta_b$, which is done in order to not include too much of the scattered intensity. In the upper half-plane, the upward propagating part of the field will consist of the reflected field from the structure and the scattered field from the scatterer:

$$H_{tot,r}(r,\theta) = H_{0,r}(r,\theta) + H_{sc}(r,\theta), \tag{2.22}$$

where the amplitude of $H_{0,r}$ is related to amplitude of the incident field by the Fresnel reflection coefficient for the layered structure. In order to find this reflected field, a Gaussian wave is used to describe the incident field, and it is given as

$$H_{0,i}(x, y = 0) = H_0 e^{-x^2/w_0^2}, \tag{2.23}$$

where w_0 is the beam waist radius. In order to propagate this beam, the Fourier transform and a propagator is used to get

$$\tilde{H}_{0,i}(k_x; y) = \frac{1}{2\pi}\int H_{0,i}(x, 0)e^{ik_y y}e^{ik_x x}\mathrm{d}x. \tag{2.24}$$

The angular spectrum, $\tilde{H}_0(k_x; y = 0)$, can be found by

$$\tilde{H}_{0,i}(k_x; y = 0) = \frac{1}{2\pi}\int H_0(x, 0)e^{-x^2/w_0^2}e^{-ik_x x}\mathrm{d}x = \frac{H_0 w_0}{2\sqrt{\pi}}e^{-k_x^2 w_0^2/4}, \tag{2.25}$$

where the last equation is found using Eq. (2.14). Now to find the reflected part of the incoming wave, the Fresnel reflection coefficient, $r(k_x)$, is used to get

$$H_{0,r}(x, y) = \int r(k_x)\tilde{H}_{0,i}(k_x; 0)e^{ik_x x}e^{ik_y y}\mathrm{d}k_x. \tag{2.26}$$

If the incident wave is a plane wave, this can be achieved by assuming a very large beam waist, and letting $w_0 \to \infty$. The far field is sought, thus the reflected wave is transformed into polar coordinates in the far field, where it is used that $k_x = k_0 n_a \sin(\theta_k)$, $k_y = k_0 n_a \cos(\theta_k)$, $x = r\sin(\theta)$ and $y = r\cos(\theta)$. These angles are defined with respect to the normal incidence, in order to simplify the calculations. This allows for the field to be written as

$$H_{0,r}^{(ff)}(r, \theta) = \frac{H_0 w_0}{2\sqrt{\pi}}\int r(k_x)e^{-\sin^2(\theta_k)\frac{(w_0 k_0 n_a)^2}{4}}e^{ik_0 n_a r\cos(\theta_k-\theta)}k_0 n_a\cos(\theta_k)\mathrm{d}\theta_k, \tag{2.27}$$

where $\cos(\theta_k)\cos(\theta) + \sin(\theta_k)\sin(\theta) = \cos(\theta - \theta_k)$ has been used. As stated earlier, this type of integral only gives something when θ is close to θ_k. Thus the slowly varying terms can be moved out, making the Fresnel coefficient $r(k_x = 0)$ and using a Taylor expansion to get $\sin^2(\theta) \approx \theta^2$. The faster varying term can be handled with a second order Taylor expansion $\cos(\theta_k - \theta) \approx 1 - 1/2(\theta_k - \theta)^2$, giving

$$H_{0,r}^{(ff)}(r, \theta) \approx H_0 w_0\sqrt{\frac{k_0 n_a}{2r}}r(k_x = 0)e^{-\frac{\theta^2}{4}(k_0 w_0 n_a)^2}e^{-i\pi/4}e^{ik_0 n_a r}, \tag{2.28}$$

where again Eq. (2.14) has been used. The reflected beam power is obtained by integrating over the time-averaged Poynting vector flux over the small angular interval

$$\begin{aligned} P_{beam,r} &= \int_{\theta=-\theta_b}^{\theta_b}\frac{1}{2}\mathrm{Re}\left(\mathbf{E}^{(ff)}(r, \theta) \times [\mathbf{H}^{(ff)}(r, \theta)]^*\right)\cdot\hat{\mathbf{r}}r\mathrm{d}\theta \\ &= \frac{1}{2}\sqrt{\frac{\mu_0}{\varepsilon_0}}\frac{1}{n_a}\int_{\theta=-\theta_b}^{\theta_b}|H_{0,r}^{(ff)}(r, \theta) + H_{sc}^{(ff)}(r, \theta)|^2 r\mathrm{d}\theta. \end{aligned} \tag{2.29}$$

In the limit for $\theta_b \to 0$ the total power can be divided into two parts

$$P_{beam,r} = P_{0,r} - P_{ext,r}, \tag{2.30}$$

where the reflected part, $P_{0,r}$, from the structure is given by

$$P_{0,r} = \frac{1}{2}\sqrt{\frac{\mu_0}{\varepsilon_0}}\frac{1}{n_a}\int_{\theta=-\theta_b}^{\theta_b} |H_{0,r}^{ff}(r,\theta)|^2 r\mathrm{d}\theta, \tag{2.31}$$

and the negative reflected extinction power, $-P_{ext,r}$ is given by

$$-P_{ext,r} = \sqrt{\frac{\mu_0}{\varepsilon_0}}\frac{1}{n_a}\int_{\theta=-\theta_b}^{\theta_b} \mathrm{Re}\left(H_{0,r}^{(ff)}(r,\theta)\left[H_{sc,r}^{(ff)}(r,\theta)\right]^*\right) r\mathrm{d}\theta. \tag{2.32}$$

The squared term of $H_{sc,r}^{(ff)}$ has been neglected as the incident field is several magnitudes stronger than the scattered field. Inserting Eq. (2.28) gives

$$-P_{ext,r} = \sqrt{\frac{\mu_0}{\varepsilon_0}}H_0 w_0\sqrt{\frac{k_0 r}{2n_a}}r(k_x=0)\int_{\theta=-\theta_b}^{\theta_b} \mathrm{Re}\left(e^{-\frac{\theta^2}{4}(k_0 w_0 n_a)^2}e^{-i\pi/4}e^{ik_0 n_a r}\left[H_{sc,r}^{(ff)}(r,\theta)\right]^*\right)\mathrm{d}\theta. \tag{2.33}$$

Again using Eq. (2.14) and using that $H_{sc}^{(ff)}(r,\theta) \approx H_{sc}^{(ff)}(r,\theta=0)$ gives

$$-P_{ext,r} = \sqrt{\frac{\mu_0}{\varepsilon_0}}\frac{H_0}{n_a}\sqrt{\frac{2\pi r}{k_0 n_a}}r(k_x=0)\mathrm{Re}\left(e^{-i\pi/4}e^{ik_0 n_a r}\left[H_{sc,r}^{(ff)}(r,\theta=0)\right]^*\right). \tag{2.34}$$

The extinction cross section is given as the fraction of the total scattered power against the power of the incident beam I_i,

$$\sigma_{ext,r} = \frac{P_{ext,r}}{I_i}, \tag{2.35}$$

where I_i is given in Eq. (2.17). A similar derivation is valid for the extinction cross section for the transmitted beam, where the refractive index has to be that of the final layer and the reflection coefficients should be changed to the transmission coefficients for the layered structure.

Field-Effect Transistors 3

This chapter is based on references [11] and [12].

In this section the concept of Field-Effect Transistors (FETs) will be presented. The idea was first proposed by Lilienfeld in 1926 when he filed for a patent for a three electrode structure in *Method and Apparatus for Controlling Electric Currents* later to be known as a field effect transistor [13]. Today a large number of variations exists but the principle, illustrated in Fig. 3.1, has remained the same. The FET is a three terminal structure, which functions as a capacitor with one plate working as a conducting channel between two of the terminals, denoted the source and drain terminal. The other plate, i.e. the gate terminal, regulates the charge carriers induces into the channel, which comes from the source and moves toward the drain.

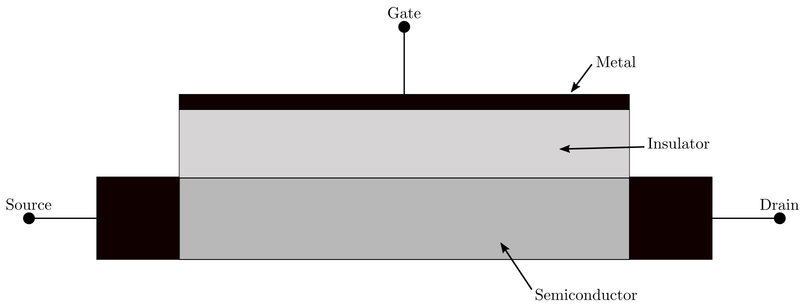

Figure 3.1. Schematic structure of the concept of a field-effect transistor.

The FET principle has been implemented in a plethora of different devices, which are mainly categorised by their gate material, location of the gate with respect to the channel, how the gate is isolated from the channel and which type of carriers is moved through the channel. This gives rise to the distinction of n-channel devices, p-channel devices, and Double Injection Field-Effect Transistors, which uses electrons, holes or both as carriers respectively. Some of the most common devices are the Metal Oxide Semiconductor Field-Effect Transistor (MOSFET), the MEtal Semiconductor Field-Effect Transistors (MESFET) and the Junction Field-Effect Transistors (JFET) [11]. The MOSFET has especially been used extensively in electronic circuit applications due to its small size, which allows for hundreds of millions transistors to be placed on a single chip [14]. As transistors continue to become smaller quantum mechanical effects starts to have a significant effect, which introduces new problems, thus limiting the scaling of conventional MOSFETs [15][16].

A relatively new FET structure is the Heterostructure Field-Effect Transistors (HFET) [17], also commonly referred to as High Electron Mobility Transistor (HEMT), has been emerging as a subject of great interest. A HEMT consists of two semiconductor materials, which forms an abrupt discontinuity in conduction and valence bands and will be the main subject of this chapter. Firstly the theory of heterojunctions will be presented with focus on their banddiagrams and the formation of a Two-Dimensional Electron Gas (2DEG) adjacent to the interface. Secondly the HEMT will be examined with special focus on the AlGaN/GaN heterostructure.

3.1 Heterojunctions

A heterojunction consist of an interface of two semiconductors with different bandgaps. This difference results in a discontinuity of the bandgaps at the junction interface. In general there are two methods of designing a heterojunction; either an abrupt junction in which the semiconductor changes abruptly from one to the other, or a graded heterojunction as for example GaN/$Al_xGa_{1-x}N$, where x may vary continuously. The graded heterojunction essentially enables one to design the bandgap energies at the interface.

3.1.1 Energy-Band Diagrams

Many of the characteristics of the heterojunction can be determined from its bandgap. Especially the formation of the 2DEG at the interface can be understood from the bending and discontinuity of the bandgaps at the interface. In the formation of a heterojunction between a narrow-bandgap material and a wide-bandgap material the alignment of their bandgap energies plays an important role for the resulting structure. In this report the type of heterojunctions called a straddling heterojunction is considered. This is the most common type of heterojunction and is the case where the wide-bandgap material completely overlaps the narrow-bandgap material as seen in Fig. 3.2. The heterojunction can further be seperated into four types based on how the two semiconductors are doped: Anisotypes in which the dopant type changes at the junction, nP or Np where the large letter indicates the wide-bandgap material. Isoptyes which have the same dopant type on both side forming either nN or pP heterojunctions.

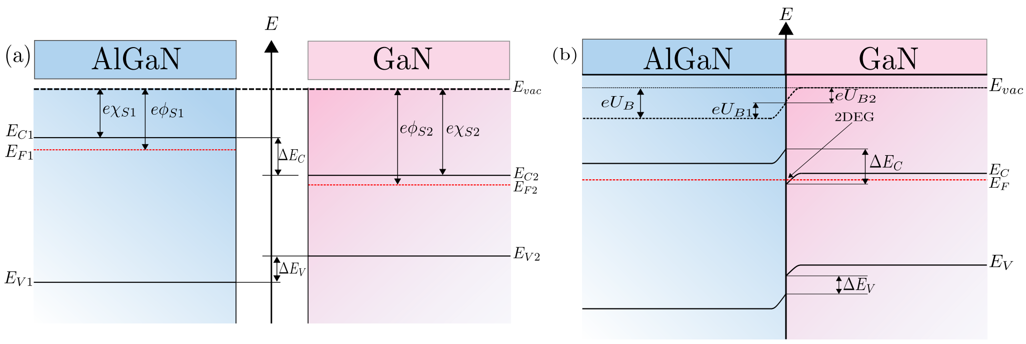

Figure 3.2. Banddiagrams of straddling heterojunctions, where (a) separated AlGaN and GaN and (b) AlGaN/GaN heterojunctions. E_V is the valence bands maximum, E_C is the conduction bands minimum, E_F is the Fermi level, ϕ_S is the work functions, χ_S is the electron affinities and U_B is the built-in voltage. A 2DEG gas is confined on in the triangular quantum well on the GaN side.

Only the bandgap of an isotype junction will be considered here, as only they can form the 2DEG at the interface [12]. In particular the AlGaN/GaN heterojunction will be examined, but this exposition should be valid for most isotype junctions. In Fig. 3.2(a) the band diagrams of isolated AlGaN and GaN, being the wide-bandgap and narrow-bandgap semiconductor respectively, are shown with the vacuum level used as reference. Here E_{C1}, E_{V1}, E_{F1}, ϕ_{S1}, χ_{S1} and U_{B1} are the conduction band minimum, the valence band maximum, the Fermi level, the semiconductor work function, the electron affinity and the built-in voltage for AlGaN and E_{C2}, E_{V2}, E_{F2}, ϕ_{S2}, χ_{S2} and U_{B2} are the respective parameters of GaN. The energy discontinuity between the conduction bands and valence bands are denoted $\Delta E_C = E_{C1} - E_{C2}$ and $\Delta E_V = E_{V1} - E_{V2}$ and from Fig. 3.2(a) it can be seen that

$$\Delta E_C = e(\chi_{S1} - \chi_{S2}) \tag{3.1}$$

and

$$\Delta E_V = \Delta E_g - \Delta E_C, \tag{3.2}$$

where $\Delta E_g = E_{g1} - E_{g2}$ is the energy gap discontinuity.

When joining two semiconductors to form a heterojunction, the atoms at the interface need to form chemical bonds. As the lattice constants of the two materials differ, atoms at the interface have to adjust, by developing strain. Depending on the degree of mismatch the interface can be far from ideal and noticeable unwanted consequences can occur. However, in order to establish the theoretical foundation, an ideal heterojunction will be assumed.

The ideal model was first developed by R. L. Anderson [18], which states that the energy bands in both materials forming the heterostructure are not affected by the combination of the two materials. This reduces the problem to alignment of the band edges at the interface. The assumption for an ideal heterojunction is that the vacuum level is continuous and thus the discontinuities in the conduction band and valence band, Eqs. (3.1) and (3.2), will exist at the interface, this is also known as the electron affinity rule.

3.1.2 Two-Dimensional Electron Gas

This subsection is based on reference [1].

When forming a heterojunction the Fermi level must be constant throughout. This is due to a varying Fermi level will induce an electric current. Therefore, under equilibrium conditions where no such current flows, the Fermi level must be constant throughout the system. This requirement leads to band bending, illustrated in Fig. 3.2(b). For the system to reach equilibrium and thus aligning the Fermi level, electrons from the wide-bandgap AlGaN region flow across the junction. This flow causes E_{F1} to lower and E_{F2} to increase, due to several factors, e.g. repulsion effects, until the two fermi levels align. This creates a depletion region in AlGaN and an accumulation region in GaN. This leads to the electrons being confined in a potential well adjacent to the interface, wherein the Fermi level is higher than the conduction band. As is known from quantum mechanics electrons in a potential well will have their energy quantised. In other words, the electrons are accumulated in the discrete quantum states of the quantum well and the 2DEG is formed at the interface on the GaN side.

The term two-dimensional electron gas is used to emphasise that the electron energy is quantised perpendicular to the interface, but are free to move in the plane parallel to the interface. A quantum mechanical description of the energy of the 2DEG based on a one-dimensional triangular well is derived in App A.1.

The main advantage of the 2DEG is its high electron mobility compared to the bulk material and large electron concentration, as a current parallel to the interface will be a function of these parameters. In bulk material, electrons are often supplied by ionised donors and will as such suffer from scattering of impurities. While, in the accumulation region a large electron concentration can be created in an lightly doped or intrinsic region, which in turn will diminish impurity scattering and the low-field mobility can therefore be much higher. This is especially the case for low temperatures, where the ionised impurity is the dominant form of scattering. In general for heterojunctions the wide-bandgap material would have to be moderately to heavily doped, while the narrow-bandgap material could be lightly doped or in some cases intrinsic. However, due to spontaneous and piezoelectric polarisations the 2DEG can be induced in the AlGaN/GaN heterojunction without intentional doping [1][19]. This among other prominent properties, e.g. high breakdown voltage [20] making them candidates for high-power applications, makes the AlGaN/GaN heterojunction excellent candidates for HEMTs [21].

3.1.3 AlGaN/GaN Field Effect Transistor

In this subsection the workings of a common HEMT will be explored, this includes a short overview of the practical operation of a typical AlGaN/GaN HEMT, derivation of a current-voltage characteristics and relating the gate voltage to the charge density of the 2DEG.

Fig. 3.3 illustrates a typical AlGaN/GaN HEMT structure. All HEMTs work by the same principle. By applying a positive voltage to the gate (here a Schottky contact) electrons are capacitively induced into the narrow channel at the GaN interface, i.e. the 2DEG, allowing for a current to flow between the source and drain.

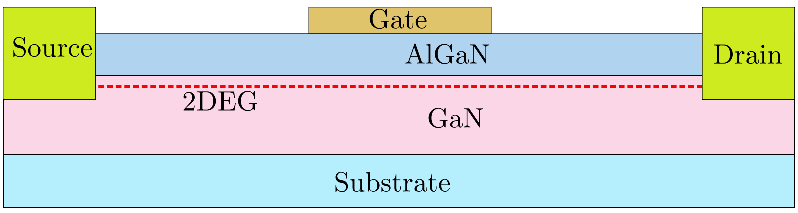

Figure 3.3. Schematic structure for a AlGaN/GaN HEMT on a substrate.

A Schottky contact is a switch utilising the Schottky barrier formed between a metal and semiconductor. The Schottky barrier is the potential barrier the electrons have to pass if they are to move from the metal into the semiconductor. In an ideal junction it is given by

$$\phi_B = \phi_m - \chi, \tag{3.3}$$

where ϕ_m is the work function of the metal and χ is the electron affinity of the adjacent semiconductor. The electrons in the conduction band of the semiconductor have a similar barrier called the built-in potential, if they are to move into the metal. It is given by

$$U_B = \phi_B - \phi_n, \tag{3.4}$$

where ϕ_n is the difference between the Fermi level and the conduction band.

Applying a negative gate bias increases the semiconductor-to-metal barrier, where for the ideal case ϕ_B is constant. This raises the triangular potential well at the GaN interface above the Fermi level and the accumulation layer is emptied. This lowers the carrier density of the 2DEG significantly, making the current in the HEMT essentially zero, switching it off, shown in Fig. 3.4(b). When zero or positive gate bias is applied the conduction band is below the Fermi level and the situation is as described in previous section, with a large electron density, shown in Fig. 3.4(a).

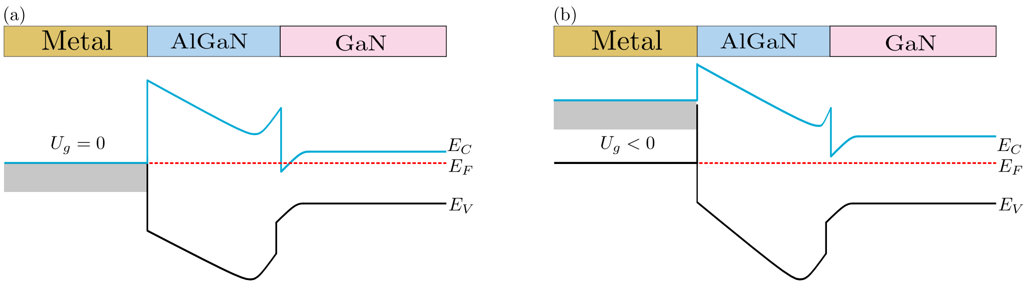

Figure 3.4. Energy band diagram of AlGaN/GaN HEMT in the direction from the gate to the substrate. (a) with zero gate bias. (b) negative gate bias.

In the remainder of this section the current-voltage characteristics of the HEMT will be developed based on a charge control model and assuming the Gradual Channel Approximation (GCA). The GCA is based on the assumption that the electric charge density which is related to the change in the electric field parallel to the conduction channel is much smaller than the change in electric field perpendicular to it, i.e. $\frac{\partial E_x}{\partial x} \ll \frac{\partial E_y}{\partial y}$. This means that the channel potential is assumed to change "gradually" and very little over distances of the order of the insulator (here the AlGaN layer) thickness along the channel length.

In the charge control model the channel carrier concentration is assumed to be given as

$$n_s(x) = \frac{\varepsilon_1}{e(d + \Delta d)}[U_g - U_{th} - U(x)], \tag{3.5}$$

where d is the thickness of the AlGaN layer, Δd is the thickness of the 2DEG, ε_1 is the dielectric permittivity of the AlGaN layer, U_g is the gate voltage, U_{th} is the threshold voltage and $U(x)$ is the potential along the channel due to the drain-to-source voltage, which varies from 0 to U_{ds} at $x = 0$ and $x = L$ respectively. The source-drain current is

$$I_{ds} = -en_s vW, \tag{3.6}$$

where v is the carrier drift velocity and W is the channel width. This expression is obtained by considering the change in the resistance at a point x in the channel

$$\mathrm{d}R = \frac{\rho \mathrm{d}x}{A(x)}, \tag{3.7}$$

where ρ is the resistivity and $A(x)$ is the cross section at x. The cross section can change along the channel by a varying thickness. For simplicity it is assumed to be constant $A(x) = A = \Delta dW$. The resistivity due to the 2DEG is

$$\rho = \frac{1}{e\mu_n n_s}, \tag{3.8}$$

where μ_n is the electron mobility, and n_s is in charge per volume. Inserting Eq. (3.8) into Eq. (3.7) yields

$$\mathrm{d}R = \frac{\mathrm{d}x}{\Delta dWe\mu_n n_s}. \tag{3.9}$$

The change in voltage across a length of dx can be written

$$\mathrm{d}U(x) = I_{ds}\mathrm{d}R(x), \tag{3.10}$$

where the source-drain current I_{ds} is constant through the channel. Isolating the source-drain current and using the expression for the resistance gives

$$I_{ds} = \Delta dWe\mu_n n_s \frac{\mathrm{d}U(x)}{\mathrm{d}x}. \tag{3.11}$$

The potential is related to the electric field as $E(x) = -\frac{\mathrm{d}U}{\mathrm{d}x}$. Furthermore the drift velocity can be obtained as $v = \mu E(x)$ and by including Δd in n_s making it charge per surface area, one obtains the source-drain current in Eq. (3.6). Using Eq. (3.5), the source-drain current can be expressed as

$$I_{ds} = -\frac{W\varepsilon_1 n_s}{(d+\Delta d)}[U_g - U_{th} - U(x)]v. \tag{3.12}$$

Applying $E(x) = -\frac{\mathrm{d}U}{\mathrm{d}x}$, the drain-source current becomes

$$I_{ds} = \frac{W\varepsilon_1 e n_s \mu}{(d+\Delta d)}[U_g - U_{th} - U(x)]\frac{\mathrm{d}U}{\mathrm{d}x}. \tag{3.13}$$

Multiplying both sides by dx and $1/L$ and integrating along the channel yields

$$\frac{1}{L}\int_{x=0}^{L} I_{ds}\mathrm{d}x = \frac{\mu W\varepsilon_1}{L(d+\Delta d)}\int_{U=0}^{U_{ds}}[U_g - U_{th} - U(x)]\mathrm{d}U. \tag{3.14}$$

Assuming that the current and mobility are constant through the channel, the source-drain current can be written as

$$I_{ds} = \frac{\mu W\varepsilon_1}{2L(d+\Delta d)}[2(U_g - U_{th})U_{ds} - U_{ds}^2]. \tag{3.15}$$

It is clear that the source-drain current can be controlled by both the source-drain voltage and the gate voltage. Furthermore, the gate voltage can be used to regulate the electron density in the 2DEG. This can be seen from Eq. (3.5) and by letting $U(x) = 0$ the system of the gate and the 2DEG can be seen as a capacitor with capacitance per area

$$C_g = \frac{\varepsilon_1}{d+\Delta d}. \tag{3.16}$$

The electron density can thus be regulated by the induced gate voltage as a capacitor by

$$C_g = \frac{e n_s}{U_g - U_{th}}. \tag{3.17}$$

3.2 Plasma Waves

This section is based on references [22] and [23].

Plasma waves are defined as oscillations in electron density. These oscillations have wavelike characteristics and can travel faster than the velocity of electrons. In order to describe the propagation of these waves a wave equation is needed. This is obtained by considering Newton's second law:

$$\mathbf{F} = m\mathbf{a}, \tag{3.18}$$

where $\mathbf{F}$ is the force, m is the mass of the particle and $\mathbf{a}$ is the acceleration. In order to simplify these waves, collisions are neglected and only the drift current, $\mathbf{v}$, is considered. In the case of an electron in an electric field, the force comes from the electric field, $\mathbf{E}$, and Newton's second law thus becomes

$$m\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\mathbf{E}e \Leftrightarrow -enm\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = e^2 n\mathbf{E}, \tag{3.19}$$

where e is the electron charge and n is the electron density. Considering an electron in a semiconductor, the mass, m, becomes the effective mass of the electron. The current density, given as $\mathbf{J} = -en\mathbf{v}$, can be inserted, yielding

$$\frac{\mathrm{d}\mathbf{J}}{\mathrm{d}t} = \frac{e^2 n}{m}\mathbf{E}. \tag{3.20}$$

Through analysis of charge conservation, the relation between the current density and the charge density, ρ, can be found. The current density through a closed surface must correspond to the change of the charge density in the volume the surface encloses, thus

$$-\frac{\partial}{\partial t}\int \rho \mathrm{d}V = \oint \mathbf{J}\cdot\hat{\mathbf{n}}\mathrm{d}A = \int \nabla\cdot\mathbf{J}\mathrm{d}V, \tag{3.21}$$

where the divergence theorem has been applied to get the last equality. As the integrals are both over the same arbitrary volume it must hold that

$$\nabla\cdot\mathbf{J} = -\frac{\partial\rho}{\partial t}. \tag{3.22}$$

Differentiating Eq. (3.22) with respect to time and inserting Eq. (3.20) yields

$$\frac{\partial^2\rho}{\partial t^2} + \frac{e^2 n}{m}\nabla\cdot\mathbf{E} = 0. \tag{3.23}$$

In order to arrive at the dispersion relation, the divergence of the electric field is needed.

3D Ungated Case

The first case studied is for a three dimensional ungated structure. Here n is the electron concentration per unit volume, and ρ is the charge per unit volume. For a 3D case Gauss's law, Eq. (2.1), given as $\nabla\cdot\mathbf{E} = \frac{\rho}{\varepsilon}$, can be inserted into Eq. (3.23), yielding

$$\frac{\partial^2\rho}{\partial t^2} + \frac{e^2 n}{m}\frac{\rho}{\varepsilon} = 0, \tag{3.24}$$

which is the equation for a harmonic oscillator, which has the solutions

$$\rho = Ae^{i\omega t} + Be^{-i\omega t}, \tag{3.25}$$

where A and B are amplitude constants and ω is the angular frequency. Inserting the solution into Eq. (3.24) gives the frequency

$$\omega = \sqrt{\frac{e^2 n}{m\varepsilon}}, \tag{3.26}$$

as such for the 3D ungated case the frequency is a constant with respect to the wavelength.

2D Gated Case

In the 2D case, $\mathbf{J}$ is the current per unit length, n is electron concentration per unit area and ρ is the charge per unit area. To correlate the gate to the 2DEG, a loop is created over the gate as shown in Fig. 3.5.

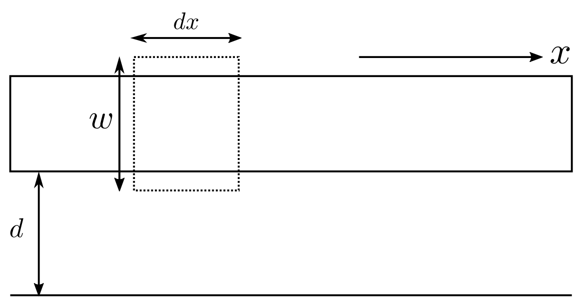

Figure 3.5. Schematic of the gate potential affecting the charge density in the 2DEG.

Through Gauss's law, Eq. (2.1), which in integral form is given as

$$\oint \mathbf{D} \cdot \hat{\mathbf{n}} \mathrm{d}A = \int \rho \mathrm{d}V, \tag{3.27}$$

the situation in Fig. 3.5 gives

$$\varepsilon E_x dxw = \rho w dx \Leftrightarrow E_x = \frac{\rho}{\varepsilon}, \tag{3.28}$$

where only the free charges are considered in ρ. Combining Eq. (3.28) with the relationship of the electric field and potential in a capacitor, $E = \frac{U}{d}$, gives

$$U = \rho \frac{d}{\varepsilon} = \frac{\rho}{C}, \tag{3.29}$$

where $C = \frac{d}{\varepsilon}$ is the capacitance per unit area, which holds when the change in the electric field parallel to the interface is much smaller than the change in the direction perpendicular to the interface. The electric field depends on the potential as $\mathbf{E} = -\nabla U$, thus combining this with Eq. (3.29) gives

$$\mathbf{E} = -\nabla U = -\frac{1}{C}\nabla \rho. \tag{3.30}$$

Inserting Eq. (3.30) into Eq. (3.23) gives

$$\frac{\partial^2 \rho}{\partial t^2} - \frac{e^2 n}{mC}\nabla^2 \rho = 0, \tag{3.31}$$

where ∇^2 is the two dimensional Laplace operator. The solution to this differential equation is on the form

$$\rho = Ae^{-i\omega t}e^{i\mathbf{k}\cdot\mathbf{r}} + Be^{-i\omega t}e^{-i\mathbf{k}\cdot\mathbf{r}}. \tag{3.32}$$

Inserting the solution into Eq. (3.31) and defining $s^2 = \frac{e^2 n}{mC}$ gives the dispersion relation

$$\omega^2 = k^2 s^2, \tag{3.33}$$

where k is the wave vector, and s is the speed of the plasma wave.

3.2.1 Density Response Theory

This subsection is based on references [24], [25] and [26].

Another approach to obtain the plasma wave dispersion relation, for the general case, the problem can be established as a many-body problem involving the electrons in the 2DEG. The plasma waves will depend on a non-local response in the dielectric function. Looking at a non-local dielectric function, $\varepsilon(\mathbf{r}-\mathbf{r}',\omega)$, the displacement field is given by

$$\mathbf{D}(\mathbf{r},\omega) = \int \varepsilon(\mathbf{r}-\mathbf{r}',\omega)\mathbf{E}(\mathbf{r}',\omega)\mathrm{d}^3 r'. \tag{3.34}$$

Taking the Fourier transform of the displacement field gives

$$\mathbf{D}(\mathbf{q},\omega) = \frac{1}{(2\pi)^3}\int \mathbf{D}(\mathbf{r},\omega)e^{-i\mathbf{q}\cdot\mathbf{r}}\mathrm{d}^3 r = \varepsilon(\mathbf{q},\omega)\mathbf{E}(\mathbf{q},\omega), \tag{3.35}$$

where $\mathbf{q}$ denotes the wavevector. Fourier transforming the vector wave equation, Eq. (2.5) yields

$$\mathbf{q}\times\mathbf{q}\times\mathbf{E}(\mathbf{q}) + k_0^2\varepsilon(\mathbf{q})\mathbf{E}(\mathbf{q}) = 0. \tag{3.36}$$

This allows for non-trivial solutions where $\mathbf{E}$ is in the same directions as $\mathbf{q}$, if $\varepsilon(\mathbf{q}) = 0$.

In order to find ε for the 2DEG, a quantum mechanical analysis of perturbations due to a external field will be made. Assuming an external field, $V_{ext}(\mathbf{r})$, the quantum mechanical Hamiltonian will be

$$\hat{H} = -e\sum_i^N V_{ext}(\mathbf{r}_i), \tag{3.37}$$

where N is the number of electrons. An electron density, $n_{ind}(\mathbf{r})$ is induced by the perturbation. The operator for this perturbation is $\sum_i^N \delta(\mathbf{r}-\mathbf{r}_i)$, which gives

$$n_{ind}(\mathbf{r}) = -\sum_{m,n} f_{nm}\frac{\langle n|\sum_i \delta(\mathbf{r}-\mathbf{r}_i)|m\rangle\langle m|-e\sum_i V_{ext}(\mathbf{r}_i)|n\rangle}{E_m - E_n - \hbar\omega}, \tag{3.38}$$

which is derived in App. A.2 and where $f_{nm} = f(E_n) - f(E_m)$, f being a probability distribution, and $\langle n|W|m\rangle = \int \phi_n^*(\mathbf{r}_1,\dots\mathbf{r}_N)W\phi_m(\mathbf{r}_1,\dots\mathbf{r}_N)\mathrm{d}^3 r_1\dots\mathrm{d}^3 r_N$, where ϕ_n is the wave function for state n and W is an operator. The states used here are true many-body states, meaning they depend on the entire system of electrons. For a symmetric function, $W = \sum_i w(\mathbf{r}_i)$ the integrals can be simplified by

$$\begin{aligned}
\langle n|W|m\rangle &= \int \phi_n^*(\mathbf{r}_1,\ldots,\mathbf{r}_N)\sum_i w(\mathbf{r}_i)\phi_m(\mathbf{r}_1,\ldots,\mathbf{r}_N)\mathrm{d}^3r_1\ldots\mathrm{d}^3r_N \\
&= N\int \phi_n^*(\mathbf{r}_1,\ldots,\mathbf{r}_N)w(\mathbf{r}_1)\phi_m(\mathbf{r}_1,\ldots,\mathbf{r}_N)\mathrm{d}^3r_1\ldots\mathrm{d}^3r_N \\
&= \int \rho_{nm}(\mathbf{r})w(\mathbf{r})\mathrm{d}^3r, \qquad (3.39)
\end{aligned}$$

where

$$\rho_{nm}(\mathbf{r}) = N\int \phi_n^*(\mathbf{r},\mathbf{r}_2,\ldots,\mathbf{r}_N)\phi_m(\mathbf{r},\mathbf{r}_2,\ldots,\mathbf{r}_N)\mathrm{d}^3r_3\ldots\mathrm{d}^3r_N. \qquad (3.40)$$

Using this on the operators from Eq. (3.38) gives

$$\langle n\mid\sum_i \delta(\mathbf{r}-\mathbf{r}_i)\mid m\rangle = \rho_{nm}(\mathbf{r}), \qquad (3.41)$$

$$\langle m\mid -e\sum_i V_{ext}(\mathbf{r}_i)\mid n\rangle = -e\int V_{ext}(\mathbf{r}')\rho_{mn}(\mathbf{r}')\mathrm{d}^3r', \qquad (3.42)$$

which inserted into Eq. (3.38) gives

$$n_{ind}(\mathbf{r}) = -e\int \chi(\mathbf{r},\mathbf{r}')V_{ext}(\mathbf{r}')\mathrm{d}^3r', \qquad (3.43)$$

where the density response function, $\chi(\mathbf{r},\mathbf{r}')$, is given as

$$\chi(\mathbf{r},\mathbf{r}') = \sum_{m,n} f_{nm}\frac{\rho_{nm}(\mathbf{r})\rho_{mn}(\mathbf{r}')}{\hbar\omega - E_m + E_n}. \qquad (3.44)$$

If instead of many-body states, single particle states are analysed, the potential affecting the perturbation is given as the external field, and an induced Coulomb potential between the electrons, thus $V_{tot}(\mathbf{r}) = V_{ext}(\mathbf{r}) + V_{ind}(\mathbf{r})$. For a single particle, the density operator becomes $\delta(\mathbf{r}-\mathbf{r}_1)$ and the Hamiltonian becomes $-eV_{tot}(\mathbf{r})$, giving the induced density as

$$n_{ind}(\mathbf{r}) = -\sum_{m,n} f_{nm}\frac{\langle n|\delta(\mathbf{r}-\mathbf{r}_1)|m\rangle\,\langle m|-eV_{tot}(\mathbf{r}_1)|n\rangle}{E_m - E_n - \hbar\omega} = -e\int \chi^s(\mathbf{r},\mathbf{r}')V_{tot}(\mathbf{r}')\mathrm{d}^3r', \qquad (3.45)$$

where $\chi^s(\mathbf{r},\mathbf{r}')$ is the single particle density response and is

$$\chi^s(\mathbf{r},\mathbf{r}') = \sum_{m,n} f_{nm}\frac{\phi_n^*(\mathbf{r})\phi_n(\mathbf{r}')\phi_m^*(\mathbf{r}')\phi_m(\mathbf{r})}{\hbar\omega - E_m + E_n}. \qquad (3.46)$$

The induced potential from the perturbations, is given through the Coulomb potential as

$$V_{ind} = -\frac{e}{4\pi\varepsilon}\int \frac{n_{ind}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}^3r'. \qquad (3.47)$$

By definition the total and external potentials are related by the dielectric function [26],

$$V_{ext}(\mathbf{r}) = \int \varepsilon(\mathbf{r}, \mathbf{r}')V_{tot}(\mathbf{r}')\mathrm{d}^3r'. \tag{3.48}$$

Combining this with $V_{tot} = V_{ext} + V_{ind}$ gives

$$\varepsilon(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') - \frac{e^2}{4\pi\varepsilon} \int \frac{1}{|\mathbf{r} - \mathbf{r}''|} \chi^s(\mathbf{r}', \mathbf{r}'')\mathrm{d}^3r''. \tag{3.49}$$

Assuming a periodic structure, the wave functions can be written as Bloch waves, giving

$$\phi_n(\mathbf{r}) = u_{n,\mathbf{k}}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}, \qquad \phi_m(\mathbf{r}) = u_{m,\mathbf{k}+\mathbf{q}}(\mathbf{r})e^{i(\mathbf{k}+\mathbf{q})\cdot\mathbf{r}}, \tag{3.50}$$

where u are periodic with the same periodicity as the lattice. As the product of u_n and u_m is also periodic with the lattice, it can be written as a Fourier series on the form

$$u^*_{n,\mathbf{k}}(\mathbf{r})u_{m,\mathbf{k}+\mathbf{q}}(\mathbf{r}) = \sum_{\mathbf{G}} \Phi_{n,m,\mathbf{k},\mathbf{q}}(\mathbf{G})e^{i\mathbf{G}\cdot\mathbf{r}}, \tag{3.51}$$

where $\mathbf{G}$ is the reciprocal lattice vector. Inserting these definitions into Eq. (3.46), and integrating over $\mathbf{q}$ and $\mathbf{k}$ to account for every solution, gives

$$\chi^s(\mathbf{r}, \mathbf{r}') = \sum_{\mathbf{G},\mathbf{G}'} \int \chi^s_{\mathbf{G},\mathbf{G}'}(\mathbf{q}, \omega)e^{i(\mathbf{G}\cdot\mathbf{r}-\mathbf{G}'\cdot\mathbf{r}')}e^{i\mathbf{q}\cdot(\mathbf{r}-\mathbf{r}')}\mathrm{d}\mathbf{q}, \tag{3.52}$$

where

$$\chi^s_{\mathbf{G},\mathbf{G}'}(\mathbf{q}, \omega) = 2\sum_{n,m} \int [f(E_{n,\mathbf{k}}) - f(E_{m,\mathbf{k}+\mathbf{q}})] \frac{\Phi_{n,m,\mathbf{k},\mathbf{q}}(\mathbf{G})\Phi^*_{n,m,\mathbf{k},\mathbf{q}}(\mathbf{G}')}{\hbar\omega + E_{n,\mathbf{k}} - E_{m,\mathbf{k}+\mathbf{q}}}\mathrm{d}\mathbf{k}, \tag{3.53}$$

where the factor 2 comes from spin. Both $\mathbf{k}$ and $\mathbf{q}$ are within the Brillouin zone. Inserting this expression for $\chi^s(\mathbf{r}, \mathbf{r}')$ into Eq. (3.49), gives

$$\varepsilon(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') - \frac{e^2}{4\pi\varepsilon} \int \frac{1}{|\mathbf{r} - \mathbf{r}''|} \sum_{\mathbf{G},\mathbf{G}'} \int \chi^s_{\mathbf{G},\mathbf{G}'}(\mathbf{q}, \omega)e^{i(\mathbf{G}\cdot\mathbf{r}'-\mathbf{G}'\cdot\mathbf{r}'')}e^{i\mathbf{q}\cdot(\mathbf{r}'-\mathbf{r}'')}\mathrm{d}\mathbf{q}\mathrm{d}\mathbf{r}''. \tag{3.54}$$

For the potential, $\frac{1}{4\pi|\mathbf{r}-\mathbf{r}''|}$ the inverse Fourier transform is used on the Fourier transform, where the derivation of the Fourier transform can be found in Appendix A.3. For the three dimensional case it gives

$$\frac{1}{4\pi|\mathbf{r} - \mathbf{r}''|} = \frac{1}{(2\pi)^3} \sum_{\mathbf{G}} \int \frac{e^{-i(\mathbf{p}+\mathbf{G})\cdot(\mathbf{r}-\mathbf{r}'')}}{|\mathbf{p} + \mathbf{G}|^2}\mathrm{d}\mathbf{p}, \tag{3.55}$$

where $\mathbf{p}$ is within the Brillouin zone. Inserting this into Eq. (3.54) gives

$$\begin{aligned}
&\varepsilon(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \\
&\quad - \frac{e^2}{(2\pi)^3\varepsilon} \sum_{\mathbf{G},\mathbf{G}'} \int\int\int \chi^s_{\mathbf{G},\mathbf{G}'}(\mathbf{q}, \omega)e^{i(\mathbf{G}\cdot\mathbf{r}'-\mathbf{G}'\cdot\mathbf{r}'')}e^{i\mathbf{q}\cdot(\mathbf{r}'-\mathbf{r}'')}\frac{e^{-i(\mathbf{p}+\mathbf{G}')\cdot(\mathbf{r}-\mathbf{r}'')}}{|\mathbf{p} + \mathbf{G}'|^2}\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}\mathrm{d}\mathbf{r}'' \\
&= \delta(\mathbf{r} - \mathbf{r}') - \frac{e^2}{(2\pi)^3\varepsilon} \sum_{\mathbf{G},\mathbf{G}'} \int\int \frac{\chi^s_{\mathbf{G},\mathbf{G}'}(\mathbf{q}, \omega)}{|\mathbf{p} + \mathbf{G}'|^2}e^{i(\mathbf{G}\cdot\mathbf{r}'-\mathbf{G}'\cdot\mathbf{r})}e^{i(\mathbf{q}\cdot\mathbf{r}'-\mathbf{p}\cdot\mathbf{r})} \int e^{i(\mathbf{q}-\mathbf{p})\cdot\mathbf{r}''}\mathrm{d}\mathbf{r}''\mathrm{d}\mathbf{p}\mathrm{d}\mathbf{q}.
\end{aligned} \tag{3.56}$$

The integral over $\mathrm{d}\mathbf{r}''$ gives $\delta(\mathbf{q}-\mathbf{p})$ thus the equation simplifies to

$$\varepsilon(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}-\mathbf{r}') - \frac{e^2}{(2\pi)^3\varepsilon}\sum_{\mathbf{G},\mathbf{G}'}\int \frac{\chi^s_{\mathbf{G},\mathbf{G}'}(\mathbf{q},\omega)}{|\mathbf{q}+\mathbf{G}|^2} e^{i(\mathbf{G}\cdot\mathbf{r}'-\mathbf{G}\cdot\mathbf{r})} e^{i\mathbf{q}\cdot(\mathbf{r}'-\mathbf{r})}\mathrm{d}\mathbf{q}, \tag{3.57}$$

where $\mathbf{G}$ and $\mathbf{G}'$ have been swapped. The Fourier transform of $\varepsilon(\mathbf{r},\mathbf{r}')$ is sought, and can be found by

$$\varepsilon(\mathbf{q},\omega) = \int\int \varepsilon(\mathbf{r},\mathbf{r}') e^{-i\mathbf{q}\cdot(\mathbf{r}'-\mathbf{r})}\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}', \tag{3.58}$$

thus

$$\begin{aligned}\varepsilon(\mathbf{q},\omega) = &\int\int e^{-i\mathbf{q}\cdot(\mathbf{r}'-\mathbf{r})}\delta(\mathbf{r}-\mathbf{r}')\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}' \\ &- \frac{e^2}{(2\pi)^3\varepsilon}\sum_{\mathbf{G},\mathbf{G}'}\int\int\int \frac{\chi^s_{\mathbf{G},\mathbf{G}'}(\mathbf{q}',\omega)}{|\mathbf{q}'+\mathbf{G}|^2} e^{i(\mathbf{G}\cdot\mathbf{r}'-\mathbf{G}\cdot\mathbf{r})} e^{i\mathbf{q}'\cdot(\mathbf{r}'-\mathbf{r})} e^{-i\mathbf{q}\cdot(\mathbf{r}'-\mathbf{r})}\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\mathrm{d}\mathbf{q}'.\end{aligned} \tag{3.59}$$

The integral over the Dirac-delta function gives 1. For the second part, the contributions can be split into the parts regarding $\mathbf{r}$ and $\mathbf{r}'$, yielding

$$\varepsilon(\mathbf{q},\omega) = 1 - \frac{e^2}{(2\pi)^3\varepsilon}\sum_{\mathbf{G},\mathbf{G}'}\int \frac{\chi^s_{\mathbf{G},\mathbf{G}'}(\mathbf{q}',\omega)}{|\mathbf{q}'+\mathbf{G}|^2}\int\int e^{i(\mathbf{q}-\mathbf{q}'-\mathbf{G})\cdot\mathbf{r}} e^{i(\mathbf{q}'-\mathbf{q}+\mathbf{G}')\cdot\mathbf{r}'}\mathrm{d}\mathbf{r}\mathrm{d}\mathbf{r}'\mathrm{d}\mathbf{q}'. \tag{3.60}$$

The integrals over $\mathbf{r}$ and $\mathbf{r}'$ gives zero unless $\mathbf{q}-\mathbf{q}'-\mathbf{G} = \mathbf{q}'-\mathbf{q}+\mathbf{G}' = 0$. As both $\mathbf{q}$ and $\mathbf{q}'$ are in the Brillouin zone $|\mathbf{q}'-\mathbf{q}| < |\mathbf{G}|$ and likewise for $\mathbf{G}'$, thus $\mathbf{G} = \mathbf{G}' = 0$ is necessary for the integral to give a value different from zero. The only solution is then $\mathbf{q} = \mathbf{q}'$. Therefore, the dielectric function becomes

$$\varepsilon(\mathbf{q},\omega) = 1 - \frac{e^2}{(2\pi)^3\varepsilon q^2}\chi^s_{0,0}(\mathbf{q},\omega). \tag{3.61}$$

In order to simplify this expression, for small $\mathbf{q}$, $\Phi_{m,n,\mathbf{k},\mathbf{q}}(\mathbf{G}) \approx \delta_{m,n}\delta_{\mathbf{G},0}$ [26] and if the only significant contribution is from a single band, then $\chi^s_{00}(\mathbf{q},\omega)$ can be approximated by the Lindhard function ,[26], given as

$$L(\mathbf{q},\omega) = 2\int \frac{f(E_\mathbf{k}) - f(E_{\mathbf{k}+\mathbf{q}})}{\hbar\omega + E_\mathbf{k} - E_{\mathbf{k}+\mathbf{q}}}\mathrm{d}\mathbf{k}. \tag{3.62}$$

Looking at the long wavelength limit, where $q \propto \frac{1}{\lambda} \to 0$, and assuming the energy follows a parabolic dispersion, $E = \frac{\hbar^2 k^2}{2m^*}$, where m^* is the effective mass, the term $E_\mathbf{k} - E_{\mathbf{k}+\mathbf{q}}$ can be approximated with

$$E_\mathbf{k} - E_{\mathbf{k}+\mathbf{q}} = \frac{\hbar^2 k^2}{2m^*} - \frac{\hbar^2}{2m^*}(k^2 + 2\mathbf{k}\cdot\mathbf{q} + q^2) \approx -\frac{\hbar^2\mathbf{k}\cdot\mathbf{q}}{m^*}. \tag{3.63}$$

The Fermi-Dirac distributions can be approximated by [25]

$$f(E_\mathbf{k}) - f(E_{\mathbf{k}+\mathbf{q}}) = f(E_\mathbf{k}) - (f(E_\mathbf{k}) + \mathbf{q}\cdot\nabla_\mathbf{k} f(E_\mathbf{k}) - \ldots) \approx -\mathbf{q}\cdot\nabla_\mathbf{k} f(E_\mathbf{k}). \tag{3.64}$$

Inserting these approximations into Eq. (3.62) gives

$$L(\mathbf{q},\omega) \approx -2\sum_i\int \frac{q_i\frac{\partial f(E_\mathbf{k})}{\partial k_i}}{\hbar\omega - \hbar^2\frac{\mathbf{k}\cdot\mathbf{q}}{m^*}}\mathrm{d}^3k = -\frac{2m^*}{\hbar^2 w}\sum_i\int \frac{q_i\frac{\partial f(E_\mathbf{k})}{\partial k_i}}{1 - \frac{\mathbf{k}\cdot\mathbf{q}}{w}}\mathrm{d}^3k, \tag{3.65}$$

where i is the components of $\mathbf{k}$ and $w = \frac{\hbar^2\omega}{m^*}$. As $\mathbf{q} \to 0$, $\frac{\mathbf{k}\cdot\mathbf{q}}{w} < 1$, thus the denominator can be written as

$$\left(1 - \frac{\mathbf{k}\cdot\mathbf{q}}{w}\right)^{-1} \approx \left(1 + \frac{\mathbf{k}\cdot\mathbf{q}}{w}\right). \tag{3.66}$$

Inserting this into Eq. (3.65) gives

$$L(\mathbf{q},\omega) \approx -\frac{2m^*}{\hbar^2 w}\sum_i \int q_i \frac{\partial f(E_{\mathbf{k}})}{\partial k_i}\left(1 + \frac{\mathbf{k}\cdot\mathbf{q}}{w}\right) \mathrm{d}^3k. \tag{3.67}$$

The first term in the bracket gives 0, as after the integration it corresponds to taking $f(E_{\mathbf{k}})$ for $k \to \infty$, which gives zero. The second term is then solved using integration by parts, yielding

$$L(\mathbf{q},\omega) \approx -\frac{2m^*}{\hbar^2 w^2}\left\{\sum_i \left[q_i f(E_{k_i})k_i q_i\right]_{k_i=-\infty}^{\infty} - \int q_i f(E_{k_i}) q_i \mathrm{d}k_i\right\}. \tag{3.68}$$

The first part is zero, as it corresponds to taking the distribution function for k at ∞, which is zero. In order to solve the integral, assume a square box containing all the possible states, with length L. Then the distance between each state is $\frac{2\pi}{L}$, thus the area occupied by each state will be $\left(\frac{2\pi}{L}\right)^3$. The amount of electrons then becomes

$$N = \frac{2}{\left(\frac{2\pi}{L}\right)^3}\int f(E_k) d^3k, \tag{3.69}$$

where the 2 comes from spin. Isolating the integral gives

$$\int f(E_k) d^3k = \frac{N}{V}\frac{(2\pi)^3}{2} = n\frac{(2\pi)^3}{2}, \tag{3.70}$$

where the box have been changed into a container with volume V. Thus the Lindhard function for long wavelengths become

$$L(\mathbf{q},\omega) \approx \frac{(2\pi)^3 m^* q^2 n}{\hbar^2 w^2} = \frac{(2\pi)^3 q^2 n}{m^*\omega^2}. \tag{3.71}$$

Inserting this expression into (3.61) gives

$$\varepsilon(\mathbf{q},\omega) = 1 - \frac{e^2}{(2\pi)^3\varepsilon q^2}\frac{(2\pi)^3 q^2 n}{m^*\omega^2} = 1 - \frac{2e^2 n}{\varepsilon m^*\omega}. \tag{3.72}$$

The transverse plasma modes are when $\varepsilon = 0$, thus

$$\omega_p = \sqrt{\frac{e^2 n}{\varepsilon m^*}}, \tag{3.73}$$

which was the same as found in Sec. 3.2. For a 2D structure, the calculation are similar, however the inserted Fourier transform of the Coulomb potential is different, the derivation can be found in Appendix A.3, giving

$$\varepsilon(\mathbf{q},\omega) = 1 - \frac{e^2}{2\varepsilon q(2\pi)^2}\frac{q^2 n(2\pi)^2}{m^*\omega^2} = 1 - \frac{e^2 q n}{2\varepsilon m^*\omega^2}, \tag{3.74}$$

which makes the plasma frequency

$$\omega_p = \sqrt{\frac{e^2 n}{2\varepsilon m^*}q}. \tag{3.75}$$

These correspond to the expressions stated in [23].

3.3 Plasma Wave Instability

This section is based on references [5], [23] and [27].

In this section the main mechanisms behind the generation of spontaneous terahertz radiation, termed the Dyakonov-Shur instability first described by M. Dyakonov and M. Shur [5], will be presented. For short HEMTs the electrons have practically zero phonon or impurity scattering as the transit time is shorter than the mean relaxation time. The main type of scattering is electron-electron collisions and due to the high electron concentration, the electrons can not be assumed as ballistic particles. Thus the electrons behave as a fluid in a channel with no external friction and can be described through hydrodynamic equations.

It will be shown that the steady state of this current-carrying HEMT can be unstable due to the growth of plasma waves at terahertz frequencies, which can be exploited for a variety of applications such as detectors, mixers and frequency multipliers [28][29].

Consider the HEMT of Fig. 3.1, but with a 2DEG at the interface of the substrate and insulator. The electron sheet and the gate electrode form a capacitor, which above the threshold voltage controls the surface concentration in the channel following Eq. (3.17), as

$$n_s = CU/e, \tag{3.76}$$

where $U = U_g - U_{th}$ and C is the capacitance per unit area given by Eq. (3.16). The Dyakonov-Shur model is a classical model of the conduction channel. The model is described by the surface concentration given by Eq. (3.76) and the electron velocity v as a function of position and time. Limiting the motion to the source-drain axis, say the x-axis, the fields obey the equation of motion

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial x} = -\frac{e}{m}\frac{\partial U}{\partial x}, \tag{3.77}$$

where $\partial U/\partial x$ is the longitudinal electric field in the channel, $v(x,t)$ is the local electron velocity and m is the effective mass. Eq. (3.77) is called the Euler equation and its derivation is done in App. A.4 . Furthermore, the fields also obey the continuity equation so Eq. (3.77) has to be solved together with

$$\frac{\partial n_s}{\partial t} + \frac{\partial (n_s v)}{\partial x} = 0, \tag{3.78}$$

which by using Eq. (3.76) can be rewritten as

$$\frac{\partial U}{\partial t} + \frac{\partial (Uv)}{\partial x} = 0, \tag{3.79}$$

which is derived in App. A.4. These equations coincide with the hydrodynamic equations governing shallow water.

The aim of this section is to study the time dependent effect a small perturbation has on the steady state of the current. The instability of the steady state occurs for a uniform flow of current through the channel, while being subject to specific boundary conditions. At first assume that the gate voltage is fixed and no current is running through the channel.

When analysing instabilities, the potential, the velocity, and the density are described as a sum of a large dc part and a small ac part [30], i.e. $(U, v, n_s) = (U_0, v_0, n_0) + (U_1, v_1, n_1)$, where $(U_1, v_1, n_1) \sim \exp(-i\omega t)$. Applying this in Eq. (3.77) and Eq. (3.79) yields

$$\frac{\partial(v_0 + v_1)}{\partial t} + (v_0 + v_1)\frac{\partial(v_0 + v_1)}{\partial x} + \frac{e}{m}\frac{\partial(U_0 + U_1)}{\partial x} = 0, \tag{3.80}$$

and

$$\frac{\partial(U_0 + U_1)}{\partial t} + \frac{\partial(v_0 + v_1)(U_0 + U_1)}{\partial x} \tag{3.81}$$

$$= \frac{\partial(U_0 + U_1)}{\partial t} + \frac{\partial v_0 U_0}{\partial x} + \frac{\partial v_0 U_1}{\partial x} + \frac{\partial v_1 U_0}{\partial x} + \frac{\partial v_1 U_1}{\partial x} = 0. \tag{3.82}$$

Linearising these equations with respect to v_1 and U_1, i.e. ignoring products of small quantities, and the assumption that no current runs through the channel ($v_0 = 0$), reduces them to

$$\frac{\partial v_1}{\partial t} + \frac{e}{m}\frac{\partial U_1}{\partial x} = 0, \tag{3.83}$$

$$\frac{\partial U_1}{\partial t} + U_0\frac{\partial v_1}{\partial x} = 0. \tag{3.84}$$

Combining these equations yields the following wave equation

$$\frac{\partial^2 U_1}{\partial t^2} - s^2\frac{\partial^2 U_1}{\partial x^2} = 0, \tag{3.85}$$

with the dispersion relation $\omega = \pm sk$, where $s = \sqrt{U_0 e/m}$, as was obtained in Sec. 3.2 for gated 2DEG. A similar equation can be obtained for v_1. The dispersion relation is the same as for shallow water with $U_0 e/m$ in place of hg [31], further emphasizing the similarities. If now the electrons move with a drift velocity, $v_0 \neq 0$, two more terms have to be kept from Eqs. (3.80) and (3.82) when linearising, which then becomes

$$\frac{\partial v_1}{\partial t} + v_0\frac{\partial v_1}{\partial x} + \frac{e}{m}\frac{\partial U_1}{\partial x} = 0, \tag{3.86}$$

$$\frac{\partial U_1}{\partial t} + v_0\frac{\partial U_1}{\partial x} + U_0\frac{\partial v_1}{\partial x} = 0 \Rightarrow \frac{\partial v_1}{\partial x} = -\frac{1}{U_0}\frac{\partial U_1}{\partial t} - \frac{v_0}{U_0}\frac{\partial U_1}{\partial x}. \tag{3.87}$$

Differentiating Eq. (3.86) with respect to x and substituting in Eq. (3.87) yields

$$s^2\frac{\partial^2 U_1}{\partial x^2} - v_0^2\frac{\partial^2 U_1}{\partial x^2} - 2v_0\frac{\partial^2 U_1}{\partial t \partial x} - \frac{\partial^2 U_1}{\partial t^2} = 0. \tag{3.88}$$

Which gives a quadratic function in k (or ω) on the form

$$k^2(v_0^2 - s^2) - 2v_0\omega k + \omega^2 = 0, \tag{3.89}$$

Using the regular quadratic formula for its roots the dispersion relation becomes

$$k = \frac{2v_0\omega \pm \sqrt{4v_0^2\omega^2 - 4\omega^2(v_0^2 - s^2)}}{2(v_0^2 - s^2)} = \omega\frac{v_0 \pm s}{v_0^2 - s^2} = \frac{\omega}{v_0 \mp s}. \tag{3.90}$$

It is clear from the change in the dispersion relation that the waves are in a way either carried along or slowed down by the flow. It will now be shown that under asymmetric boundary conditions this velocity difference leads to an instability, giving that the steady state is unstable against spontaneous generation of plasma waves. The boundary conditions under consideration are those given in Ref. [5], which are fixed voltage at the source and fixed current at the drain: $U = U_0$ at $x = 0$ and $J = J_0$ at $x = L$. Applying Eq. (3.76) and the assumption that the potential, velocity and density can be written on the form $(U, v, n_s) = (U_0, v_0, n_0) + (U_1, v_1, n_1)$, the boundary conditions can restated as

$$U_1(0) = 0, \tag{3.91}$$

$$U_0 v_1(L) + v_0 U_1(L) = 0. \tag{3.92}$$

Solutions for v_1 and U_1 to the linearised equations are then sought after as the sum of two waves propagating from the source to the drain and the drain to the source (under the assumption that $s > v_0$), with wave vectors k_+ and k_-

$$v_1 = Ae^{ik_+x} + Be^{ik_-x}, \tag{3.93}$$

$$U_1 = Ce^{ik_+x} + De^{ik_-x}. \tag{3.94}$$

Applying Eq. (3.86), the complexity of the problem can be reduced by relating v_1 to U_1, therefore Eqs. (3.93) can be written as

$$v_1 = \frac{ek_+}{m(\omega - v_0k_+)} Ce^{ik_+x} + \frac{ek_-}{m(\omega - v_0k_-)} De^{ik_-x}, \tag{3.95}$$

The constants can now be determined from the boundary conditions. The first boundary condition $U_1(x = 0) = 0$ yields

$$D = -C, \tag{3.96}$$

which can be applied for the second condition $U_0 v_1(L) + v_0 U_1(L) = 0$ resulting in

$$\frac{s^2k_+}{(\omega - v_0k_+)} Ce^{ik_+L} - \frac{s^2k_-}{(\omega - v_0k_-)} Ce^{ik_-L} + v_0Ce^{ik_+L} - v_0Ce^{ik_-L} = 0. \tag{3.97}$$

The goal is to find an expression for ω. To this end the exponential terms are isolated in the following manner

$$\begin{aligned} &\frac{s^2k_+}{(\omega - v_0k_+)} e^{i(k_+-k_-)L} - \frac{s^2k_-}{(\omega - v_0k_-)} + v_0e^{i(k_+-k_-)L} - v_0 \\ &\qquad = e^{i(k_+-k_-)L}\left(\frac{s^2k_+}{\omega - v_0k_+} + v_0\right) - \frac{s^2k_-}{\omega - v_0k_-} - v_0 = 0 \end{aligned} \tag{3.98}$$

$$\Rightarrow e^{i(k_+-k_-)L} = \frac{\frac{s^2k_-}{\omega - v_0k_-} + v_0}{\frac{s^2k_+}{\omega - v_0k_+} + v_0} = \frac{\omega - v_0k_+}{\omega - v_0k_-}\frac{s^2k_- + v_0(\omega - v_0k_-)}{s^2k_+ + v_0(\omega - v_0k_+)}. \tag{3.99}$$

By applying Eq. (3.90) on the last fraction of Eq. (3.99) can be reduced

$$\frac{s^2k_- + v_0(\omega - v_0k_-)}{s^2k_+ + v_0(\omega - v_0k_+)} = \frac{s^2\frac{\omega}{v_0-s} + v_0(\omega - v_0\frac{\omega}{v_0-s})}{s^2\frac{\omega}{v_0+s} + v_0(\omega - v_0\frac{\omega}{v_0+s})} =$$
$$\frac{s^2\omega(v_0+s) + v_0\omega(v_0+s)(v_0-s) - \omega v_0^2(v_0+s)}{s^2\omega(v_0-s) + v_0\omega(v_0+s)(v_0-s) - \omega v_0^2(v_0-s)} =$$
$$\frac{s^3\omega - s\omega v_0^2}{s\omega v_0^2 - s^3\omega} = -1. \tag{3.100}$$

This results in the expression

$$e^{i(k_+-k_-)L} = -\frac{\omega - v_0k_+}{\omega - v_0k_-}. \tag{3.101}$$

By inserting the expressions for in k_+ and k_-, the exponent on the left side becomes

$$(k_+ - k_-)L = \left(\frac{\omega}{v_0+s} - \frac{\omega}{v_0-s}\right)L = \frac{2sL}{s^2 - v_0^2}\omega, \tag{3.102}$$

and the right side becomes

$$-\frac{\omega - v_0k_+}{\omega - v_0k_-} = -\frac{\omega - v_0\frac{\omega}{v_0+s}}{\omega - v_0\frac{\omega}{v_0-s}} = -\frac{\omega s v_0 - \omega s^2}{\omega(v_0+s)(v_0-s) - \omega v_0(v+0-s)}$$
$$= -\frac{\omega v_0 s - \omega s^2}{-\omega s^2 - v_0 s\omega} = \frac{v_0 - s}{v_0 + s}. \tag{3.103}$$

Eq. (3.101) can be written on the form

$$e^{i\frac{2sL}{s^2-v_0^2}\omega} = \frac{v_0 - s}{v_0 + s}, \tag{3.104}$$

which defines the complex frequency as a function of v_0 and s. Finding an explicit expression for ω is done by first taking the natural logarithm of Eq. (3.104). However, when taking the logarithm of a complex number a bit of care has to be taken, especially on the right-hand side. First note that any complex number can be written on the form $w = re^{i\phi}$, where $r = |w|$ and ϕ, called the argument of w or the phase, is the angle between the point in the complex plane and the positive real axis. The left side written in this form is self explanatory. The angle of the right side will, however, depend on whether or not the fraction yields a positive or negative number, more precisely on the relative size between v_0 and s. Since s is positive the case will either be $|v_0| < s$, in which case the fraction will be negative and the phase will be $\phi = \pi$, or $|v_0| > s$ and thus $\phi = 0$. Since adding $2\pi n$, where n is a integer, to the phase of a complex number gives the same complex number, a complex number will also have an infinite number of solutions, which are logarithms of it, stated as

$$\ln(w) = \ln(r) + i(\phi + 2\pi n). \tag{3.105}$$

With this in mind taking the logarithm of Eq. (3.104) becomes either

$$i\left(\frac{2sL}{s^2 - v_0^2}\omega + 2\pi n'\right) = \ln\left|\frac{v_0 - s}{v_0 + s}\right| + i(\pi + 2\pi m) \Leftrightarrow$$
$$\omega = -i\frac{s^2 - v_0^2}{2Ls}\ln\left|\frac{v_0 - s}{v_0 + s}\right| + \frac{s^2 - v_0^2}{2Ls}\pi(1 + 2(m - n'))$$
$$= i\frac{s^2 - v_0^2}{2Ls}\ln\left|\frac{v_0 + s}{v_0 - s}\right| + \frac{s^2 - v_0^2}{2Ls}\pi(1 + 2n) \tag{3.106}$$

for $|v_0| < s$, where $n = m - n'$ is an integer, or

$$\omega = i\frac{s^2 - v_0^2}{2Ls}\ln\left|\frac{v_0 + s}{v_0 - s}\right| + \frac{s^2 - v_0^2}{2Ls}2\pi n \tag{3.107}$$

for $|v_0| > s$. Eq. (3.106) and (3.107) can be combined into a single expression and by writing the real and imaginary part separately yields

$$w' = \frac{s^2 - v_0^2}{2Ls}\pi n, \tag{3.108}$$

$$\gamma = \frac{s^2 - v_0^2}{2Ls}\ln\left|\frac{v_0 + s}{v_0 - s}\right|, \tag{3.109}$$

where n is an odd integer for $|v_0| < s$ and an even integer for $|v_0| > s$. There are two regions of stable flow and two regions of unstable flow, which can be deduced from Eq. (3.109). For positive v_0 the steady flow will become unstable if $s > v_0$ and remain stable if $s < v_0$. The stable region can be difficult to reach as the electron drift velocity will normally saturate at the order of 10^7 cm/s due to emission of optical phonons [5], whereas s is typically on the order of 10^8 cm/s [23]. In the case of negative v_0, which is obtained simply by interchanging the boundary conditions at the source and drain [5], the flow is stable when $s > |v_0|$ and unstable for $s < |v_0|$.

As the negative v_0 is symmetric to positive v_0 and only the unstable region is of interest, only the case with $s > v_0 > 0$ will be analysed.

Writing the wave increment, i.e. γ, in units of $s/2L$ it only becomes dependent on v_0/s, which makes it clear that in the limit $v_0/s \ll 1$, γ reduces to

$$\frac{\gamma}{(s/2L)} = \left(1 - \frac{v_0^2}{s^2}\right)\ln\left|\frac{1 + \frac{v_0}{s}}{1 - \frac{v_0}{s}}\right| \approx \ln(1) + \frac{v_0}{s} - \left(\ln(1) - \frac{v_0}{s}\right) = 2\frac{v_0}{s} \Rightarrow \gamma \approx \frac{v_0}{L}, \tag{3.110}$$

which is the inverse of the electron transit time τ. Therefore shorter channels should be more prone to this instability, which also is consistent with the fact that the hydrodynamic description best describes short HEMTs.

3.3.1 Instability Origin

This subsection is based on reference [32].

The source of the instability can be deduced by considering the reflection of the wave at the two channel boundaries and evaluating the amplitude of the reflection coefficient at the boundaries. Here the reflection coefficient is termed as the amplitude ratio of the reflected and oncoming waves. In order to simplify the analysis, a new notation for the first order terms U_1, v_1 and n_1 is introduced, represented by the linearly independent superposition of a downstream term and an upstream term, e.g. U_+ and U_-, i.e. for the potential

$$U_1 = U_{1+} + U_{1-}, \tag{3.111}$$

and similar expressions for v_1 and n_1. Eq. (3.111) is equivalent to Eq. (3.94). Examining the reflection coefficient of the potential at the source ($x = 0$), and applying Eq. (3.96), then

$$R_{U,s} = \frac{U_{1+}(0)}{U_{1-}(0)} = -\frac{Ce^{ik_+0}}{Ce^{ik_-0}} = -1 \tag{3.112}$$

shows that the reflection at the boundary does not change the amplitude of the wave, but it does change its sign. While the reflection coefficient at the drain side ($x = L$) is equal to

$$R_{U,d} = \frac{U_{1-}(L)}{U_{1+}(L)} = -\frac{Ce^{ik_-L}}{Ce^{ik_+L}} = \frac{s+v_0}{s-v_0}, \tag{3.113}$$

where Eq. (3.104) have been used for the last equality. Hence, in the case of $s > v_0 > 0$, Eq. (3.113) will always be greater than unity which results in wave amplification from the reflection of the boundary with fixed current. The wave amplification is not solely a result of the reflection with the drain contact, which can be shown by considering the current at the boundaries, for the source:

$$R_{J,s} = \frac{J_{1+}(0)}{J_{1-}(0)} = -\frac{v_0 + s^2\frac{k_+}{\omega - v_0k_+}}{v_0 + s^2\frac{k_-}{\omega - v_0k_-}} = \frac{s+v_0}{s-v_0}, \tag{3.114}$$

where the results follows from Eq. (3.99) and Eq. (3.101) and at the drain side

$$R_{J,d} = \frac{J_{1-}(L)}{J_{1+}(L)} = -\frac{\left(v_0 + s^2\frac{k_-}{\omega - v_0k_-}\right)e^{ik_-L}}{\left(v_0 + s^2\frac{k_+}{\omega - v_0k_+}\right)e^{ik_+L}} = -1, \tag{3.115}$$

by Eq. (3.99). and expressions for $J_{1\pm}$ are obtained by

$$\begin{aligned} J_1 &= U_0v_1 + v_0U_1 \\ &= C\frac{s^2k_+}{(\omega - v_0k_+)}e^{ik_+x} - C\frac{s^2k_-}{(\omega - v_0k_-)}e^{ik_-x} + v_0Ce^{ik_+x} - v_0Ce^{ik_-x} \\ &= C\left(\frac{s^2k_+}{(\omega - v_0k_+)} + v_0\right)e^{ik_+x} - C\left(\frac{s^2k_-}{(\omega - v_0k_-)} + v_0\right)e^{ik_-x} = J_{1+} + J_{1-}. \end{aligned} \tag{3.116}$$

The reflection coefficients are reversed compared to the potential. However, the product of the reflection coefficients stays the same and this product is greater than unity if a dc current is present and becomes unity if not. As such for every reflection with either boundary the resulting wave is amplified. This is, according to Dyakonov and Shur [5][22], the mechanism for the wave amplification in the channel, which then leads to an instability of the steady state with respect to the generation of plasma waves.

3.3.2 Instability Conditions

There are two main decay mechanics which inhibit plasma wave growth from the instability: First external friction due to electron scattering by impurities or phonons, and second internal friction caused by the viscosity of the electron fluid. The external friction can be accounted for by introducing the term $-v/\tau$ to the right side of equation Eq. (3.77). This leads to the addition of $-1/(2\tau)$ to the wave increment of Eq. (3.109) [5][32], as such setting a minimum velocity for the wave instability to grow or a requirement of few scattering events doing the transit time. The viscosity, ν adds an additional damping of the increment by νk^2 [5], where k is the length of the wave vector, which in the limit

$v_0/s \ll 1$ is reduces to

$$\begin{aligned}\nu k^2 &= \nu\frac{\omega^2}{(v_0-s)^2(v_0+s)^2}(2v_0^2+2s^2) = \nu\frac{w'^2-\gamma^2+2i\gamma\omega'}{(v_0-s)^2(v_0+s)^2}(2v_0^2+2s^2)\\ &= \nu\frac{\left(\frac{s^2-v_0^2}{2Ls}\pi n\right)^2-\left(\frac{s^2-v_0^2}{2Ls}\ln\left|\frac{v_0+s}{v_0-s}\right|\right)^2+2i\frac{(s^2-v_0^2)^2}{(2Ls)^2}\pi n\ln\left|\frac{v_0+s}{v_0-s}\right|^2}{(v_0-s)^2(v_0+s)^2}(2v_0^2+2s^2)\\ &\approx \nu\frac{\pi^2n^2}{2L^2}, \end{aligned} \quad (3.117)$$

where $\ln\left|\frac{v_0+s}{v_0-s}\right| \approx 2v_0/s$ and $v_0/s \ll 1$ has been applied repeatedly. Hence, its effect becomes more prevalent for higher order modes. By comparing γ with νk^2 in the same limit for the first order modes:

$$\frac{\gamma}{\nu k^2} = \frac{2v_0L^2}{\pi^2\nu L} = \frac{Lv_0}{\nu}\frac{2}{\pi^2}, \quad (3.118)$$

the effect of the viscosity can be considered small when $Lv_0/\nu \gg 1$.

The instability will occur when the electron velocity, and hence, the current, exceeds a certain threshold value, which in the limit $v_0/s \ll 1$ is given by

$$\frac{v_0}{L} > \frac{1}{2\tau} + \nu\frac{\pi^2n^2}{2L^2}. \quad (3.119)$$

Once the threshold is exceeded, plasma waves oscillations should grow, resulting in a periodic variation of the channel charge, i.e. a periodic change of the dipole moment. This variation can then lead to electromagnetic radiation [33].

Method 4

4.1 Guided Modes

This section is based on reference [10].

Scattered light near a waveguide can result in the excitation of guided modes in the waveguide. As the 2DEG exhibits metal-like properties, it is possible to find guided modes for the structure in the same manner as for metal strips. The guided modes are described through their mode index, n_m, which is the refractive index for the x-dependent part of the wave. In order to determine the mode index for these guided modes, the structure is analysed. The guided modes for an ungated are found using transfer matrices, and is explained in App. A.5.

4.1.1 Gated Structure

For the gated structure, the gate is assumed to be a perfect conductor, thus no field exist in and above the gate. The gated structure is illustrated in Fig. 4.1.

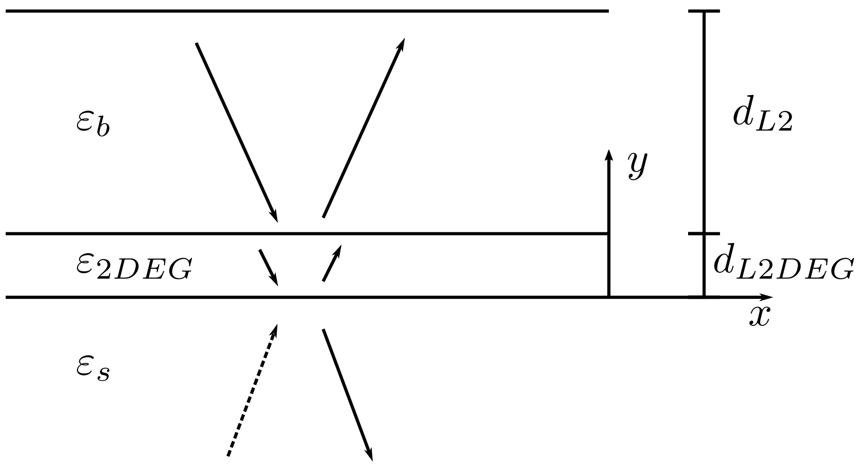

Figure 4.1. The HEMT structure with a gate, which is assumed to be a perfect conductor. The dashed lines are the fields which are zero for guided modes.

The 2DEG needs a dielectric constant of its own in order to describe the response in the layer. This is done by considering the Drude model, derived in App. A.6, with the effective mass of the electron, in order to account for the material, is used to describe the conductivity of the 2DEG. Using Eq. (2.4) and that $\mathbf{J}_f = \sigma\mathbf{E}$, the relative permittivity can be found as

$$\nabla \times \mathbf{H} = \mathbf{J}_f + \frac{\partial \mathbf{D}}{\partial t} = \sigma\mathbf{E} - i\omega\varepsilon_0\varepsilon\mathbf{E} = -i\omega\varepsilon_0 \left(\varepsilon + i\frac{\sigma}{\omega\varepsilon_0}\right)\mathbf{E}, \tag{4.1}$$

making the relative permittivity

$$\varepsilon_{2DEG} = \varepsilon_s + i\frac{\sigma}{\omega\varepsilon_0}. \tag{4.2}$$

As σ is found through the sheet conductivity, when applying it to a layer with a finite thickness, σ becomes σ/d_{L2DEG}.

The field can now be split into three regions,

$$H = Ae^{-ik_{y,s}y}e^{ik_x x} \qquad \text{for} \quad y < 0, \tag{4.3}$$

$$H = \left(B_1 e^{ik_{y,2DEG}y} + B_2 e^{-ik_{y,2DEG}y}\right) e^{ik_x x} \qquad \text{for} \quad 0 < y < d_{L2DEG}, \tag{4.4}$$

$$H = \left(C_1 e^{ik_{y,b}y} + C_2 e^{-ik_{y,b}y}\right) e^{ik_x x} \qquad \text{for} \quad d_{L2DEG} < y < d_{tot}, \tag{4.5}$$

where $d_{tot} = d_{L2} + d_{L2DEG}$. For the boundary conditions, the tangential component of the electric and magnetic field must be continuous, where the electric field can be found using Eq. (4.8). For the boundary at $y = d_{tot}$, $E_x = 0$, as the gate is a perfect conductor, leading to

$$C_1 e^{ik_{y,b}d_{tot}} - C_2 e^{-ik_{y,b}d_{tot}} = 0 \Leftrightarrow C_1 = C_2 e^{-2ik_{y,b}d_{tot}}. \tag{4.6}$$

For the boundary at $y = d_{L2DEG}$ the boundary condition for the tangential component of the magnetic field gives

$$C_1 e^{ik_{y,b}d_{L2DEG}} + C_2 e^{-ik_{y,b}d_{L2DEG}} = B_1 e^{ik_{y,2DEG}d_{L2DEG}} + B_2 e^{-ik_{y,2DEG}d_{L2DEG}}. \tag{4.7}$$

The other boundary condition is for the tangential component of the electric field, which should be conserved. The electric field is related to the magnetic field through Eq. (2.4), where $\mathbf{J}_f = 0$, as

$$E_x = \frac{i}{\omega\varepsilon_0\varepsilon_r}\frac{\partial H_z}{\partial y}, \tag{4.8}$$

giving the boundary condition as

$$\frac{k_{y,b}}{\varepsilon_b}\left(C_2 e^{-ik_{y,b}d_{L2DEG}} - C_1 e^{ik_{y,b}d_{L2DEG}}\right) = \frac{k_{y,2DEG}}{\varepsilon_{2DEG}}\left(B_2 e^{-ik_{y,2DEG}d_{L2DEG}} - B_1 e^{ik_{y,2DEG}d_{L2DEG}}\right). \tag{4.9}$$

The tangential component of the magnetic field at $y = 0$ gives

$$A = B_1 + B_2, \tag{4.10}$$

and the electric field

$$\frac{k_{y,s}}{\varepsilon_s}A = \frac{k_{y,2DEG}}{\varepsilon_{2DEG}}\left(B_2 - B_1\right). \tag{4.11}$$

Combining Eqs. (4.10) and (4.11) yields

$$B_1 = B_2\frac{k_{y,2DEG}\varepsilon_s - k_{y,s}\varepsilon_{2DEG}}{k_{y,2DEG}\varepsilon_s + k_{y,s}\varepsilon_{2DEG}} = B_2 K_1. \tag{4.12}$$

Inserting Eqs. (4.6) and (4.12) into Eq. (4.7) gives

$$C_2\left(e^{-2ik_{y,b}d_{tot}}e^{ik_{y,b}d_{L2DEG}}+e^{-ik_{y,b}d_{L2DEG}}\right) = B_2\left(K_1e^{ik_{y,2DEG}d_{L2DEG}}+e^{-ik_{y,2DEG}d_{L2DEG}}\right), \quad (4.13)$$

from which the correlation $C_2 = K_2B_2$ can be found. Inserting Eqs. (4.6), (4.12), and (4.13) into Eq. (4.9) yields

$$\frac{k_{y,b}}{\varepsilon_b}K_2\left(e^{-ik_{y,b}d_{L2DEG}}-e^{-2ik_{y,b}d_{tot}}e^{ik_{y,b}d_{L2DEG}}\right) = \frac{k_{y,2DEG}}{\varepsilon_{2DEG}}\left(e^{-ik_{y,2DEG}d_{L2DEG}}-K_1e^{ik_{y,2DEG}d_{L2DEG}}\right). \quad (4.14)$$

Subtracting the left side from Eq. (4.14) gives the function $f(n_m)$, where guided modes are found when the equation $f(n_m) = 0$ is fulfilled. The same methods as used for the ungated structure in order to locate these mode indices, as described in App. A.5, can be used for the gated structure. The contour plot is illustrated in Fig. 4.2.

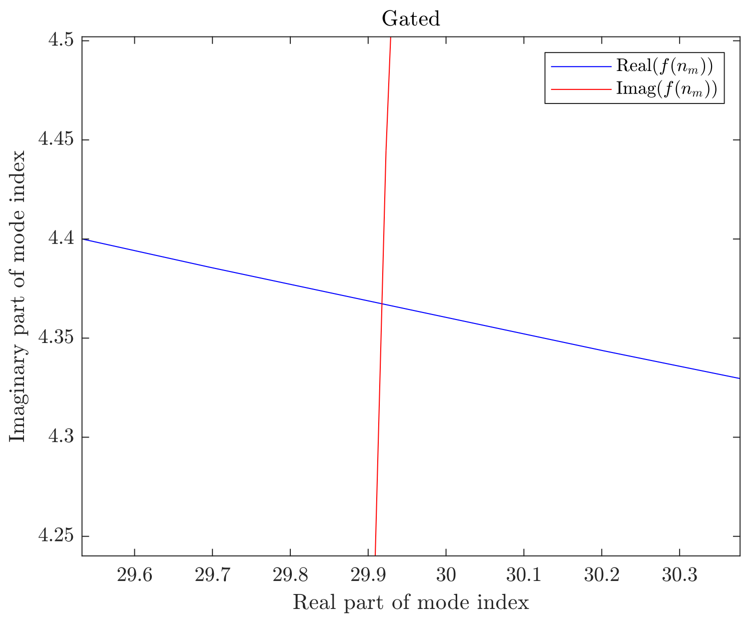

Figure 4.2. The contour plot of $f(n_m) = 0$ for a HEMT structure with a 2DEG.

From the contour plot the mode index for the guided mode was found to be $n_{m,1} \approx 29.9 + 4.4i$. Using the Newton-Raphson algorithm, described in App. A.5, gives a value of $n_m = 29.9174 + 4.3672i$ which corresponds very well with the contour plot.

Infinitely Thin 2DEG

For the gated structure, a model where the 2DEG is infinitely thin has also been constructed. The structure is illustrated in Fig. 4.3.

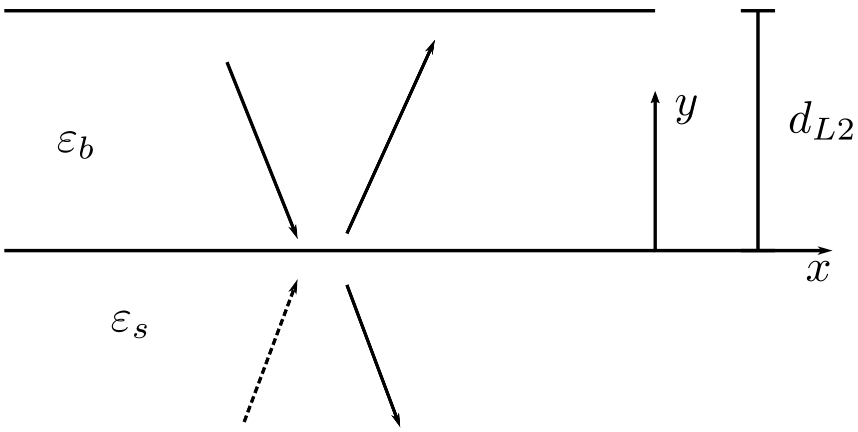

Figure 4.3. The structure of a gated HEMT, with an infinitely thin 2DEG placed at $y = 0$.

The field can be split into two separate sections,

$$H = Ae^{-ik_{y,a}y}e^{ik_x x} \qquad \text{for} \quad y < 0, \tag{4.15}$$

$$H = \left(Be^{ik_{y,b}y} + Ce^{-ik_{y,b}y}\right)e^{ik_x x} \qquad \text{for} \quad 0 < y < d_{L2}. \tag{4.16}$$

For the boundary at $y = d_{L2}$ the electric field must be 0, thus

$$C = Be^{2ik_{y,b}d_{L2}}. \tag{4.17}$$

For the boundary at $y = 0$, the change in magnetic field must be equal to the current in the 2DEG, which can be described through Ohm's law as

$$I_x(k_x, y = 0) = \sigma(\omega)E_x(k_x, y = 0), \tag{4.18}$$

where $\sigma(\omega)$ is the sheet conductivity of the plasma. The sheet conductivity can be approximated using the Drude model, described in App. A.6, giving

$$\sigma(\omega) = \frac{e^2 N\tau}{m^*(1 - i\omega\tau)}, \tag{4.19}$$

where e is the electron charge, N is the electron density per area, m^* is the effective mass of the electrons and τ is the electron-scattering time. The electric field is found through Eq. (4.8). Thus the boundary gives

$$B + C - A = \sigma E_x = \frac{\sigma k_{y,b}}{\omega\varepsilon_0\varepsilon_b}(C - B), \tag{4.20}$$

where E_x is found through Eq. (4.8). The tangential component of the electric field must also be continuous, giving

$$\frac{k_{y,b}}{\varepsilon_b}(C - B) = \frac{k_{y,s}}{\varepsilon_s}A \Leftrightarrow A = \frac{k_{y,b}\varepsilon_s}{k_{y,s}\varepsilon_b}B\left(e^{2ik_{y,b}d_{L2}} - 1\right). \tag{4.21}$$

Combining Eq. (4.20) and Eq. (4.21) gives

$$1 + e^{2ik_{y,b}d_{L2}} - \frac{k_{y,b}\varepsilon_s}{k_{y,s}\varepsilon_b}\left(e^{2ik_{y,b}d_{L2}} - 1\right) = \frac{\sigma k_{y,b}}{\omega\varepsilon_0\varepsilon_b}\left(e^{2ik_{y,b}d_{L2}} - 1\right). \tag{4.22}$$

Subtracting one side from the other gives $f(n_m)$ and solutions are found when $f(n_m) = 0$. The contour plot is illustrated in Fig. 4.4

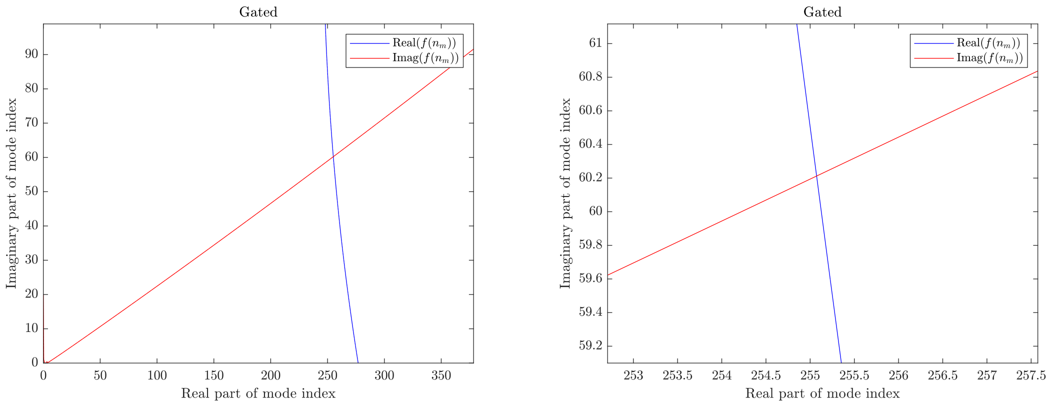

Figure 4.4. The contour plot of $f(n_m) = 0$ for a HEMT structure with an infinitely thin 2DEG. The distance between the gate and the 2DEG, is 10 nm. The right figure is focused on the mode index.

The value of the mode index is $n_m \approx 255.1 + 60.2i$. Through the Newton-Raphson algorithm the mode index is found to be $n_m = 255.07 + 60.21i$. This differs greatly from the mode index for the gated structure where the gate was assumed to be of a finite thickness.

4.2 Integral Equation Method for 2DEG

This section is based on reference [34].

In this section a method for numerical modelling the electromagnetic radiation pattern of a HEMT with an active 2DEG layer will be established. In order to construct such a model, Maxwell's equations with appropriate boundary conditions will be used to describe how the electromagnetic wave behaves through the structure. The analysed situation is illustrated in Fig. 4.5.

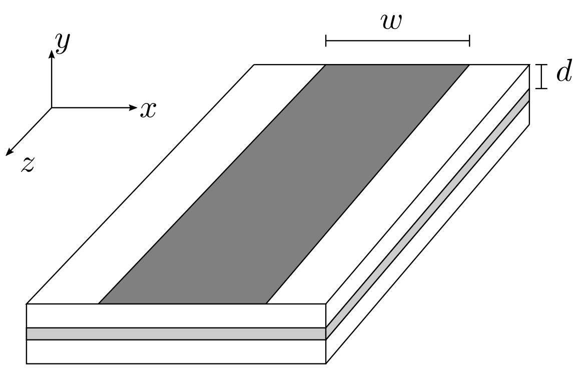

Figure 4.5. Schematic of the gated 2DEG, positioned distance d below the gate where the width of the gate is W.

Assuming an incident plane wave polarised along the x-axis, i.e. p-polarised, is travelling along the negative y-axis. Such a field is given as $E_{0,x} = E_0 e^{-ik_0 y}$. For a standard

HEMT structure, the 2DEG is in the interface between 2 different materials with different dielectric constants. Setting the 2DEG as $y = 0$, the relative dielectric constants will be

$$\varepsilon_r = \varepsilon_a = 1 \quad \text{for} \quad y > d, \tag{4.23}$$

$$\varepsilon_r = \varepsilon_b \quad \text{for} \quad 0 < y < d, \tag{4.24}$$

$$\varepsilon_r = \varepsilon_s \quad \text{for} \quad y < 0, \tag{4.25}$$

for the ambient medium, the insulator layer and the substrate respectively. The gate electrode is considered to be a perfect electric conductor and infinitely long along the z-axis, with a width of w and assumed to be infinitely thin, placed a $y = d$. The HEMT, except the gate, is assumed to be infinitely long and homogeneous along the x- and z-axes. This includes the 2DEG, which is a rough approximation, but can still provide some insight in the radiation pattern. Boundary conditions are formulated at the interfaces at, $y = 0$ and $y = d$. The tangential part of the electric field, E_x has to be continuous across the boundaries, while the difference in the tangential magnetic field is $H_z(x, y = (0^+, d^+)) - H_z(x, y = (0^-, d^-)) = I_x(x, y = (0, d))$, where $I_x(x, y = (0, d))$ is the sheet current density. The correlation between the electric and magnetic fields, is described through Eq. (4.8). Thus the incident magnetic field can be written as

$$H_{0,z} = \frac{i}{\omega\mu_0}\frac{\partial E_{0,x}}{\partial y} = \frac{k_0}{\omega\mu_0}E_0 e^{-ik_0 y} = \sqrt{\frac{\varepsilon_0}{\mu_0}}E_0 e^{-ik_0 y}, \tag{4.26}$$

where $k_0 = \frac{\omega}{c} = \omega\sqrt{\varepsilon_0\mu_0}$ has been inserted. The structure can be regarded as three separate regions with dielectric constants given by Eqs. (4.23), (4.24), and (4.25), each with its own electric and magnetic field components. In the ambient medium, $y > d$, the total field will be a combination of the incoming field and the field reflected from the structure, thus

$$H_z = H_{z,0} + H_{z,a} \quad \text{for} \quad y > d. \tag{4.27}$$

The reflected field can be represented as a sum over all plane waves propagating in the positive y-axis:

$$H_{z,a}(x, y) = \int A(k_x)e^{ik_x x}e^{ik_{y,a}y}\mathrm{d}k_x, \tag{4.28}$$

where $k_{y,a} = \sqrt{k_0^2 - k_x^2}$. In the substrate, $y < 0$, only a downwards propagating wave makes physical sense, as such the magnetic field can likewise be represented by

$$H_{z,s}(x, y) = \int D(k_x)e^{ik_x x}e^{-ik_{y,s}y}\mathrm{d}k_x, \tag{4.29}$$

where $k_{y,s} = \sqrt{k_0^2\varepsilon_s - k_x^2}$. In the insulator, $0 < y < d$, The field will have a component propagating in the positive y direction and a part propagating in the negative y direction. Such a field can be written on the form

$$H_{z,b}(x, y) = \int \left(iB(k_x)\sin(k_{y,b}y) + C(k_x)\cos(k_{y,b}y)\right)e^{ik_x x}\mathrm{d}k_x, \tag{4.30}$$

where $k_{y,b} = \sqrt{k_0^2 \varepsilon_b - k_x^2}$. The fields are analysed in k space, by using the Fourier transform on the fields, yielding

$$H_z(k_x, y) = \frac{1}{2\pi} \int H_z(x, y) e^{-ik_x x} \mathrm{d}x. \tag{4.31}$$

In order to describe the field, the coefficients, $A(k_x)$, $B(k_x)$, $C(k_x)$ and $D(k_x)$ must be found through the four boundary conditions of the system.

Boundary at $y = 0$

For the boundary at the 2DEG, $y = 0$, applying the boundary condition for the magnetic field gives the first relation between two of the coefficients as

$$H_z(k_x, y = 0^+) - H_z(k_x, y = 0^-) = C(k_x) - D(k_x) = I_x(k_x, y = 0). \tag{4.32}$$

Furthermore using the other boundary condition regarding the tangential component of the electric field and that the electric field is related to the magnetic field through Eq. (4.8), which allows the boundary condition $E_x(k_x, y = 0^+) = E_x(k_x, y = 0^-)$ to be expressed as

$$B(k_x) \frac{k_{y,b}}{\varepsilon_b} = -D(k_x) \frac{k_{y,s}}{\varepsilon_s}. \tag{4.33}$$

Isolating for $D(k_x)$ and inserting this into Eq. (4.32) gives

$$C(k_x) + \frac{k_{y,b} \varepsilon_s}{k_{y,s} \varepsilon_b} B(k_x) = I_x(k_x, y = 0). \tag{4.34}$$

The current in the 2DEG can be described through Ohm's law as Eq. (4.18), where σ is found through Eq. (4.19). Inserting the electric field given from Eq. (4.8), using the magnetic field from Eq. (4.30), into Eq. (4.18) gives

$$I_x(k_x, y = 0) = \sigma(\omega) E_x(k_x, y = 0) = \sigma(\omega) \frac{i^2 k_{y,b}}{\omega \varepsilon_0 \varepsilon_b} B(k_x) = -\sigma(\omega) \frac{k_{y,b}}{k_0 \varepsilon_b} \sqrt{\frac{\mu_0}{\varepsilon_0}} B(k_x). \tag{4.35}$$

Inserting this expression for $I_x(k_x, y = 0)$ into Eq. (4.34) gives

$$C(k_x) = -B(k_x) \left[\frac{k_{y,b} \varepsilon_s}{k_{y,s} \varepsilon_b} + \sigma(\omega) \frac{k_{y,b}}{k_0 \varepsilon_b} \sqrt{\frac{\mu_0}{\varepsilon_0}} \right]. \tag{4.36}$$

This makes the Fourier representation

$$H_z(k_x, y) = B(k_x) \left[i \sin(k_{y,b} y) - \left\{ \frac{k_{y,b} \varepsilon_s}{k_{y,s} \varepsilon_b} + \sigma(\omega) \frac{k_{y,b}}{k_0 \varepsilon_b} \sqrt{\frac{\mu_0}{\varepsilon_0}} \right\} \cos(k_{y,b}) \right], \tag{4.37}$$

for $0 < y < d$.

Boundary at $y = d$

The last two boundary conditions are at the gate on top of the structure, at $y = d$. The first is obtained from the discontinuity across the surface of the tangential part of the magnetic field, $H_z(k_x, y = d^+) - H_z(k_x, y = d^-) = I_x(k_x, y = d)$, giving

$$\begin{aligned} I_x(k_x, y = d) = A(k_x) e^{ik_{y,a} d} &+ \sqrt{\frac{\varepsilon_0}{\mu_0}} E_0 e^{-ik_0 d} \delta(k_x) \\ &- B(k_x) \left[i \sin(k_{y,b} d) - \left\{ \frac{k_{y,b} \varepsilon_s}{k_{y,s} \varepsilon_b} + \sigma(\omega) \frac{k_{y,b}}{k_0 \varepsilon_b} \sqrt{\frac{\mu_0}{\varepsilon_0}} \right\} \cos(k_{y,b} d) \right], \end{aligned} \tag{4.38}$$

where $\sqrt{\frac{\varepsilon_0}{\mu_0}}E_0 e^{-ik_0 d}\delta(k_x)$ is the Fourier transform of the incident field, Eq. (4.26). The last boundary condition is that the tangential component of the electric field is conserved across the boundary. The electric field is found through Eq. (4.8), yielding

$$A(k_x) i k_{y,a} e^{ik_{y,a}d} - ik_0\sqrt{\frac{\varepsilon_0}{\mu_0}}E_0 e^{-ik_0 d}\delta(k_x) = \frac{k_{y,b}}{\varepsilon_b}B(k_x)\left[i\cos(k_{y,b}d) + \left\{\frac{k_{y,b}\varepsilon_s}{k_{y,s}\varepsilon_b} + \sigma(\omega)\frac{k_{y,b}}{k_0\varepsilon_b}\sqrt{\frac{\mu_0}{\varepsilon_0}}\right\}\sin(k_{y,b}d)\right]. \tag{4.39}$$

Isolating $A(k_x)e^{ik_{y,a}d}$ in Eq. (4.38) and inserting that into Eq. (4.39) leads to

$$ik_{y,a}\left(I_x(k_x, y=d) - \sqrt{\frac{\varepsilon_0}{\mu_0}}E_0 e^{-ik_0 d}\delta(k_x) + B(k_x)\left[i\sin(k_{y,b}d) - \left\{\frac{k_{y,b}\varepsilon_s}{k_{y,s}\varepsilon_b} + \sigma(\omega)\frac{k_{y,b}}{k_0\varepsilon_b}\sqrt{\frac{\mu_0}{\varepsilon_0}}\right\}\cos(k_{y,b}d)\right]\right) - ik_0\sqrt{\frac{\varepsilon_0}{\mu_0}}E_0 e^{-ik_0 d}\delta(k_x) = \frac{k_{y,b}}{\varepsilon_b}B(k_x)\left[i\cos(k_{y,b}d) + \left\{\frac{k_{y,b}\varepsilon_s}{k_{y,s}\varepsilon_b} + \sigma(\omega)\frac{k_{y,b}}{k_0\varepsilon_b}\sqrt{\frac{\mu_0}{\varepsilon_0}}\right\}\sin(k_{y,b}d)\right]. \tag{4.40}$$

Using that $k_{y,a}\delta(k_x) = k_0\delta(k_x)$ and gathering all terms containing $B(k_x)$ on the right-hand side gives

$$ik_{y,a}I_x(k_x, y=d) - 2ik_0\sqrt{\frac{\varepsilon_0}{\mu_0}}E_0 e^{-ik_0 d}\delta(k_x) = B(k_x)\left(\sin(k_{y,b}d)\left[k_{y,a} + \frac{k_{y,b}}{\varepsilon_b}\left\{\frac{k_{y,b}\varepsilon_s}{k_{y,s}\varepsilon_b} + \sigma(\omega)\frac{k_{y,b}}{k_0\varepsilon_b}\sqrt{\frac{\mu_0}{\varepsilon_0}}\right\}\right] + i\cos(k_{y,b}d)\left[\frac{k_{y,b}}{\varepsilon_b} + k_{y,a}\left\{\frac{k_{y,b}\varepsilon_s}{k_{y,s}\varepsilon_b} + \sigma(\omega)\frac{k_{y,b}}{k_0\varepsilon_b}\sqrt{\frac{\mu_0}{\varepsilon_0}}\right\}\right]\right). \tag{4.41}$$

In order to simplify the expression, four variables are defined, $Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}}$, $\chi_a = \frac{k_0}{k_{y,a}}$, $\chi_b = \frac{\varepsilon_b k_0}{k_{y,b}}$, and $\chi_s = \frac{\varepsilon_s k_0}{k_{y,s}} + Z_0\sigma(\omega)$. Using these variables and multiplying both the numerator and denominator with $\frac{1}{ik_0}$, the expression for $B(k_x)$ can be found as

$$B(k_x) = \frac{\frac{1}{Z_0}2E_0 e^{-ik_0 d}\delta(k_x) - \frac{1}{\chi_a}I_x(k_x, y=d)}{i\sin(k_{y,b}d)\left[\frac{1}{\chi_a} + \frac{1}{\chi_b}\frac{\chi_s}{\chi_b}\right] - \cos(k_{y,b}d)\left[\frac{1}{\chi_b} + \frac{1}{\chi_a}\frac{\chi_s}{\chi_b}\right]}. \tag{4.42}$$

The electric field in the region $0 < y < d$ can be found through Eq. (4.37) and Eq. (4.8), which gives

$$E_x(k_x, y) = B(k_x)\frac{1}{\chi_b}Z_0\left[i\sin(k_{y,b}y)\frac{\chi_s}{\chi_b} - \cos(k_{y,b}y)\right]. \tag{4.43}$$

Evaluating the field at the $y = d$, the electric field becomes

$$E_x(k_x, y=d) = \frac{\left\{i\sin(k_{y,b}d)\frac{\chi_s}{\chi_b} - \cos(k_{y,b}d)\right\}\left\{2E_0 e^{-ik_0 d}\delta(k_x) - Z_0 I_x(k_x, y=d)\right\}}{\chi_a\chi_b\left\{i\sin(k_{y,b}d)\left[\frac{1}{\chi_a} + \frac{1}{\chi_b}\frac{\chi_s}{\chi_b}\right] - \cos(k_{y,b}d)\left[\frac{1}{\chi_b} + \frac{1}{\chi_a}\frac{\chi_s}{\chi_b}\right]\right\}}, \tag{4.44}$$

where $\delta(k_x)\chi_a = \delta(k_x)$ has been used. This can be simplified by multiplying by $\frac{1}{\sin(k_{y,b}d)}$ in both the numerator and denominator, yielding

$$E_x(k_x, y = d) = \frac{i\frac{\chi_s}{\chi_b} - \cot(k_{y,b}d)}{i\left[\chi_b + \chi_a\frac{\chi_s}{\chi_b}\right] - \cot(k_{y,b}d)\left[\chi_a + \chi_s\right]} \left\{2E_0 e^{-ik_0 d}\delta(k_x) - Z_0 I_x(k_x, y = d)\right\}. \quad (4.45)$$

Now as the field is zero at the gate, assuming it is a perfect conductor, the only unknown will be the current at gate. By determining this current the coefficients for the magnetic and electric field can be found and thus the radiation pattern of the HEMT. To find the current at the gate, a new variable is defined,

$$Z_k(k_x) = \frac{i\frac{\chi_s}{\chi_b} - \cot(k_{y,b}d)}{i\left[\chi_b + \chi_a\frac{\chi_s}{\chi_b}\right] - \cot(k_{y,b}d)\left[\chi_a + \chi_s\right]}. \quad (4.46)$$

The current is given as

$$I_x(k_x, y = d) = \frac{1}{2\pi}\int_{-\frac{w}{2}}^{\frac{w}{2}} I_x(x', y = d)e^{-ik_x x'}\mathrm{d}x'. \quad (4.47)$$

Thus the field at the gate can be found as

$$\begin{aligned} E_x(x, y = d) &= \int_{-\infty}^{\infty} E_x(k_x, y = d)e^{ik_x x}\mathrm{d}k_x \\ &= \int_{-\infty}^{\infty} Z_k(k_x)\left[2E_0 e^{-ik_0 d}\delta(k_x) - Z_0 I_x(k_x, y = d)\right] e^{ik_x x}\mathrm{d}k_x \quad (4.48) \end{aligned}$$

where inserting Eq. (4.47) gives

$$E_x(x, y = d) = 2E_0 e^{-ik_0 d}Z_k(0) - \int_{-\frac{w}{2}}^{\frac{w}{2}}\int_{-\infty}^{\infty} \frac{Z_0}{2\pi}Z_k(k_x)I(x', y = d)e^{ik_x(x-x')}\mathrm{d}k_x\mathrm{d}x'. \quad (4.49)$$

At the gate, $|x| \leq \frac{w}{2}$, the electric field is zero, as the gate is a perfect conductor, yielding

$$2E_0 e^{-ik_0 d}Z_k(0) = \int_{-\frac{w}{2}}^{\frac{w}{2}} Z(x - x')I_x(x', y = d)\mathrm{d}x', \quad (4.50)$$

where

$$Z(x - x') = \int_{-\infty}^{\infty} \frac{Z_0}{2\pi}Z_k(k_x)e^{ik_x(x-x')}\mathrm{d}k_x. \quad (4.51)$$

Now the current can be found using numerical methods.

4.3 Reference Field

Another approach follows a similar procedure to the one in Sec. 4.2, but instead of looking at the total field, the field is split into two parts: a reference field for the system without the scatterer and the field generated from the scatterer, i.e. the gate in this case. The reference structure is seen in Fig. 4.6.

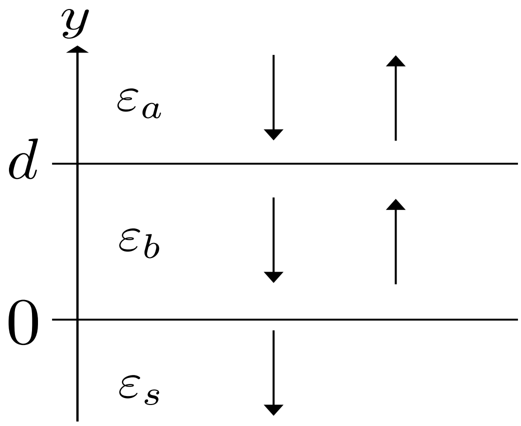

Figure 4.6. Reference structure for how the field would propagate without the scatterer at surface $y = d$.

Assuming a p-polarised wave, ($\mathbf{H} = \hat{\mathbf{z}}H$), The corresponding reference field can be stated as

$$H_{ref}(x,y) = H_0 \left(e^{-ik_{y,a}(y-d)} + r_{as}e^{ik_{y,a}(y-d)}\right) e^{ik_x x} \qquad y > d, \tag{4.52}$$

$$H_{ref}(x,y) = \left(B_{ref}e^{-ik_{y,b}(y-d)} + C_{ref}e^{ik_{y,b}(y-d)}\right) e^{ik_x x} \qquad 0 < y < d, \tag{4.53}$$

$$H_{ref}(x,y) = H_0 t_{as} e^{-ik_{y,s}y} e^{ik_x x} \qquad 0 > y. \tag{4.54}$$

Thus expressions for B_{ref}, C_{ref}, r_{as} and t_{as} are needed.

This can be done by considering the boundary conditions for the two surfaces. The same boundary conditions as in Sec. 4.2 are valid, i.e. the tangential electric field component is continuous and the tangential component of the magnetic field has a jump equal to surface current present. For the reference field this means that the magnetic field is continuous at the interface $y = d$ and the difference is equal to σE_x at $y = 0$, where σ is given by Eq. (4.19).

First the single interface reflection and transmission coefficients for each interface will be derived as these are needed to establish three layer coefficients. Considering a single interface at $y = d$ one can state that the field above the interface consists of a wave propagating downwards and upwards due to reflection at the interface and below the interface the field is what has been transmitted through the surface illustrated in Fig. 4.7.

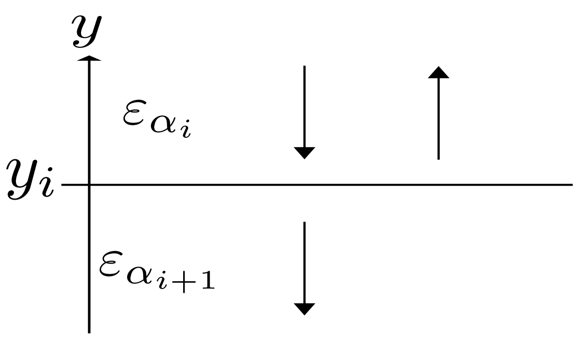

Figure 4.7. Field components for a single layer under consideration with an incident wave from layer α, where $\alpha_i = [a, b, s]$ and $y_i = [0, d]$, for $i = 1, 2$.

The boundary conditions for the surface $y = d$ results in the relations

$$1 + r_{ab} = t_{ab} \tag{4.55}$$

$$\frac{k_{y,a}}{\varepsilon_a}(1 - r_{ab}) = \frac{k_{y,b}}{\varepsilon_b} t_{ab}, \tag{4.56}$$

which gives the expression for the single layer reflection coefficient:

$$r_{ab} = \frac{k_{y,a}\varepsilon_b - k_{y,b}\varepsilon_a}{k_{y,a}\varepsilon_b + k_{y,b}\varepsilon_a}, \tag{4.57}$$

and the transmission coefficient can be obtained from Eq. (4.55). A similar consideration can be done for the surface $y = 0$, which gives the relations

$$H_0(1 + r_{bs} - t_{bs}) = \sigma E_x \tag{4.58}$$

$$\frac{k_{y,b}}{\varepsilon_b}(1 - r_{bs}) = \frac{k_{y,s}}{\varepsilon_s} t_{bs}, \tag{4.59}$$

for the tangential magnetic and electric field boundary condition respectively, where E_x is given by Eq. (4.8). Again searching for an expression for the single interface reflection coefficient leads to

$$r_{bs} = \frac{k_{y,b}\varepsilon_s(1 + \sigma\frac{k_{y,s}}{\omega\varepsilon_0\varepsilon_s}) - k_{y,s}\varepsilon_b}{k_{y,b}\varepsilon_s(1 + \sigma\frac{k_{y,s}}{\omega\varepsilon_0\varepsilon_s}) + k_{y,s}\varepsilon_b}, \tag{4.60}$$

and a transmission which can be obtained from

$$t_{bs} = (1 - r_{bs})\frac{k_{y,b}\varepsilon_s}{\varepsilon_b k_{y,s}}. \tag{4.61}$$

The reflection and transmission coefficient for the whole structure, r_{as} and t_{as}, can be found by considering propagation throughout the entire structure. For the reflection coefficient the path can be described as

$$r_{as}(k_x) = r_{ab} + t_{ab}e^{ik_{y,b}d}r_{bs}e^{ik_{y,b}d}t_{ba}\sum_{n=0}^{\infty}\left(r_{ba}e^{2ik_{y,b}d}r_{bs}\right)^n, \tag{4.62}$$

The first term is the reflection at the first interface. The following term for $n = 0$ the wave is firstly transmitted through the first interface, propagates down through the second layer to the second interface where it is reflected back up to the first interface and gets transmitted into the first layer. For $n > 0$, the situation is the same but the wave does n extra turns back and forth in the second layer before being transmitted back through the upper interface and into the first layer. The sum in Eq. (4.62) can be reduced by noting that it is in fact a geometric series, which converges to

$$\sum_{n=0}^{\infty} ax^n = \frac{a}{1-x}, \quad \text{for } |x| < 1. \tag{4.63}$$

Eq. (4.62) can therefore be simplified to

$$r_{as} = \frac{r_{ab} + r_{bs}e^{2ik_{y,b}d}}{1 + r_{ab}r_{bs}e^{2ik_{y,b}d}}, \tag{4.64}$$

where $r_{ba} = -r_{ab}$ and $t_{ab}t_{ba} = 1 - r_{ab}^2$ has been used. The three layer transmission coefficient can be obtained by a similar line of thought:

$$t_{as} = \frac{t_{ab}t_{bs}e^{ik_{y,b}d}}{1 + r_{ab}r_{bs}e^{2ik_{y,b}d}}. \tag{4.65}$$

The coefficients B_{ref} and C_{ref} can be obtained by considering the fields of Eq. (4.54) at the interface $y = d$. Here, the magnetic and electric field boundary conditions yield

$$H_0(1 + r_{as}) = B_{ref} + C_{ref}, \tag{4.66}$$

$$\frac{k_{y,a}\varepsilon_b}{\varepsilon_a k_{y,b}} H_0(1 - r_{as}) = B_{ref} - C_{ref}. \tag{4.67}$$

By respectively adding and subtracting these equations one obtains the following expressions for the coefficients:

$$B_{ref} = \frac{1}{2}H_0\left((1 + r_{as}) + (1 - r_{as})\frac{k_{y,a}\varepsilon_b}{\varepsilon_a k_{y,b}}\right), \tag{4.68}$$

$$C_{ref} = \frac{1}{2}H_0\left((1 + r_{as}) - (1 - r_{as})\frac{k_{y,a}\varepsilon_b}{\varepsilon_a k_{y,b}}\right). \tag{4.69}$$

The reference field can now be calculated, which leaves the contribution from the scattered field, illustrated in Fig. 4.8.

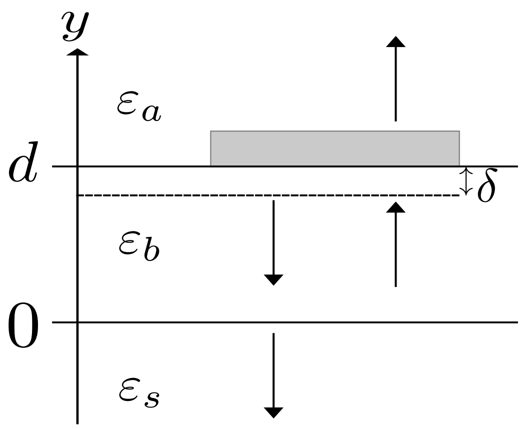

Figure 4.8. Field emitted from the scatterer placed on the interface $y = d$.

As in Sec. 4.2 the electromagnetic fields can be in expressed through the Fourier transform as follows

$$H_{sc}(x, y) = \int A_{sc}(k_x)e^{ik_{y,a}(y-d)}e^{ik_x x}\mathrm{d}k_x \qquad y > d, \tag{4.70}$$

$$H_{sc}(x, y) = \int \left(B_{sc}(k_x)e^{-ik_{y,b}(y-d)} + C_{sc}(k_x)e^{ik_{y,b}(y-d)}\right)e^{ik_x x}\mathrm{d}k_x \qquad 0 < y < d, \tag{4.71}$$

$$H_{sc}(x, y) = \int D_{sc}(k_x)e^{-ik_{y,s}y}e^{ik_x x}\mathrm{d}k_x \qquad y < 0. \tag{4.72}$$

Likewise the current in the gate can be represented by its Fourier transform:

$$I_x(x, y = d) = \int I_x(k_x, y = d)e^{ik_x x}\mathrm{d}k_x. \tag{4.73}$$

Similarly to the reference field the coefficients for the scattered field must be found from the boundary conditions. Unlike before, for the reference structure, the current in the gate is also an unknown, which must be determined, which will be done in a similar vein to what was done in Sec. 4.2. From the boundary conditions at the gate, $y = d$, the following relations must be satisfied

$$A_{sc} - (B_{sc} + C_{sc}) = I_x, \tag{4.74}$$

$$\frac{k_{y,a}}{\varepsilon_a} A_{sc} = -\frac{k_{y,b}}{\varepsilon_b}(B_{sc} - C_{sc}). \tag{4.75}$$

In the second layer the upward propagating field can be related to the downward propagating field, by realising that the only contribution to the upward propagating field is from what has been reflected field at the interface at $y = 0$, hence the fields are related as

$$r_{bs} B_{sc} e^{ik_{y,b}d} = C_{sc} e^{-ik_{y,b}d} \Rightarrow C_{sc} = r_{bs} B_{sc} e^{2ik_{y,b}d}. \tag{4.76}$$

Using this expression for C_{sc} in Eq. (4.75) two equations for A_{sc} can be obtained as

$$A_{sc} = B_{sc}(1 + r_{bs} e^{2ik_{y,b}d}) + I_x, \tag{4.77}$$

$$A_{sc} = -B_{sc}(1 - r_{bs} e^{2ik_{y,b}d}) \frac{k_{y,b}\varepsilon_a}{\varepsilon_b k_{y,a}}, \tag{4.78}$$

which can be combined to find an expression for B_{sc}

$$\begin{aligned} B_{sc}(1 + r_{bs} e^{2ik_{y,b}d}) + I_x &= -B_{sc}(1 - r_{bs} e^{2ik_{y,b}d}) \frac{k_{y,b}\varepsilon_a}{k_{y,a}\varepsilon_b} \\ \Leftrightarrow B_{sc} &= -\frac{I_x}{(1 + r_{bs} e^{2ik_{y,b}d}) + (1 - r_{bs} e^{2ik_{y,b}d}) \frac{k_{y,b}\varepsilon_a}{k_{y,a}\varepsilon_b}}. \end{aligned} \tag{4.79}$$

All the coefficients can now be determined provided I_x is known. In order to determine I_x, the tangential component of the electric field is evaluated a small distance, δ, below the gate, yielding

$$\begin{aligned} E_{x,sc}(k_x, y = d - \delta) &= \frac{k_{y,b}}{\omega\varepsilon_0\varepsilon_b} B_{sc} \left(e^{ik_{y,b}\delta} - r_{bs} e^{ik_{y,b}(2d-\delta)} \right) \\ &= -I_x \frac{\frac{k_{y,b}}{\omega\varepsilon_0\varepsilon_b} \left(e^{ik_{y,b}\delta} - r_{bs} e^{ik_{y,b}(2d-\delta)} \right)}{(1 + r_{bs} e^{2ik_{y,b}d}) + (1 - r_{bs} e^{2ik_{y,b}d}) \frac{k_{y,b}\varepsilon_a}{k_{y,a}\varepsilon_b}} \\ &= -I_x(k_x) Z(k_x) \end{aligned} \tag{4.80}$$

where Eq. (4.76) has been used for the first equality and

$$Z(k_x) = \frac{\frac{k_{y,b}}{\omega\varepsilon_0\varepsilon_b} \left(e^{ik_{y,b}\delta} - r_{bs} e^{ik_{y,b}(2d-\delta)} \right)}{(1 + r_{bs} e^{2ik_{y,b}d}) + (1 - r_{bs} e^{2ik_{y,b}d}) \frac{k_{y,b}\varepsilon_a}{k_{y,a}\varepsilon_b}}. \tag{4.81}$$

Using that the gate is a perfect conductor, then for $|x| < \frac{w}{2}$ the tangential component of the field slightly below the gate can be approximated to be zero, i.e.

$$E_{x,ref}(x, y = d - \delta) + E_{x,sc}(x, y = d - \delta) \simeq 0. \tag{4.82}$$

Therefore by transforming Eq. (4.80) to real space one can write

$$E_{x,ref}(x, y = d - \delta) - \int I_x(k_x) Z(k_x) e^{ik_x x} \mathrm{d}k_x = 0, \tag{4.83}$$

the Fourier transform of the current is written as the inverse Fourier transform, where it should be noted that the integral limits can be restricted to run from $\pm\frac{w}{2}$, as the current is restricted to the gate:

$$E_{x,ref}(x, y = d - \delta) = \frac{1}{2\pi} \int_{-\frac{w}{2}}^{\frac{w}{2}} \int_{-\infty}^{\infty} I_x(x') Z(k_x) e^{ik_x(x-x')} \mathrm{d}k_x \mathrm{d}x' \tag{4.84}$$

$$= \frac{1}{2\pi} \int_{-\frac{w}{2}}^{\frac{w}{2}} I_x(x') Z(x - x') \mathrm{d}x', \tag{4.85}$$

where

$$Z(x - x') = \int_{-\infty}^{\infty} Z(k_x) e^{ik_x(x-x')} \mathrm{d}k_x. \tag{4.86}$$

Analysis and Discussion 5

In the following analysis an AlGaN/GaN HEMT structure will be used for calculations gate current calculations following the methods described in Sec. 4.2 and Sec. 4.3, which will be referred to as RF (Reference Field) and PPS (Popov, Polischuk, Shur). The general structure parameters required for the analysis is first of all the dielectric constants of AlGaN and GaN respectively in the THz range. The dielectric constant for GaN has been acquired from references [35][36] and is set to 5.3. For the insulator layer of AlGaN the dielectric has been set to 9.5 based on references [37][38]. Parameters such as frequency, f, electron concentration, N_c, momentum relaxation time, τ, gate length, w, and distance from the gate to the 2DEG, d, will in general be variable parameters and their influence will be analysed. Furthermore the amplitude of the incident field is set to $H_0 = 1$. However, for the initial analysis, which is the calculation of the current, these parameters will be constant, with the exception of gate length and frequency in order to determine how convergence is affected by these parameters. The values used for these parameters are $f = 1.44$ Thz, $N_c = 10^{17}$ m^{-2} [19], $\tau = 2.27 \times 10^{-13}$ s [38], $w = 400$ nm and $d = 10$ nm when the 2DEG layer is infinitely thin and $d = 6$ nm and $d_{2DEG} = 4$ nm when it has a finite thickness, furthermore the value for the effective mass is taken as $m = 0.22m_0$ from Ref. [33][19].

5.1 Mode Indices

Based on the theory established in Sec. 4.1 and App. A.5, mode indices for different HEMT structures can be found and analysed. In Sec. 4.1 the mode index for the gated structure are found. First the model with a finite thickness of the 2DEG will be analysed. The mode index depends on the parameters of the setup, the dependency on the frequency for an ungated structure is shown in Fig. 5.1.

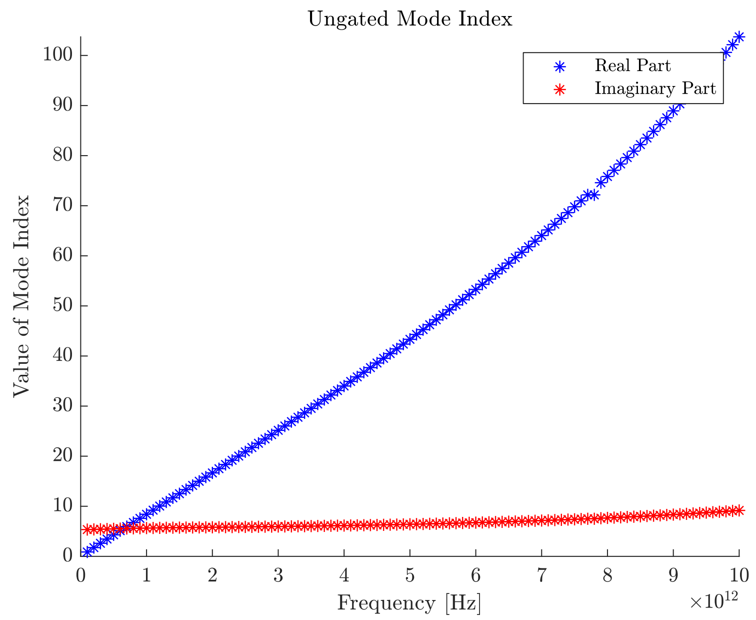

Figure 5.1. The mode index for a ungated structure, versus different frequencies.

The imaginary part of the mode index is almost constant, while the real part increases almost linearly as the frequency increases. The frequencies tested are from 0.1 THz to 10 THz. For the gated structure the dependency on frequency is shown in Fig. 5.2.

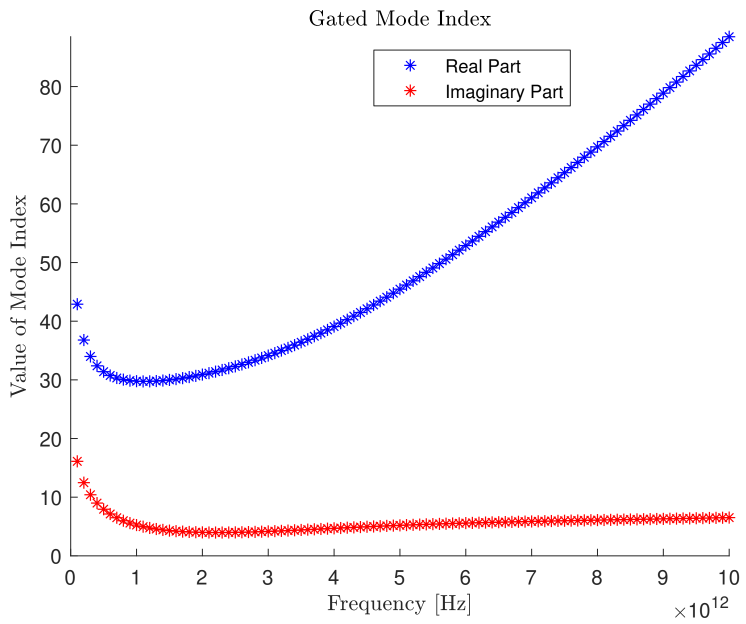

Figure 5.2. The mode index for a gated structure, versus different frequencies.

Compared to the ungated structure, the gated structure has the same linear tendency for higher frequencies, however for low frequencies, lower than 1 THz, the mode index increases.

By varying the distance from the gate to the electron gas, the mode index also varies. This has been illustrated in Fig. 5.3

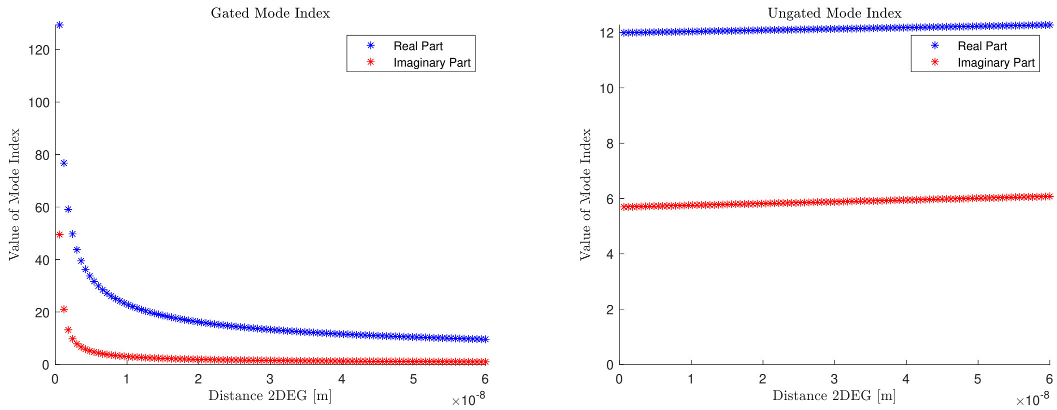

Figure 5.3. The mode index for both the gated and ungated structure, versus different distances to the electron gas from the gate.

The ungated mode index does not change significantly with changing distance. As the distance to the 2DEG is relatively small compared to the wavelength, for the ungated structure, the difference in nm scale, the effective material above the 2DEG is largely unaffected, which could explain the irrelevance of the distance to the 2DEG. However for the gated structure, the mode index rises very high for low distances, which could be explained by the gate drastically changing the material above the 2DEG, as it does not have an infinite amount of ambient material above.

When changing the thickness of the 2DEG, the mode index also changes. For the models in Secs. 4.2 and 4.3, the thickness of the 2DEG has been assumed to be zero. The mode index as the thickness of the gas goes to zero, should show which plasma modes can be expected in the models. In Fig. 5.4 the dependency of thickness is illustrated.

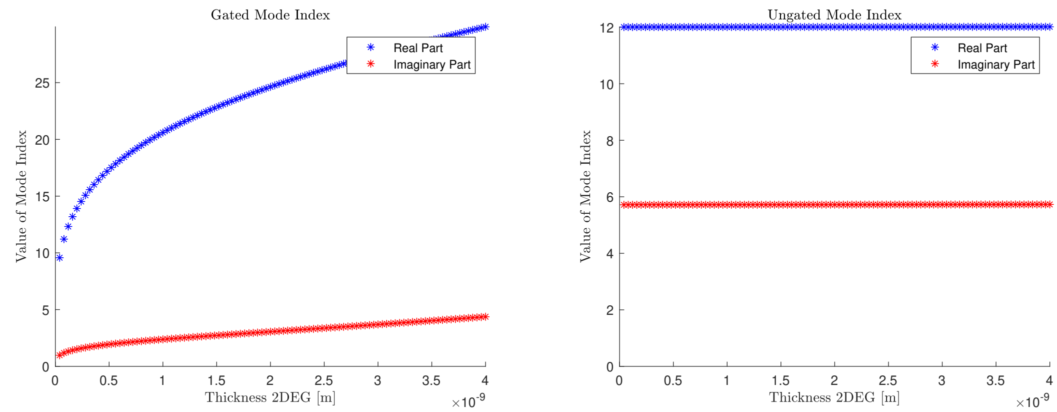

Figure 5.4. The mode index for both the gated and ungated structure, versus different thicknesses of the electron gas.

It can be seen that as the 2DEG becomes thinner, the mode index decreases for the gated structure. The models where the 2DEG was considered to be infinitely thin, the distance to the 2DEG is set to be 10 nm, instead of the 6 nm used in the standard calculations for the mode index. To account for this, the mode index have been found for even thinner electron gas, with a distance from the gate of 10 nm, which is shown in Fig. 5.5.

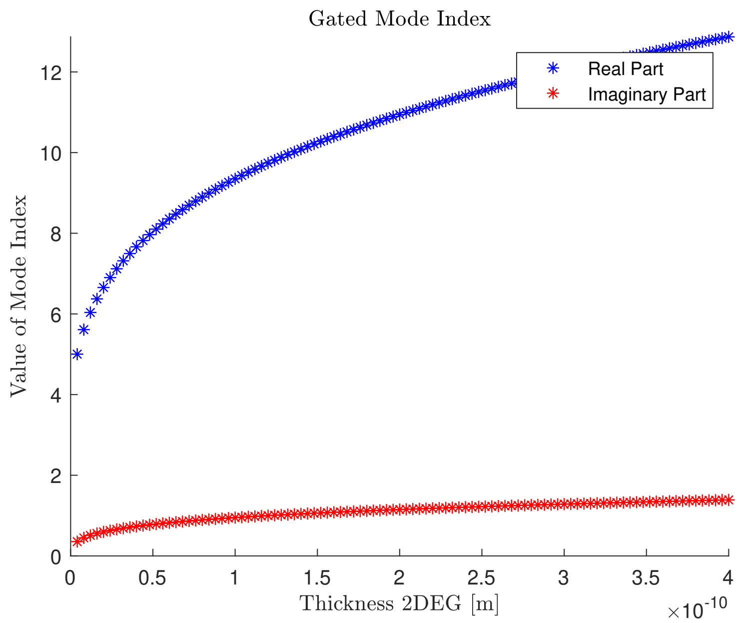

Figure 5.5. The mode index for both the gated structure, versus different thicknesses the electron gas. The distance from the gate to the 2DEG is 10 nm.

The mode index tends to zero as the thickness of the electron gas decreases. This disagrees with the method for the infinitely thin 2DEG. For the infinitely thin 2DEG the mode index was found to be $n_m = 255.07 + 60.21i$ in Sec. 4.1, however, that method incorporated the surface currents of the 2DEG, which might explain the difference in the supported modes as the thickness goes to zero. For the mode index found through the infinitely thin model, the optimal gate length can be calculated, assuming the gate operates as a dipole antenna, where the optimal length is

$$\left(\frac{1}{2} + m\right)\lambda \approx w, \tag{5.1}$$

where w is the gate length [39]. Using the relation between the plasma wavelength and the incident wavelength

$$\lambda_p = \frac{\lambda}{n_m}, \tag{5.2}$$

the fundamental peak will be for a gate length of

$$w \approx \frac{\lambda}{2n_m}. \tag{5.3}$$

For an incoming wave with a frequency of 1.44 THz, the optimal gate length becomes

$$w \approx \frac{c}{2n_m \times 1.44\,\mathrm{THz}} \approx 408\,\mathrm{nm}. \tag{5.4}$$

Thus a gate length of around 400 nm should be have a resonant response to the plasma waves, if the 2DEG is assumed to be infinitely thin. The mode index for an infinitely thin 2DEG for different frequencies is shown in Fig. 5.6.

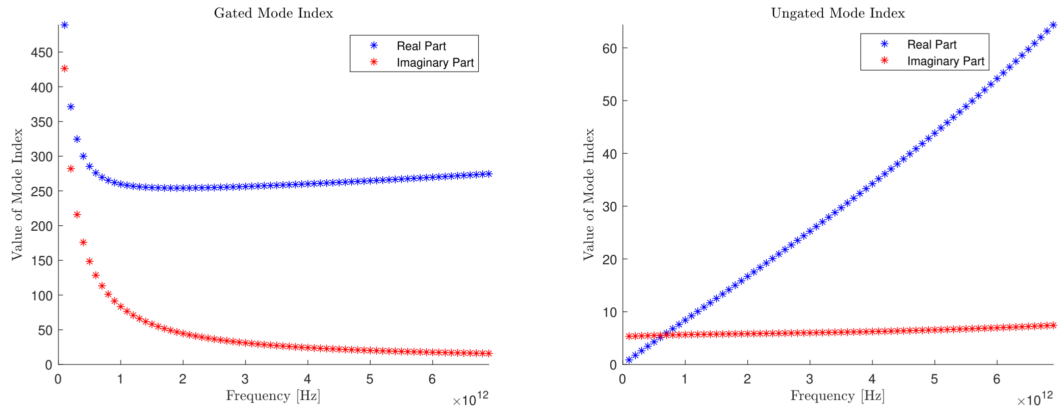

Figure 5.6. The mode index for both the gated infinitely thin 2DEG structure, versus different frequencies. The distance from the gate to the 2DEG is 10 nm.

The real part of the mode index for the gated structure seems to have a minimum, which would correspond to the longest plasma modes the structure can support, for a frequency of around 1.9 THz, where the mode index is $n_m \approx 250.01+46.83i$. For the gated structure, the mode increases rapidly for frequencies lower than 1 THz, both for the real and imaginary part. The real part of the mode index for the ungated structure seems to increase almost linearly as was seen for the model with a finite thickness. The mode index for a 10 THz radiation is $n_m = 294.65 + 12.89i$. For this wavelength the optimal gate length, again assuming that dipole antenna theory holds, would be $w = 50.91$ nm. The amount of electrons in the 2DEG can be affected by applying a potential on the gate as mentioned in Sec. 3.1.3, thus the mode index for varying charge carrier concentrations are shown in Fig. 5.7.

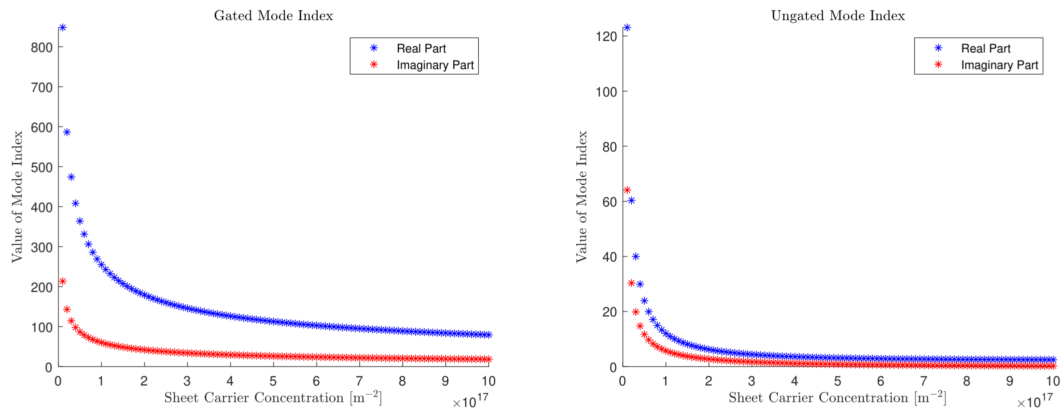

Figure 5.7. The mode index for both the gated infinitely thin 2DEG structure, versus different sheet carrier concentrations. The distance from the gate to the 2DEG is 10 nm.

The mode index seems to rise as N_c decreases. This was not as expected from Eq. (3.75), where ω_p^2 should be proportional to N_c, which would suggest $n_m \propto \sqrt{N_c}$.

The distance from the 2DEG to the gate, can also have an effect on the possible plasma modes. Thus the mode indices for an interval of 2DEG depths are shown in Fig. 5.8.

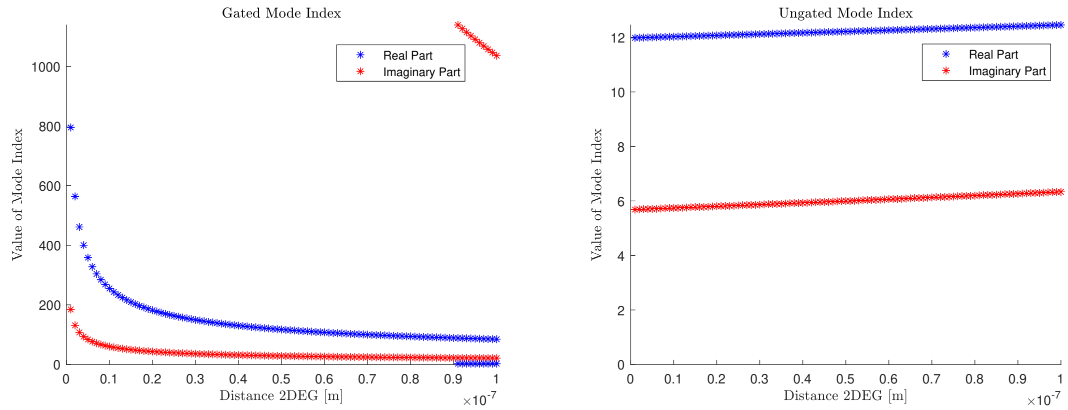

Figure 5.8. The mode index for both the gated infinitely thin 2DEG structure, versus different 2DEG depths. The distance from the gate to the 2DEG varies and the rest of parameters are the same as in Fig. A.4.

For the gated structure, the mode index increases as the distance decreases. For large distances, another mode appears, with a low real part and a high imaginary part. For the ungated structure, the mode index increases linearly.

5.2 The PPS Method

In order to model the electromagnetic radiation, as described in Sec. 4.2, the current in the gate is described through point matching. The gate is split into N sections, each with length Δ and the current at each section is assumed to be constant. Thus at point i the current needs to satisfy,

$$E_{0,i} = \sum_{j=1}^{N} Z(x_i - x_j) I_j \Delta, \tag{5.5}$$

where both x_i and x_j are points along the gate and $E_{0,i} = 2E_0 e^{ik_0 d} Z_k(0)$. This can be written as a matrix equation,

$$\mathbf{E}_0 = \mathbf{ZI}, \tag{5.6}$$

where the elements of $\mathbf{Z}$ can be found using Eq. (4.51). However the $Z(x - x')$ does not converge, as can be seen in Fig. 5.9 as the linear tendency continues when increasing $|k_x|$.

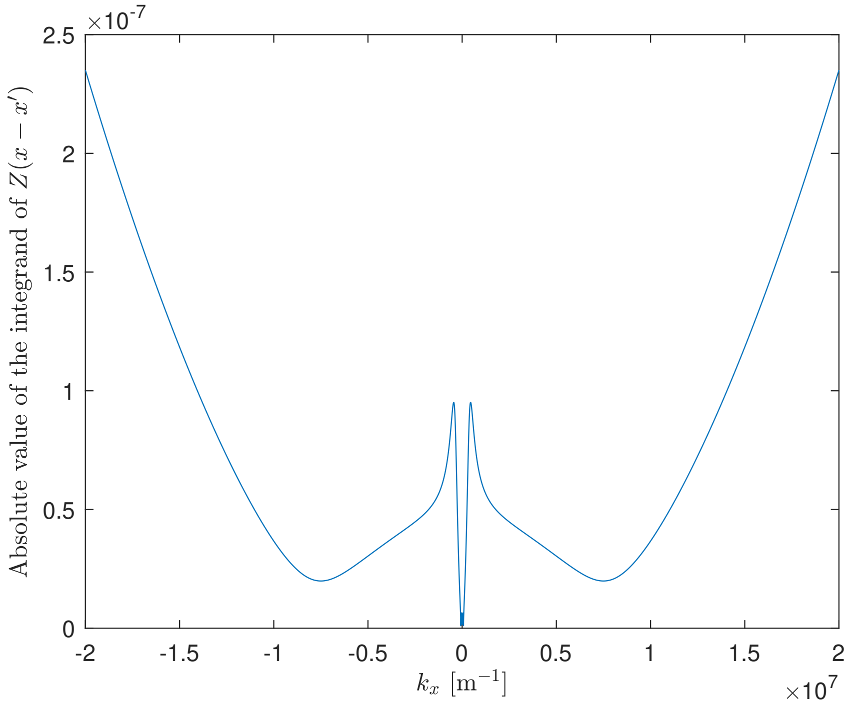

Figure 5.9. Plot of the integrand of $Z(x-x')$ versus k_x.

A possible solution to this is letting the electric field be zero a small distance, δ, beneath the gate, so that $E_x(|x| \leq \frac{w}{2}, d-\delta) = 0$, where δ will depend on the division of the gate. The electric field can be found from Eq. (4.43) which leads to

$$E_x(k_x, y = d-\delta) = \frac{i\sin(k_{y,b}(d-\delta))\frac{\chi_s}{\chi_b} - \cos(k_{y,b}(d-\delta))}{i\sin(k_{y,b}d)\left[\chi_b + \chi_a\frac{\chi_s}{\chi_b}\right] - \cos(k_{y,b}d)\left[\chi_a + \chi_s\right]} \left[2E_0 e^{-ik_0 d}\delta(k_x) - Z_0 I_x(k_x, y=d)\right], \quad (5.7)$$

Using the same approach as in Sec. 4.2, the field at the gate gives

$$2E_0 e^{ik_0 d} Z_{\delta,k}(0) = \int_{-\frac{w}{2}}^{\frac{w}{2}} Z_\delta(x-x') I(x', y=d)\mathrm{d}x', \quad (5.8)$$

where

$$Z_{\delta,k}(k_x) = \frac{i\sin(k_{y,b}(d-\delta))\frac{\chi_s}{\chi_b} - \cos(k_{y,b}(d-\delta))}{i\sin(k_{y,b}d)\left[\chi_b + \chi_a\frac{\chi_s}{\chi_b}\right] - \cos(k_{y,b}d)\left[\chi_a + \chi_s\right]}, \quad (5.9)$$

and

$$Z_\delta(x-x') = \int_{-\infty}^{\infty} \frac{Z_0}{2\pi} Z_{\delta,k}(k_x) e^{ik_x(x-x')}\mathrm{d}k_x. \quad (5.10)$$

The integral of Z_δ does converge, the integrand is illustrated in Fig. 5.10, where it converges to zero for increasing $|k_x|$. The integral should be calculated for all for all $x_n - x'_m$, where both n and m runs over N. However, due to the translational symmetry of Eq. (5.10), only the N integrals for the first measuring point and the N integrals for the last measuring point has to be calculated, as this covers all the possible exponents in the integrand. The remaining integrals can simply be extracted from these $2N$ integrals.

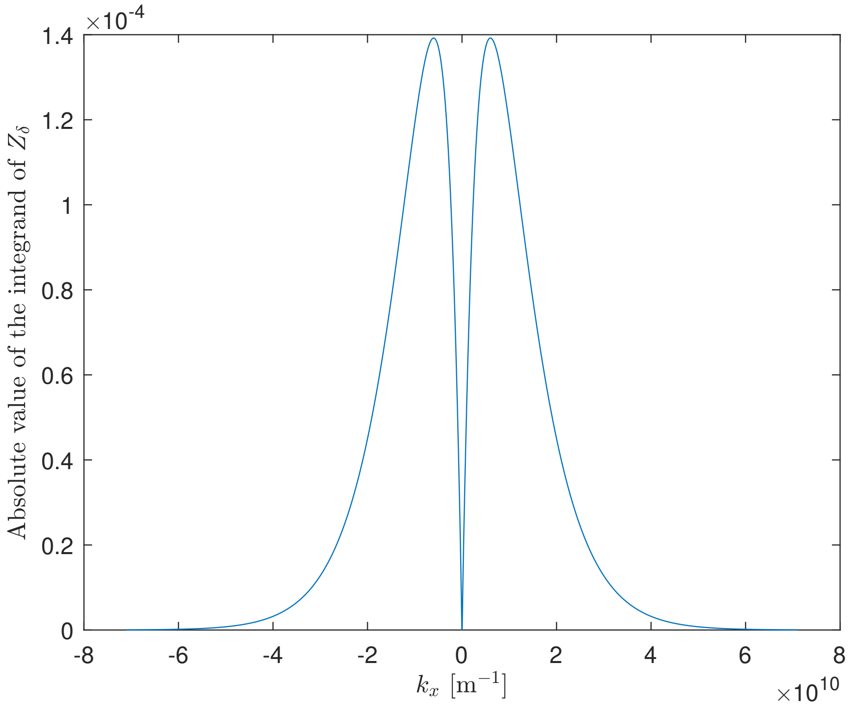

Figure 5.10. Plot of the integrand of $Z_\delta(x - x')$ versus k_x.

However, letting k_x go towards infinity results in numerical problems for MATLAB, as both the numerator and denominator tends to infinity. The denominator does go to infinity quicker than the numerator and as such the integrand should tend towards 0. MATLAB, however, seems to evaluate the numerator and denominator separately (at least partially) and for very large values of k_x the integrand becomes ∞/∞, which is indeterminate and the integral becomes undefined. The integral of Eq. (5.10) is therefore done over a limited amount of k_x from which the integrand has become vanishing small but are within MATLABs numerical limits, these limits are denoted from $-k_{x,lim}$ to $k_{x,lim}$. There is in fact an upper limit on how high k_xs needs to be included in the integral. This is due to MATLABs floating-point number limit, being about 10^{308}, and that $1/x$ in the limit $x \lim \infty \rightarrow 0$. When the denominator of Eq. (5.10) reaches the limit it simply sets the value to zero even though the fraction itself might not be close to zero. This causes a discontinuity, and if the function is not sufficiently close to zero at this point problems can occur when integrating. The discontinuity can be seen in Fig. 5.11 for $N = 700$, where the real part of the integrand of Eq. (5.10) is taken as an example.

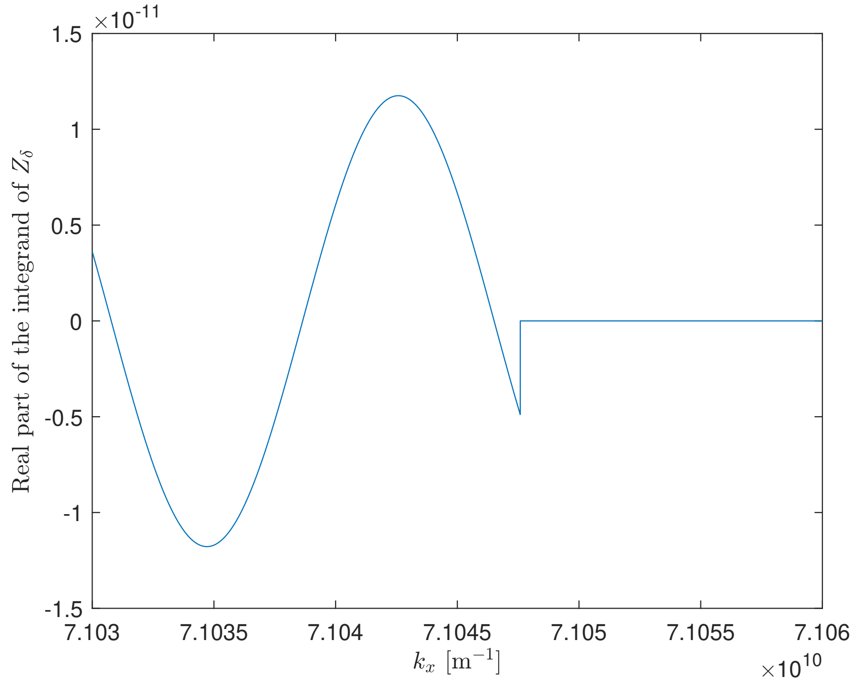

Figure 5.11. Plot of the real part of the integrand of $Z_\delta(x - x')$ versus k_x showing the discontinuity due to MATLABs floating-point number limit for $N = 700$.

Since the denominator is independent of the number of divisions, N, in the gate this limit is constant for any number and is ca. $|k_x| \sim 7.10476 \times 10^{10}$. This is not a problem for low numbers of N as the function is practically zero before reaching this value, but it does become increasingly relevant when increasing number of gate divisions. For high division of the gate, $N \geq 1300$, another problem occurs due to the size of δ. As N increases δ decreases and the numerator of Eq. (5.10) becomes less suppressed and increases more rapidly. At about $N \geq 1300$ it reaches MATLABs numerical limit before $|k_x| \sim 7.10476 \times 10^{10}$, where the denominator is set equal to ∞ and the function is set to ∞ and cannot be integrated. The effect of this is shown in Fig. 5.12 for $N = 3000$, where one can observe that the x-axis simply ends at about $|k_x| = 7.10476 \times 10^{10}$. The further effect of δ will be analysed further at the end of this section.

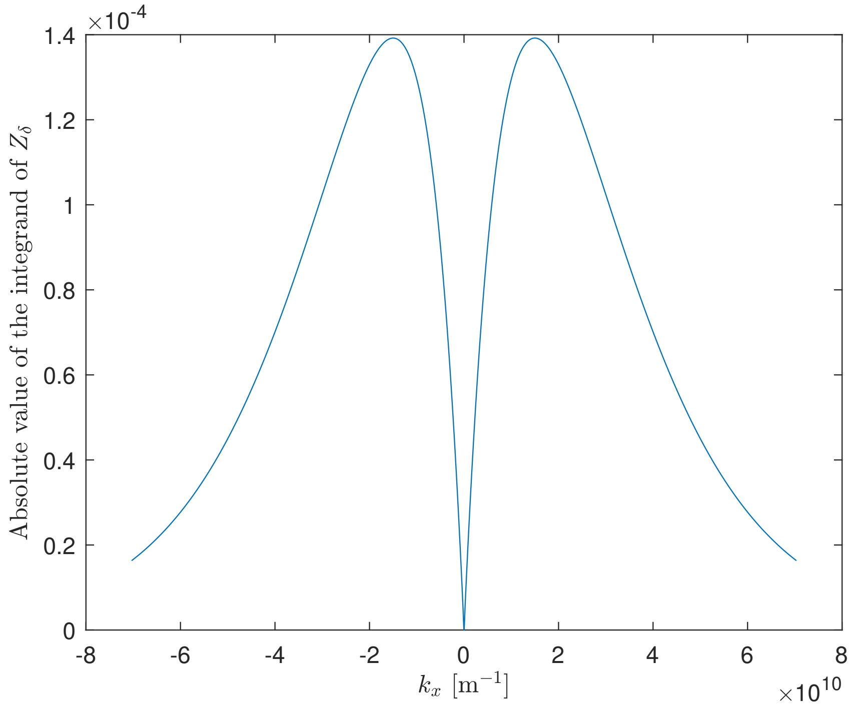

Figure 5.12. Plot of the integrand of $Z_\delta(x - x')$ versus k_x, with the number of gate divisions being $N = 3000$.

Following the preceding analysis the integration is limited to $|k_x| = 7.10476 \times 10^{10}$ and $N = 1200$. When the integral is computed the matrix equation, Eq. (5.6), can be used to find the current at the gate.

If the current is converged, its value should go towards a constant with increasing number of gate divisions and if the change in between increasing number of segmentation becomes acceptably small, while the amount of segments is small enough to allow for the integrand to go to zero, the result is usable.

5.2.1 Calculating the current

The current has been calculated for an increasing number of gate divisions, from $N = 100$ to $N = 1200$ in steps of 100. In Fig. 5.13 the current for N =100, 200, 500, 800, 1000, 1200 has been plotted.

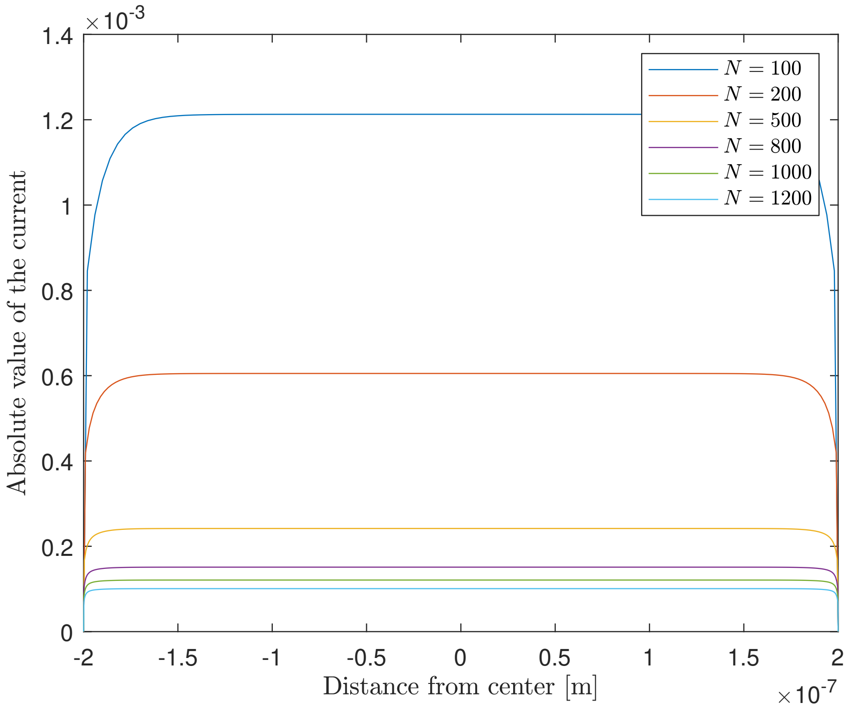

Figure 5.13. Convergence plot for the absolute value of the current from Eq. (5.6) for N =100, 200, 500, 800, 1000, 1200.

As the number of gate divisions increase, the current across the gate drops. The max value of the current drops to about half the value when the amount of segments is doubled. At about $N = 1000$ this tendency becomes less obvious as the function seem to start overlapping due to the scale of the axis, however focusing on the higher amounts of segments, the same tendency can be observed, which is illustrated in Fig. 5.14.

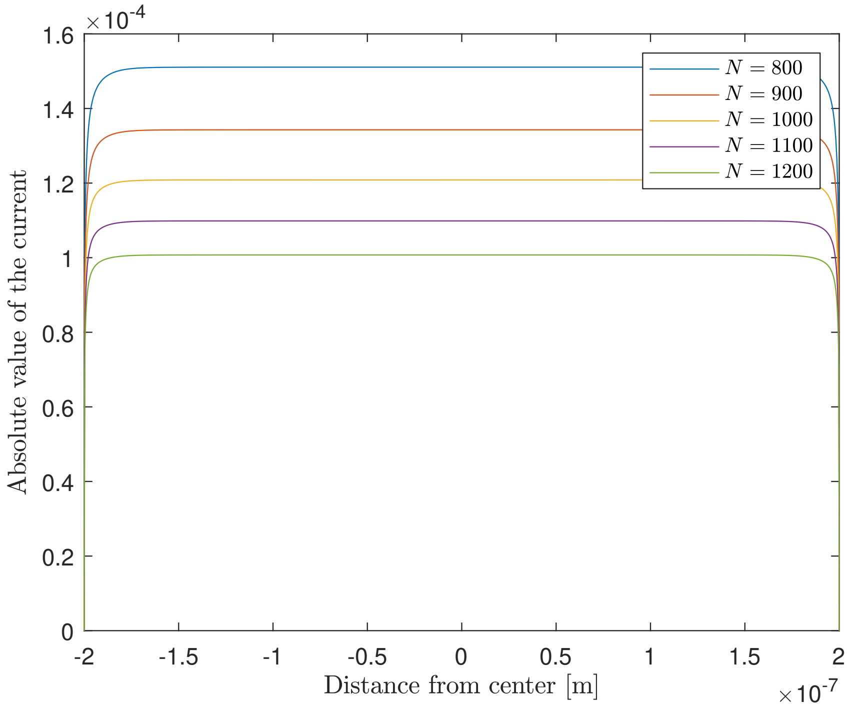

Figure 5.14. Convergence plot for the absolute value of the current from Eq. (5.6) for N =800, 900, 1000, 1100, 1200.

The pattern with a halving in current when doubling in N also seems to continue if the calculated values are compared. If this tendency is consistent for all values of N, one should be able to observe a strong correlation between the value of the current and $1/N$, this has been plotted in Fig. 5.15.

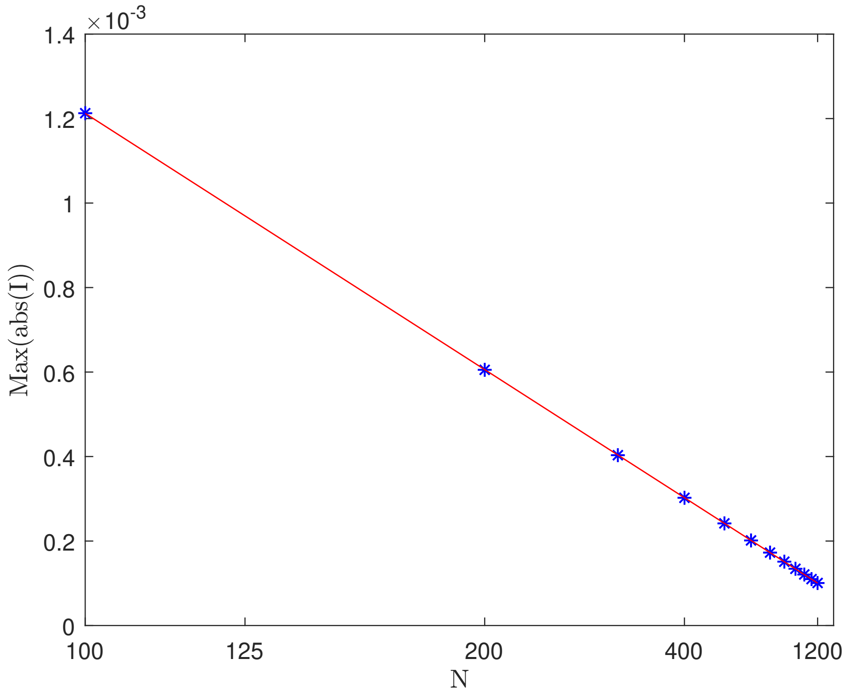

Figure 5.15. The blue 'stars' are the calculated maximum value of the absolute value of the current against $-1/N$. The red line is the regression line obtained from the calculated values. Plotting against $-1/N$ is done in order to have increasing N in the positive x direction. The points on the x-axis is given in values of N.

The plot shows a very strong linear relation. Therefore, by making a linear regression of the function $y = az + b$, where y is Max(abs(I)) and $z = 1/N$, the intersection with the y-axis can be found, which corresponds to the limit for $N \to \infty$, under the assumption that the function is valid outside the calculated interval.

Performing the linear regression one obtains

$$y = z \times (-0.121293436) - 0.000000651, \tag{5.11}$$

which practically says that the absolute value of the current converges to a negative number. As this cannot be the case the relation cannot be valid outside the observed interval. This indicates that this method is poor at describing the current in the gate.

One option that is worth exploring is the dependency of δ, the small displacement in y, and Δ the length of the segmentation. So far δ has been calculated as

$$\delta = \frac{\Delta}{2} = \frac{w}{2N}, \tag{5.12}$$

where w is the width of the gate. The reasoning behind this specific distance is showcased in Fig. 5.16, where the black dots represents the measuring points. By moving the measuring points further away from the gate will make neighbouring sections and its own contributions more equal, as the distance between its own section and its neighbours becomes comparable. Therefore the value of δ should at most be on the order of the segmentation length.

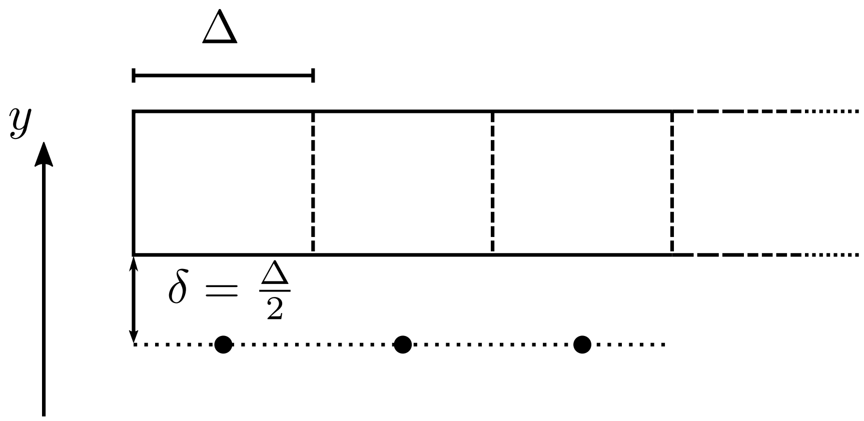

Figure 5.16. Schematic illustration of the displacement of measuring points from the gate strip.

However, the function will obviously also become a problem for increasing N as the measuring points moves closer to the gate and integrand shown in Fig. 5.9 is restored. The value for N for which the displacement becomes too small is ca. $N = 1300$ as was found earlier. It is clear that having a dependency as in Eq. 5.12 cannot be used for the method in the general case. So in an attempt to improve the convergence different constant values of δ has been used in calculating the current. It should be noted that δ also has a maximum and minimum value limited by the distance to the 2DEG, as placing the measuring point for the gate in the 2DEG is nonsensical, and placing the measuring points closer than 1.5385×10^{-10} m, which corresponds to $N = 1300$, as δ becomes too small, which causes numerical instability. If δ is a set value, N is no longer restricted in value. Results for $\delta = w/300$, $w/1000$, $w/1800$, $w/2400$ are shown in Fig. 5.17.

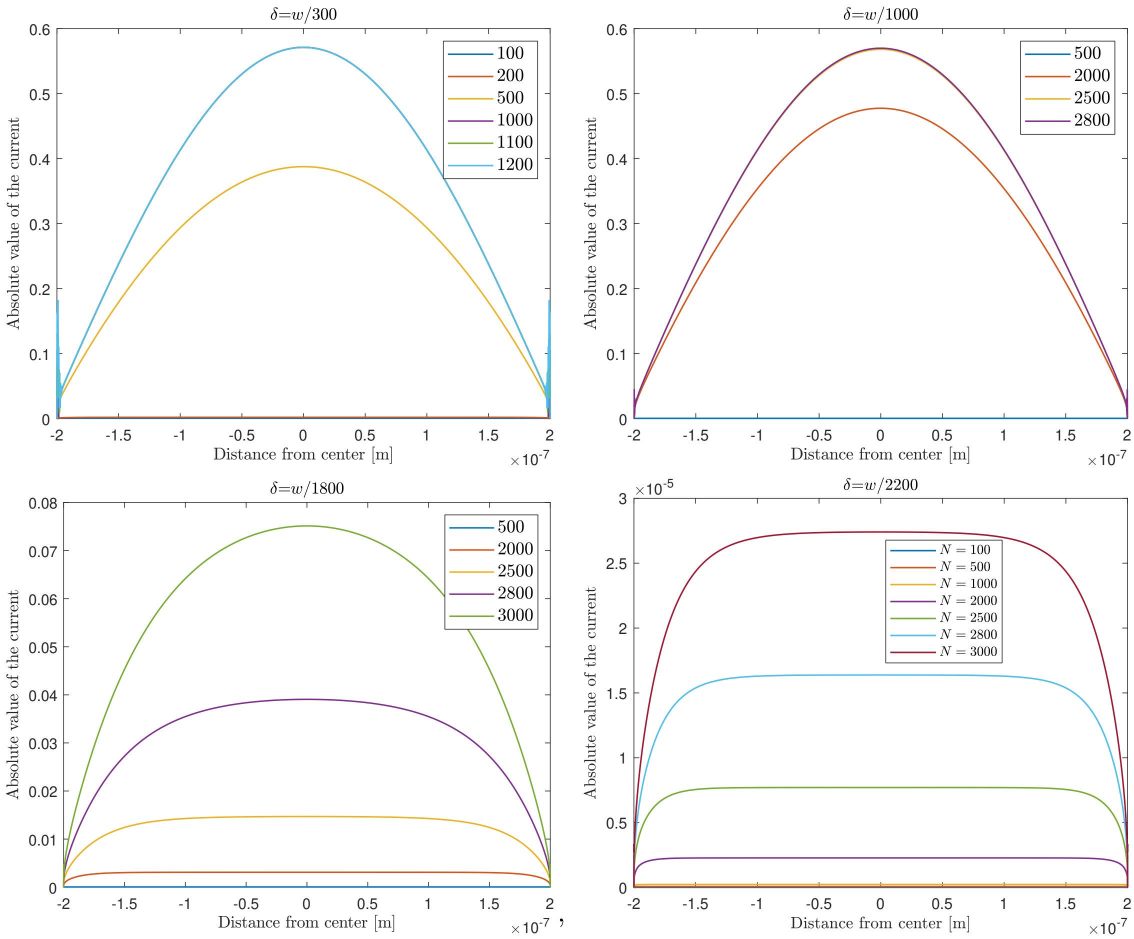

Figure 5.17. Current plots for constant value of displacement δ where the legends indicate the number of segmentation, for values $\delta = w/300$, $w/1000$, $w/1800$, $w/2200$.

It would seem that for $\delta = w/1000$ the current converges. For a gate length of 400 nm this corresponds to a distance of 0.4 nm. Decreasing the distance further restores the more square current profile seen in Fig. 5.17 for $\delta = w/1800$ and $w/2200$ and in Fig. 5.14 for the variable δ. Increasing the distance to about a third of the distance between the gate and 2DEG, seen in the top left of Fig. 5.17, the corners are poorly described, which can be due to the concern expressed in connection with Fig. 5.16, i.e. the contribution from nearby compared to far away points becomes comparable. Increasing the number of segments for this distance simply breaks the model, which might be due to an increasingly poor description of the neighbouring contributions as the segments becomes smaller.

The corners for $\delta = w/1000$ also have some problems and again becomes worse as the segmentation increases. By the previous reasoning one should able to improve this by moving the measuring points slightly closer to the gate. In Fig. 5.18 different values of δ has been used for the same number of segments, $N = 3000$, in order to examine this assertion.

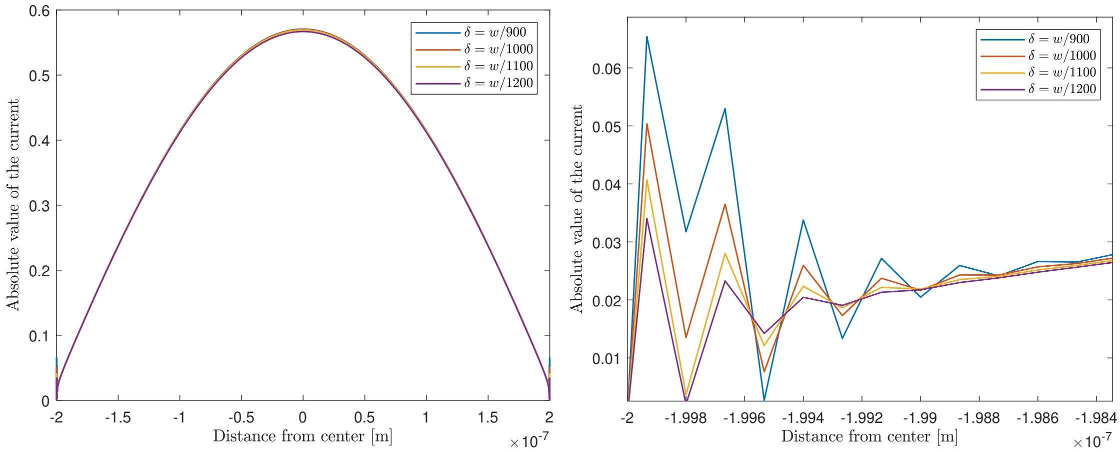

Figure 5.18. Left figure shows a current plot as function of δ, with constant segmentation, $N = 3000$. The legend is the value, K, used for $\delta = w/K$. The right figure is a close up of the end of the gate.

From Fig. 5.18 it is clear that moving the measuring points closer to the gate the consistency of the increase. The cost is, however, a weaker convergence and past $\delta = w/1200 \approx 0.33$ nm it becomes increasingly difficult to say with certainty that it has in fact converged with the tested number of segmentation. For $\delta = w/1100$ the current still converges strongly before $N = 2800$ and the sharp edges has been reduced from $\delta = w/1000$, which will therefore be as the value of δ for the rest of this section.

If the gate is shortened the correlation between convergence and segmentation does not seem to change considerably. But the measuring points have to be moved closer to the gate to find the minimum distance for convergence in order to account for the now inherently smaller segments. The same relation can be found when increasing the gate length, where the measuring points have to placed further away to obtain the minimum distance required for convergence.

Before examining the various cross sections and the radiation patterns from the structure, it would be of interest to test whether the current is actually calculated correctly or if the more numerically stable method derived in Sec. 4.3 is able to reduce some of the difficulties of this method and if the calculated current is consistent with what is obtained through the PPS model.

5.3 The RF Method

In this section the model derived in Sec. 4.3 will be used in to determine the current in the gate strip. The advantage of this model should be that in contrast to the integral of Eq. (5.10), the integrand of Eq. (4.86), $Z(k_x)$, goes towards zero in a more numerical stable manner due to the exponential functions in the numerator naturally going to zero for large k_x, as

$$k_{y,b} = i\sqrt{k_x^2 - k_0^2\varepsilon_b} \simeq ik_x\left(1 - \frac{1}{2}\frac{k_0^2\varepsilon_b}{k_x^2}\right), \tag{5.13}$$

which would give a exponential decreasing function in the numerator of Eq. (4.86). This is indeed confirmed by plotting the $Z(k_x)$, shown in Fig. 5.19.

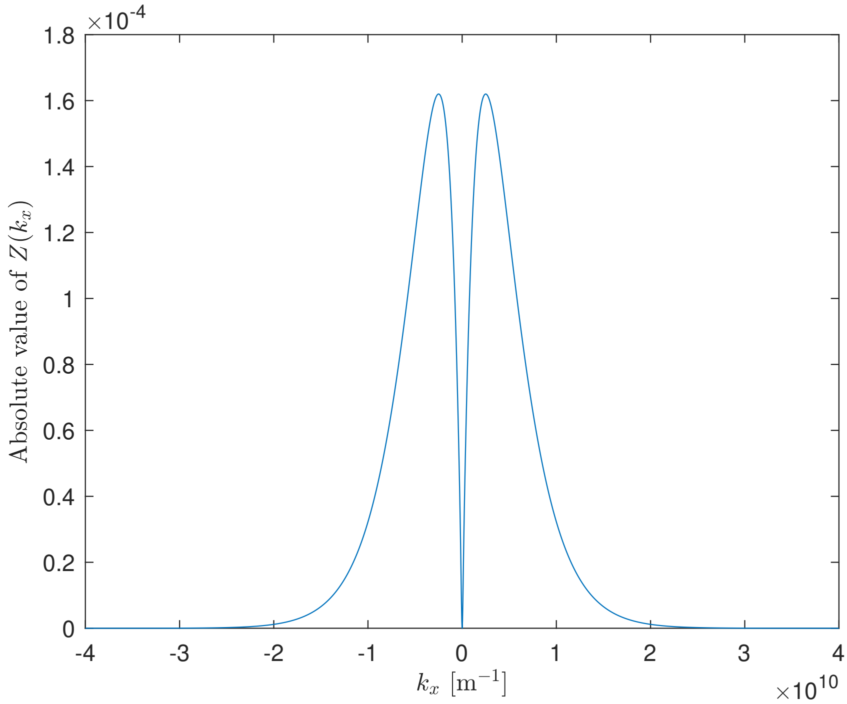

Figure 5.19. The function $Z(k_x)$ from Eq. (4.81) exemplifying the exponentially decreasing nature of the function.

Similarly to the method previous method the current in the gate is described through point matching. The same set of equations are obtained, i.e. Eq. (5.5) and Eq. (5.6), but the function $Z(x - x')$ are that of Eq. (4.86). Initially the distance δ is calculated from Eq. (5.12), in order to examine if this method has the same issues as the method analysed in Sec. 5.2. The current has been plotted for a number of segmentation in Fig. 5.20.

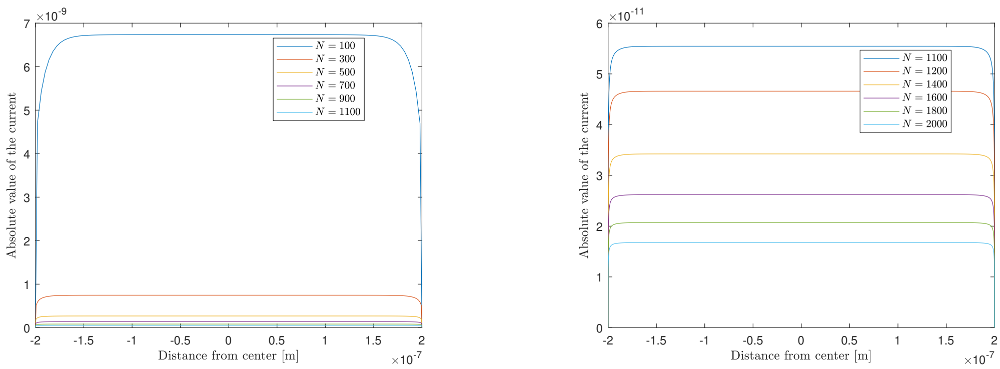

Figure 5.20. Convergence plot for the absolute value of the current. On the left N=100, 200, 500, 700, 900, 1100. On the right N=1100, 1200, 1400, 1600, 1800, 2000.

The same problem of no convergence as was previously observed is still present if the displacement δ follows the segmentation. But, as is clear from Fig. 5.20, the limitations of the number of segmentation have been removed with the new integrand. This is in itself a large improvement to the previous method.

The displacement δ is now set to a constant value and the current is again determined for an increasing number of segments, the results for a 400 nm gate is shown in Fig. 5.21.

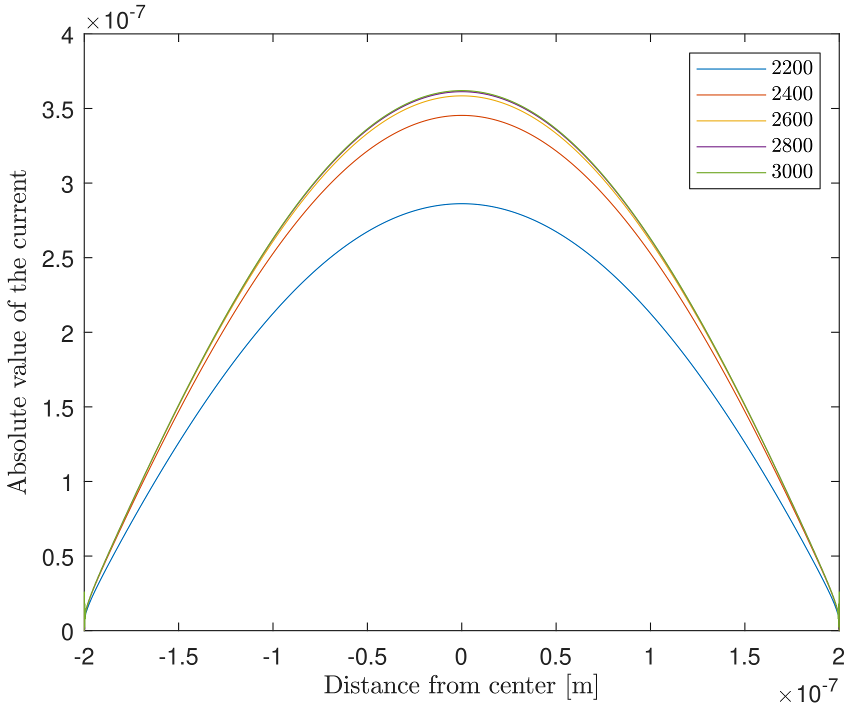

Figure 5.21. Convergence plot for a 400 mn gate with constant displacement of $\delta = w/1100 = 0.36364$ nm with different number of segmentation N = 2200, 2400, 2600, 2800, 3000.

Again with the constant displacement the current converges and to the same shape as seen in Fig. 5.17. However, the value for the current distribution has dropped by a factor of 10^{-6}, which is a large discrepancy between the two approaches. There is no obvious reason for this difference as the parameters used for both calculations are identical.

In order to improve the consistency of the RF method, it is expanded upon by using second order basis functions to possible give a better description of the current and to improve the convergence as having to use up to 3000 segments for convergence for sub-micrometer gates cannot be said to be great. The RF is chosen as this method omits some of the numerical difficulties of the PPS method and are in this sense more versatile.

5.4 Second Order Basis functions

In order to improve the convergence of the method, instead of assuming the current is constant in each segment, second order basis functions are used to improve the method. The basis functions are given as

$$f^{(0)}(x) = 2x^2 - 3x + 1, \tag{5.14}$$

$$f^{(1)}(x) = -4x^2 + 4x, \tag{5.15}$$

$$f^{(2)}(x) = 2x^2 - x. \tag{5.16}$$

$$\tag{5.17}$$

Starting from Eq. (4.84), the current can be written as

$$I_x(x') = \sum_{n=1}^{N}\sum_{\nu=0}^{2} a_n^{(\nu)} f_n^{(\nu)}, \tag{5.18}$$

where the basis functions are given as

$$f_n^{(\nu)}(x) = f^{(\nu)}\left(\frac{x - x_n^{(s)}}{x_n^{(e)} - x_n^{(s)}}\right) \quad \text{for} \quad x_n^{(s)} < x < x_n^{(e)}, \tag{5.19}$$

where $x_n^{(s)}$ is the start of element n and $x_n^{(e)}$ is the end of element n, and $f_n^{(\nu)} = 0$ everywhere else. Inserting this into Eq. (4.84) gives

$$E_{x,ref}(x, y = d - \delta) = \sum_{n=1}^{N}\sum_{\nu=0}^{2} \frac{a_n^{(\nu)}}{2\pi} \int_{-\frac{w}{2}}^{\frac{w}{2}} \int_{-\infty}^{\infty} f_n^{(\nu)}(x') Z(k_x) e^{ik_x(x-x')} \mathrm{d}k_x \mathrm{d}x'. \tag{5.20}$$

As the basis functions are zero when not in the appropriate region, the integral over x' can be split into integrals over each section. Looking at a specific section, n, gives

$$E_{x,ref}(x, y = d - \delta) = \sum_{\nu=0}^{2} \frac{a_n^{(\nu)}}{2\pi} \int_{x_n^{(s)}}^{x_n^{(e)}} \int_{-\infty}^{\infty} f_n^{(\nu)}(x') Z(k_x) e^{ik_x(x-x')} \mathrm{d}k_x \mathrm{d}x'. \tag{5.21}$$

The integral over x' can be identified as the Fourier transform into k-space:

$$\tilde{f}_n^{(\nu)}(k_x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x') e^{-ik_x x'} \mathrm{d}x' = \frac{1}{2\pi} \int_{x_n^{(s)}}^{x_n^{(e)}} f(x') e^{-ik_x x'} \mathrm{d}x'. \tag{5.22}$$

Inserting this into Eq. (5.21) gives

$$E_{x,ref}(x, y = d - \delta) = \sum_{n=1}^{N}\sum_{\nu=0}^{2} a_n^{(\nu)} \int_{-\infty}^{\infty} \tilde{f}_n^{(\nu)}(k_x) Z(k_x) e^{ik_x x} \mathrm{d}k_x. \tag{5.23}$$

In order to evaluate Eq. (5.22), the variable is changed from x' to $z = \frac{x' - x_n^{(s)}}{\Delta}$, where $\Delta = |x_n^{(e)} - x_n^{(s)}|$, which gives

$$\tilde{f}_n^{(\nu)}(k_x) = \frac{1}{2\pi} \int_0^1 f^{(\nu)}(z) e^{-ik_x\left(\Delta z + x_n^{(s)}\right)} \Delta \mathrm{d}z. \tag{5.24}$$

The basis functions are then inserted from Eqs.(5.14), (5.15), and (5.16). The first basis function becomes

$$\tilde{f}_n^{(0)}(k_x) = \frac{\Delta e^{-ik_x x_n^{(s)}}}{2\pi} \int_0^1 \left(2z^2 - 3z + 1\right) e^{-ik_x \Delta z} \mathrm{d}z = A_n \int_0^1 \left(2z^2 - 3z + 1\right) e^{-ik_x \Delta z} \mathrm{d}z, \tag{5.25}$$

where $A_n = \frac{\Delta e^{-ik_x x_n^{(s)}}}{2\pi}$. The integral can be split into three separate integrals as

$$\tilde{f}_n^{(0)}(k_x) = A_n \left(\int_0^1 2z^2 e^{-ik_x \Delta z} \mathrm{d}z - \int_0^1 3z e^{-ik_x \Delta z} \mathrm{d}z + \int_0^1 e^{-ik_x \Delta z} \mathrm{d}z \right). \tag{5.26}$$

The last integral can be readily integrated, giving

$$\int_0^1 e^{-ik_x\Delta z}\mathrm{d}z = \left[-\frac{e^{-ik_x\Delta z}}{ik_x\Delta}\right]_0^1 = \frac{1-e^{-ik_x\Delta}}{ik_x\Delta}. \tag{5.27}$$

The second integral is evaluated using integration by parts giving,

$$\begin{aligned}\int_0^1 3ze^{-ik_x\Delta z}\mathrm{d}z &= \left[-\frac{3ze^{-ik_x\Delta z}}{ik_x\Delta}\right]_0^1 + \int_0^1 3\frac{e^{-ik_x\Delta z}}{ik_x\Delta}\mathrm{d}z \\ &= \left[-\frac{3ze^{-ik_x\Delta z}}{ik_x\Delta}\right]_0^1 + \left[\frac{3e^{-ik_x\Delta z}}{k_x^2\Delta^2}\right]_0^1 = -\frac{3e^{-ik_x\Delta z}}{ik_x\Delta} + 3\frac{e^{-ik_x\Delta z}-1}{k_x^2\Delta^2}.\end{aligned} \tag{5.28}$$

The first integral is also handled by integration by parts, yielding

$$\begin{aligned}&\int_0^1 2z^2e^{-ik_x\Delta z}\mathrm{d}z = \left[-2z^2\frac{e^{-ik_x\Delta z}}{ik_x\Delta}\right]_0^1 + \int_0^1 4z\frac{e^{-ik_x\Delta z}}{ik_x\Delta}\mathrm{d}z \\ &= -\frac{2e^{ik_x\Delta}}{ik_x\Delta} + \left[4z\frac{e^{-ik_x\Delta z}}{k_x^2\Delta^2}\right]_0^1 - \int_0^1 4\frac{e^{-ik_x\Delta z}}{k_x^2\Delta^2} = -\frac{2e^{-ik_x\Delta}}{ik_x\Delta} + \frac{4e^{-ik_x\Delta}}{k_x^2\Delta^2} + \frac{4e^{-ik_x\Delta}-4}{ik_x^3\Delta^3}.\end{aligned} \tag{5.29}$$

Inserting these into Eq. (5.26) gives the Fourier transformed basis function as

$$\begin{aligned}\tilde{f}_n^{(0)}(k_x) = A_n\Bigg(&-\frac{2e^{-ik_x\Delta}}{ik_x\Delta} + \frac{4e^{-ik_x\Delta}}{k_x^2\Delta^2} \\ &+\frac{4e^{-ik_x\Delta}-4}{ik_x^3\Delta^3} + \frac{3e^{-ik_x\Delta z}}{ik_x\Delta} - \frac{3e^{-ik_x\Delta z}-3}{k_x^2\Delta^2} + \frac{1-e^{-ik_x\Delta}}{ik_x\Delta}\Bigg) \\ &= A_n\left(\frac{1}{ik_x\Delta} + \frac{e^{-ik_x\Delta}+3}{k_x^2\Delta^2} + \frac{4e^{-ik_x\Delta}-4}{ik_x^3\Delta^3}\right).\end{aligned} \tag{5.30}$$

The other Fourier transformed basis functions are similarly derived in Appendix, A.7, and are

$$\tilde{f}_n^{(1)}(k_x) = 4A_n\left(-\frac{e^{-ik_x\Delta}+1}{k_x^2\Delta^2} - \frac{2e^{-ik_x\Delta}-2}{ik_x^3\Delta^3}\right), \tag{5.31}$$

and

$$\tilde{f}_n^{(2)}(k_x) = A_n\left(-\frac{e^{-ik_x\Delta}}{ik_x\Delta} + \frac{3e^{-ik_x\Delta}+1}{k_x^2\Delta^2} + \frac{4e^{-ik_x\Delta}-4}{ik_x^3\Delta^3}\right). \tag{5.32}$$

These basis functions can be inserted into Eq. (5.23) and the current can be found. The basis functions are dependent on $1/k_x^3$, however they are still stable in the limit $k_x \to 0$, which can be shown by using the Taylor expansion on the exponential function $e^{-ik_x\Delta}$, giving

$$e^{-ik_x\Delta} \approx 1 - ik_x\Delta - \frac{1}{2}k_x^2\Delta^2. \tag{5.33}$$

Inserting this into Eq. (5.30) gives

$$\tilde{f}_n^{(0)}(k_x) \approx A_n\left(\frac{1}{k_x\Delta}[-i-i+2i] + \frac{1}{(k_x\Delta)^2}[1+3-4] + \frac{1}{(k_x\Delta)^3}[4-4] - \frac{1}{2}\right) = -\frac{A_n}{2}, \tag{5.34}$$

and for Eq. (5.31) it gives

$$\tilde{f}_n^{(1)}(k_x) \approx 4A_n \left(\frac{1}{k_x\Delta}[i-i] + \frac{1}{(k_x\Delta)^2}[-2+2] + \frac{1}{(k_x\Delta)^3}[2-2] + \frac{1}{2} \right) = 2A_n, \quad (5.35)$$

and for Eq. (5.32)

$$\tilde{f}_n^{(2)}(k_x) \approx A_n \left(\frac{1}{k_x\Delta}[i-3i+2i] + \frac{1}{(k_x\Delta)^2}[3+1-4] + \frac{1}{(k_x\Delta)^3}[4-4] + 1 - i\frac{1}{2}k_x\Delta - \frac{3}{2} \right)$$
$$= A_n \left(-\frac{1}{2}ik_x\Delta - \frac{1}{2} \right). \quad (5.36)$$

Thus for $k_x \to 0$ the integrals are not singular.

To properly handle the extra unknowns introduced from the second order basis functions, the middle point of each segment is also incorporated in the calculations, thus for N segments, there are $2N+1$ points.

5.4.1 Current Analysis

The current from the second order basis functions are then found through the integral in Eq. (5.23). Again the translational symmetry can be used to decrease computation time. In order to avoid the singular behaviour of the basis functions at $k_x = 0$, a distance in k_x, δ_k, is introduced, where the second order Taylor expansion for the basis functions are used. The Taylor expansion in Eq. (5.33) is expanded for $\Delta k_x \approx 0$, and as Δ depends on the segmentation, N, δ_k is set to $\delta_k = A/\Delta$, in order to keep δ_k constant for changing segmentation. The current is illustrated for different amounts of segments in Fig. 5.22.

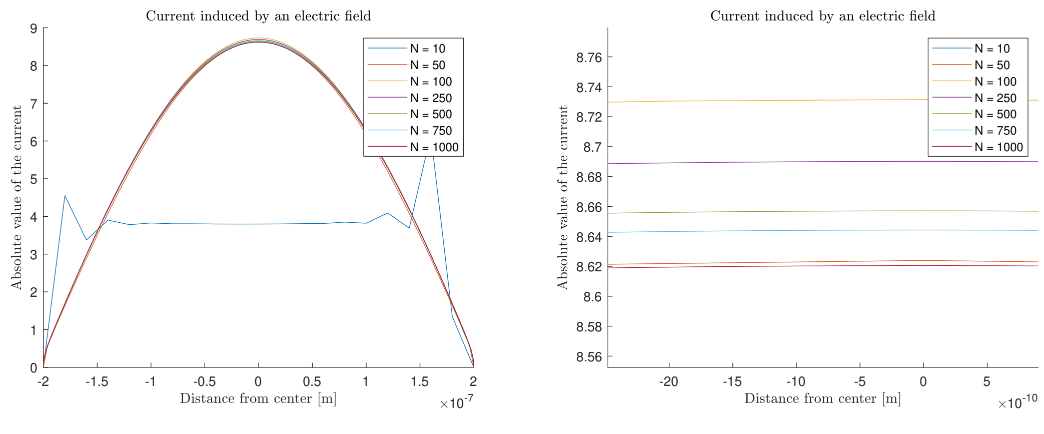

Figure 5.22. The current in the gate for varying segmentation. The values used for the calculations are for a AlGaN/GaN structure. As the segments now have a middle point, the displacement is set to $\delta = \Delta/4$, $\delta_k = 10^{-4}/\Delta$. The right plot is zoomed on the top of the curve.

It can be seen from the figure, the value seems to converge decently for a relatively low segmentation, for $N \geq 50$ the values seem to differ very little. After reaching a plateau, the values seem to decrease as N, the correlation between the max value of the current and segmentation is seen in Fig. 5.23.

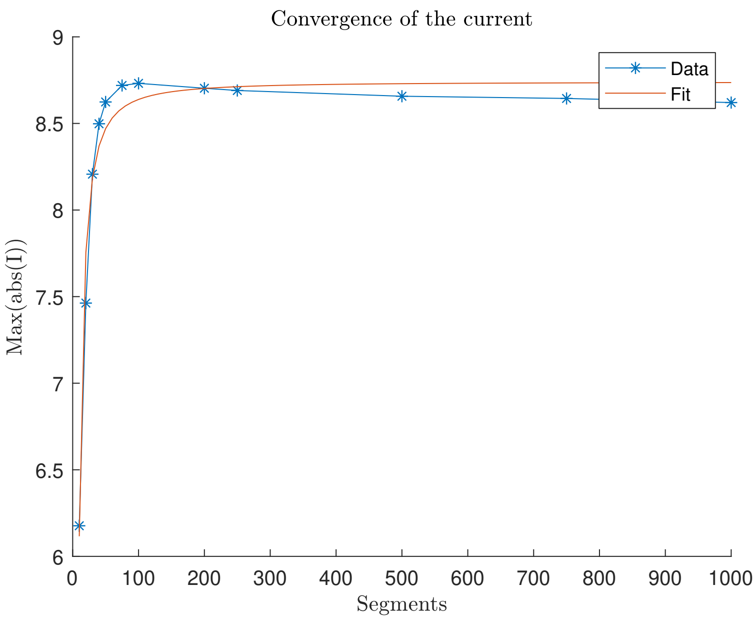

Figure 5.23. The maximum value of the current in the gate for varying segmentation with a fit of $aN^{-b} + c$.

The convergence in the plot seem to follow a $aN^{(-b)} + c$ tendency, with coefficients $a = -67.8805$, $b = 1,4128$ and $c = 8.7399$. The current fit does seem to follow the convergence decently, however for higher amounts of segments, the current has a lower value than the fit, thus the approximated converged current of $I_c = 8.7399$ is likely higher than the real converged value. The percentage difference in the maximum value of the current between $N = 100$ and $N = 1000$, is $\approx 0,64\%$, thus there is little difference in the current for $N > 100$, thus $N = 100$ seems a reasonably convergent value. For the rest of the convergence tests, $N = 100$ has been chosen. In order to test the dependency of δ_k, the function have been plotted with different values of δ_k/Δ. The dependency on δ_k which is shown in Fig. 5.24.

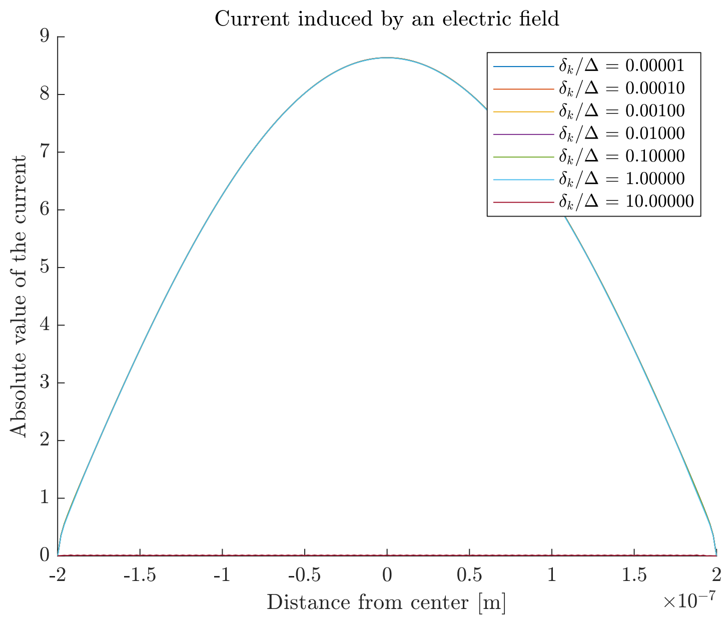

Figure 5.24. The absolute value of the current along the gate for different δ_k.

It seems that as long as $\delta_k/\Delta \leq 1$ the approximation is valid. Setting the $\delta_k/\Delta = 10$, gives a much lower current that all the other δ_k/Δ. The current for $\delta_k/\Delta = 1$ is almost identical to the values for lower δ_k/Δ. The computation time needed for these solutions did not differ significantly, thus $\delta_k = 10^{-4}/\Delta$ is chosen to be the standard δ_k.

The electric field is set to be 0 a small distance, δ, below the gate in order to have the integrals converge. As δ is a significant variable in the expression for the current it would be of interest to examine its effect on the calculated current. The current for different values of δ is shown in Fig. 5.25

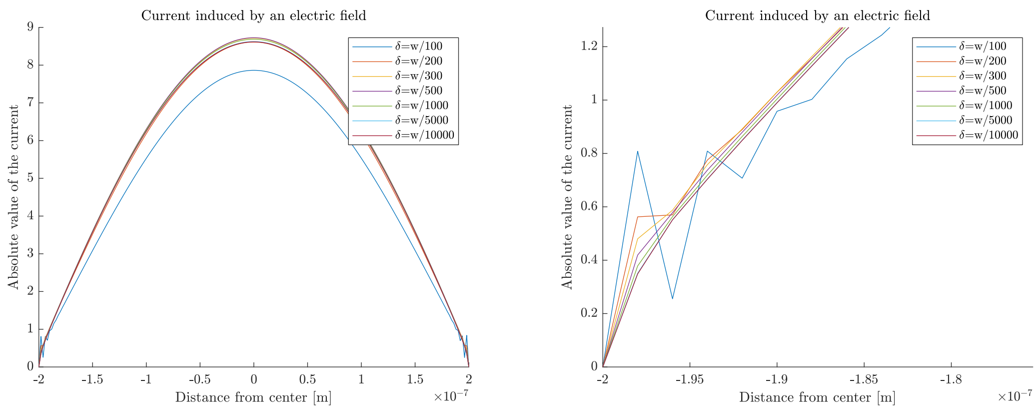

Figure 5.25. The absolute value of the current along the gate for different δ and $N = 100$. The right picture is focused on the corner of the gate.

The overall value does not seem to change much depending on δ, as long as δ is not too big, however the corners seem to have problems with lower values of δ. For $\delta \geq w/500$ the corners seem to be completely smooth. For higher values of δ the corners continues to be smooth, however the overall value of the current seems to decrease. For $\delta \leq w/100$ the model starts to break as is illustrated in Fig. 5.26.

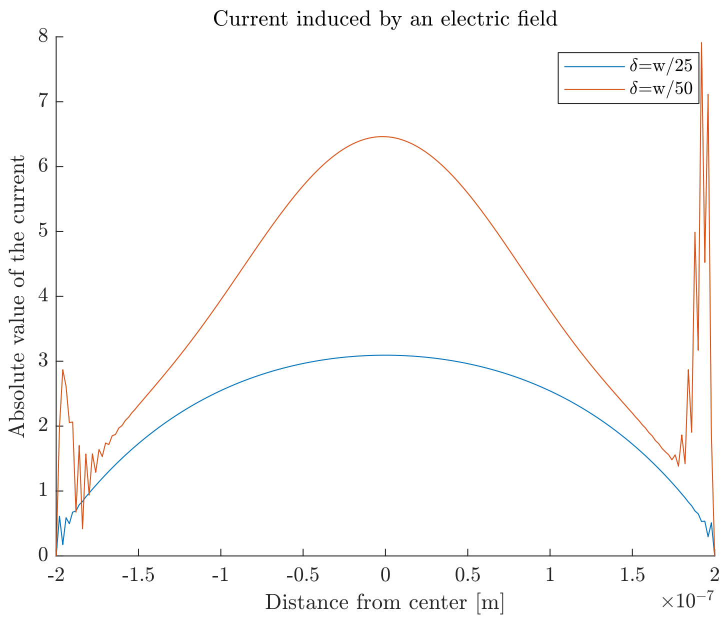

Figure 5.26. The absolute value of the current along the gate for $\delta \leq w/50$ and $N = 100$.

This is likely due to the observation point being to far from the gate, thus the model breaks and the contributions begin to be poorly weighted as was discussed at Fig. 5.16. In Fig. 5.26, the furthest distance was $\delta = 400$ nm$/25 = 16$ nm which corresponds to the measuring point being below the 2DEG. The total value seems to have a peak and then decrease as δ decreases, which is shown in Fig. 5.27.

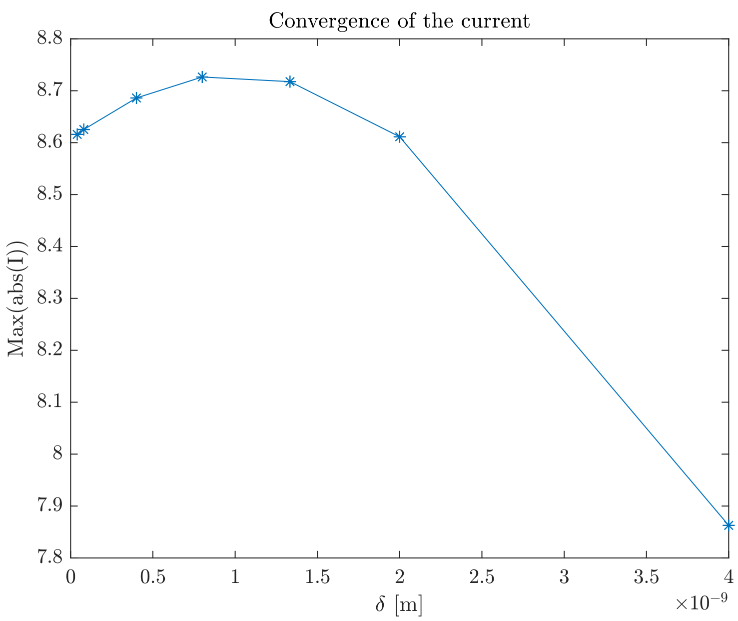

Figure 5.27. The maximum absolute value of the current along the gate for varying δ and $N = 100$.

For low δ the current decreases as δ decreases. This indicates that the model does not converge with respect to δ. Even for $\delta = w/10^5$ the current continues to follow the decreasing tendency shown in Fig 5.27. Using a fixed δ the model seems to converge with respect to N. For a fixed $\delta = w/1500$ the current for increasing segmentation is shown in Fig. 5.28.

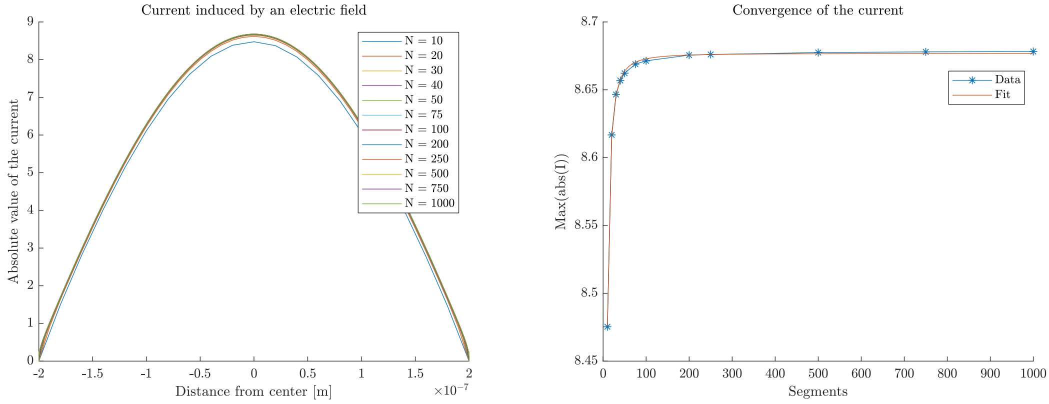

Figure 5.28. The absolute value of the current along the gate for $\delta = w/1500$ and for increasing N. The right picture the convergence of the absolute value of the current with an approximation of $aN^{-b} + c$ plotted.

The current does not seem to differ significantly for different values of N. The convergence of the current seems to follow the form $aN^{-b} + c$ with the values $a = -10.1811$, $b = 1.7043$ and $c = 8.6768$ which suggests that the converged value for the current is $I_c = 8.6768$, however as can be seen from Fig. 5.28, the fit seem to diverge slightly for higher values of N ($N \geq 750$). For high values of N, the current at the corners of the gate starts to break and show strange shapes, which is consistent with Fig. 5.25, thus for high N a smaller δ is required for the corners to be calculated properly. Thus the convergence for a smaller δ is relevant in order to see if the convergence can be consistent. In Fig. 5.29 the current for increasing N and $\delta = w/3000$ is shown.

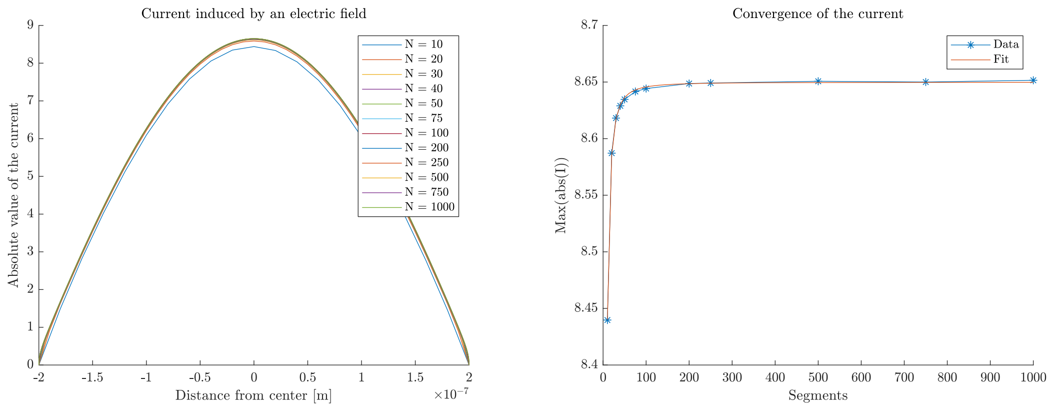

Figure 5.29. The absolute value of the current along the gate for $\delta = w/3000$ and for increasing N. The right picture the convergence of the absolute value of the current with an approximation of $aN^{-b} + c$ plotted.

Again the current seems to differ very little for increasing N, however the overall value of the current is lower than for $\delta = w/1500$, which is to be expected from the tendency seen in Fig. 5.27. The approximation of $aN^{-b} + c$ with $a = -10.6419$, $b = 1.7053$ and $c = 8.6498$

is decent. According to the fit, the converged value of the current is $I_c = 8.6498$, which is lower than the converged value for $\delta = w/1500$. For this smaller δ, the corners are smooth for all N shown in the figure. To see if this trend continues, the convergence for N with $\delta = w/4500$ is shown in Fig. 5.30.

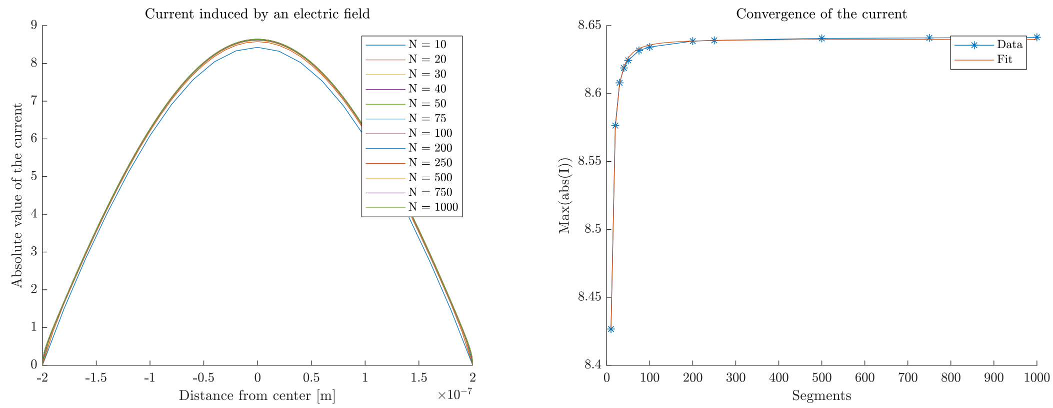

Figure 5.30. The absolute value of the current along the gate for $\delta = w/4500$ and for varying N. The right picture the convergence of the absolute value of the current with an approximation of $aN^{-b} + c$ plotted.

The model seems to converge very quickly with respect to N. The corners are still smooth and the current has dropped a tiny amount. The fit of $aN^{-b} + c$ seems decent with $a = -10.8281$, $b = 1.7063$ and $c = 8.6400$ which would correspond to a converged current of $I_c = 8.6400$, which again is lower than the larger values of δ, which is to be expected.

The model should also be capable of handling different wavelengths, so far only a wave with the frequency of 1.44 THz has been used, but in order to check convergence, a wave with a frequency of 10 THz is used for different N with $\delta = w/3000$ is shown in Fig. 5.31.

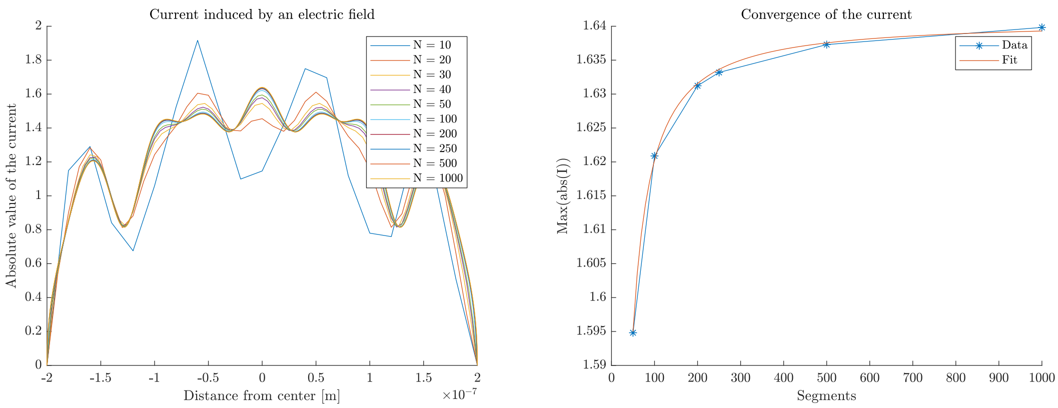

Figure 5.31. The absolute value of the current along the gate for $\delta = w/3000$, a frequency of 10 THz and for varying N. The right picture the convergence of the absolute value of the current with an approximation of $aN^{-b} + c$ plotted.

For 10 THz, the model seems to converge a bit slower, possibly because of the more complex pattern of the current. For low N, the current is not symmetric and does not represent the pattern well. The approximation of $aN^{-b} + c$ was again used to check for convergence, and the values used are $a = -4.2742$, $b = 1.1598$ and $c = 1.6407$. For this situation, $N \geq 50$ seems to be necessary for a reasonably converged value.

In order to check if the convergence depends on the gate length, different lengths of gates are checked in order to insure convergence for several cases. First a gate with a length of 100 nm and $\delta = w/1500$ is used, and the convergence plot is shown in Fig. 5.32.

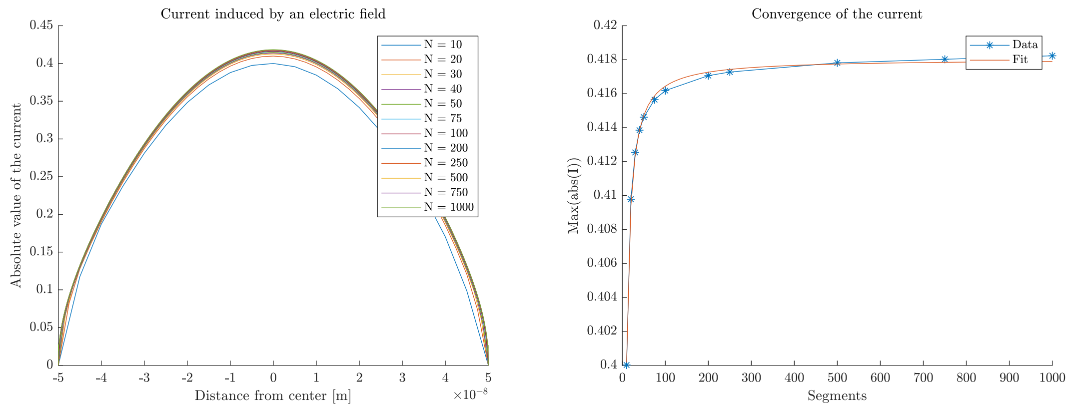

Figure 5.32. The absolute value of the current along the gate for $\delta = w/1500$, a gate length of 100 nm and for varying N. The right picture the convergence of the absolute value of the current with an approximation of $aN^{-b} + c$ plotted.

The current seems to converge very quickly for this gate length as well. The approximation of $aN^{-b} + c$ with values $a = -0.1994$, $b = 1.0477$ and $c = 0.4180$ seems to fit decently. The convergence has also been tested for $\delta = w/3000$ and $\delta = w/4500$, which seems to follow the same tendencies, as δ gets lower, the current in the gate drops, however the convergence for N still seems strong, this is shown in Fig. 5.33.

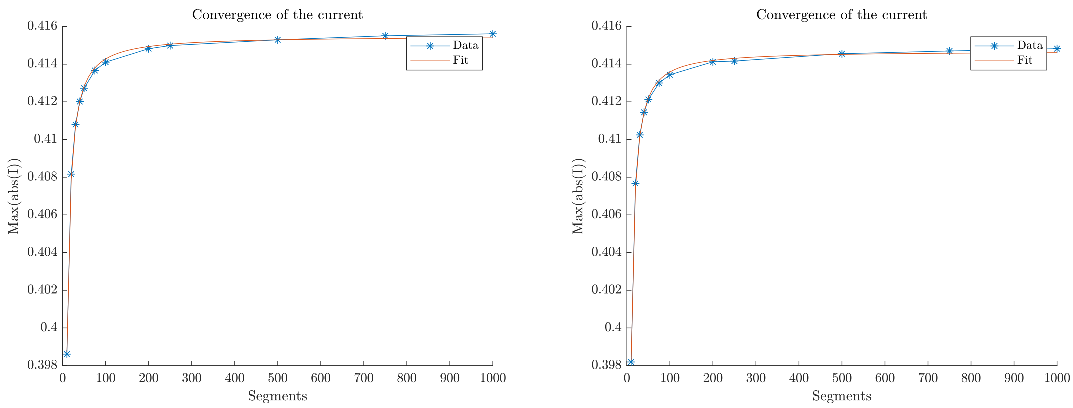

Figure 5.33. Both plots are of the maximum of the absolute value of the current for an antenna with length $w = 100$ nm and with an approximation of $aN^{-b} + c$ plotted. The left picture is for $\delta = w/3000$ and the right for $\delta = w/4500$.

The fit for $\delta = w/4500$ is $a = -0.2333$, $b = 1.1433$ and $c = 0.4155$ and for $\delta = w/3000$ it is $a = -0.1994$, $b = 1.0477$ and $c = 0.4180$, thus the lower value of δ has a lower approximated converged value, which is consistent with the previous results.

In order to check the convergence for longer gates, a gate length of $w = 1000$ nm is checked. For $\delta = w/1500$ and varying N is shown in Fig. 5.34.

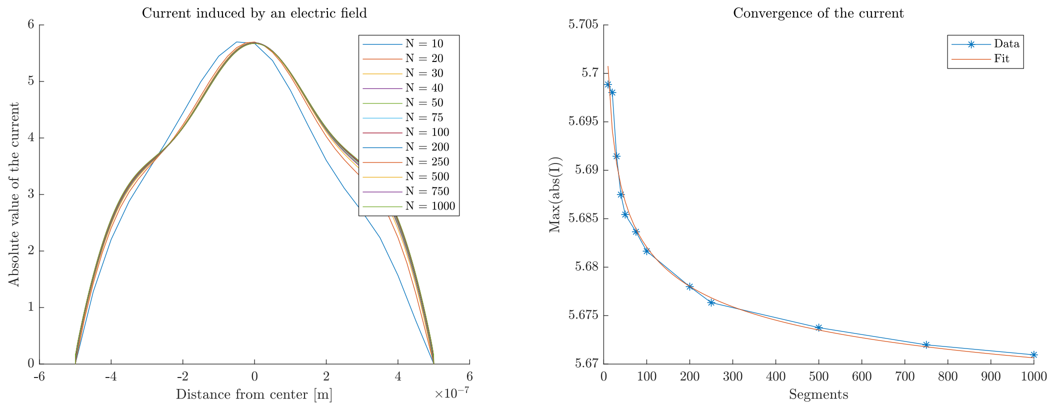

Figure 5.34. The absolute value of the current along the gate for $\delta = w/1500$, a gate length of 1000 nm and for varying N. The right picture the convergence of the absolute value of the current with an fit of $aN^{-b} + c$ plotted.

The current seems decently converged for $N \approx 30$. The shape seems to shift to the right as the number of segments increase. The power fit describes this setup well, and the current seems to converge. The values used are $a = 0.0787$, $b = 0.2138$ and $c = 5.6527$. The current does seem to converge towards a higher value than the fit suggests. The convergence for different values of δ have been done as well and the convergence is shown in Fig. 5.35.

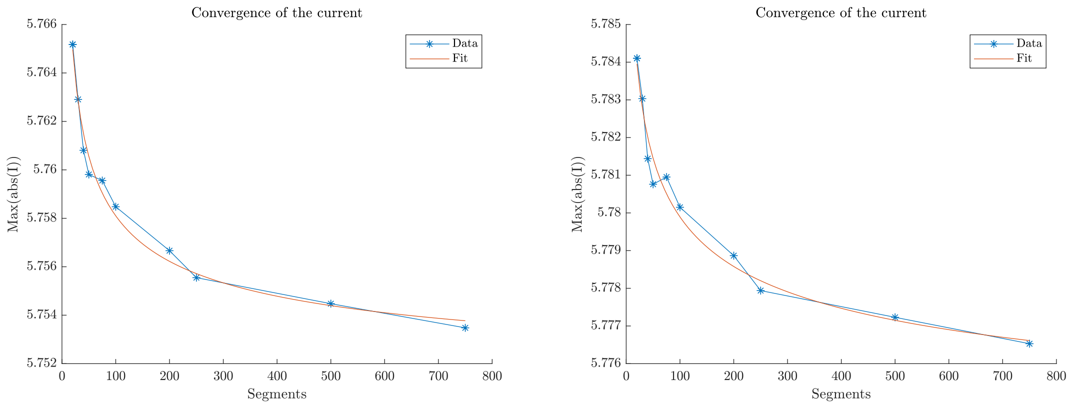

Figure 5.35. Both plots are of the maximum of the absolute value of the current for an antenna with length $w = 1000$ nm and with an fit of $aN^{-b} + c$ plotted. The left picture is for $\delta = w/3000$ and the right for $\delta = w/4500$.

The power fit for $\delta = w/3000$ have the coefficients $a = 0.0471$, $b = 0.3840$ and $c = 3.1074$ and for $\delta = w/4500$ have the coefficients $a = 0.0260$, $b = 0.2446$ and $c = 5.7715$. It seems as δ decreases, the current increases, which differs from earlier results. The convergence pattern is also significantly different.

In order to see better analyse the dependency on δ, the gate with $w = 1000$ nm has been tested for $N = 100$ and varying δ, in order to test the convergence. The result is shown in Fig. 5.36.

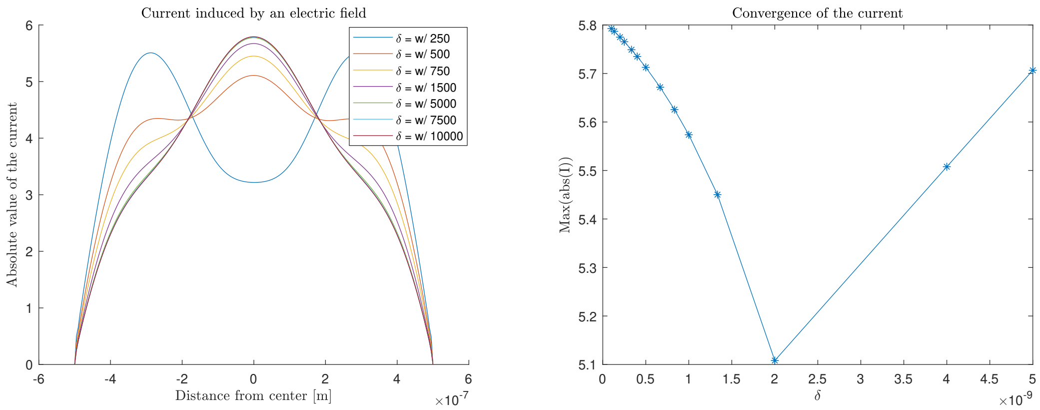

Figure 5.36. The absolute value of the current along the gate for varying δ, a gate length of 1000 nm and for $N = 100$. The right picture the convergence of the absolute value of the current.

It seems that for high values of δ, the pattern is inconsistent, however as δ decreases the pattern seems to converge as well as the maximum value of the current.

When taking $\delta \to 0$ the model starts to break and MATLAB is unable to solve the integrals properly, thus limiting the possible δ to check. As the model seems reasonably convergent within the tested parameters the current of the gate found for structures within these tested parameters, can be used for e.g. calculating radiation patterns, cross sections and determining plasma wave resonances.

5.5 Radiation Pattern Analysis

For the following analysis the far field radiation pattern, as well as optical cross sections, introduced in Sec. 2.2, will be analysed for possible resonances as function of frequencies and gate lengths, with the effect of parameters such as gate to 2DEG distance examined. At first an InGaAs HEMT structure will be analysed in order to compare the RF model with a similar analyses given in Ref. [34]. Afterwards an AlGaN/GaN HEMT, with parameters given in the analysis introduction, will be studied and compared to the InGaAs for similarities and differences.

5.5.1 InGaAs HEMT

For this section the RF model will be used on the InGaAS HEMT analysed in Ref. [34] in order to compare it with the inspiration for the PPS model. The parameters used for the InGaAs HEMT are: dielectric constants $\varepsilon_b = 13.88$, $\varepsilon_s = 13.88$, and sheet carrier density $N_c = 3 \times 10^{12}$ m^{-2}, relaxation time $\tau = 2.8 \times 10^{-12}$ s, and the effective mass is taken as $m = 0.042 m_0$ [34].

First the far field radiation pattern calculated from Eq. (2.21), using the different models for obtaining the gate current, is compared to that of Fig. 7 in Ref. [34], which is added for easier comparison. The radiation patterns are shown in Fig. 5.37.

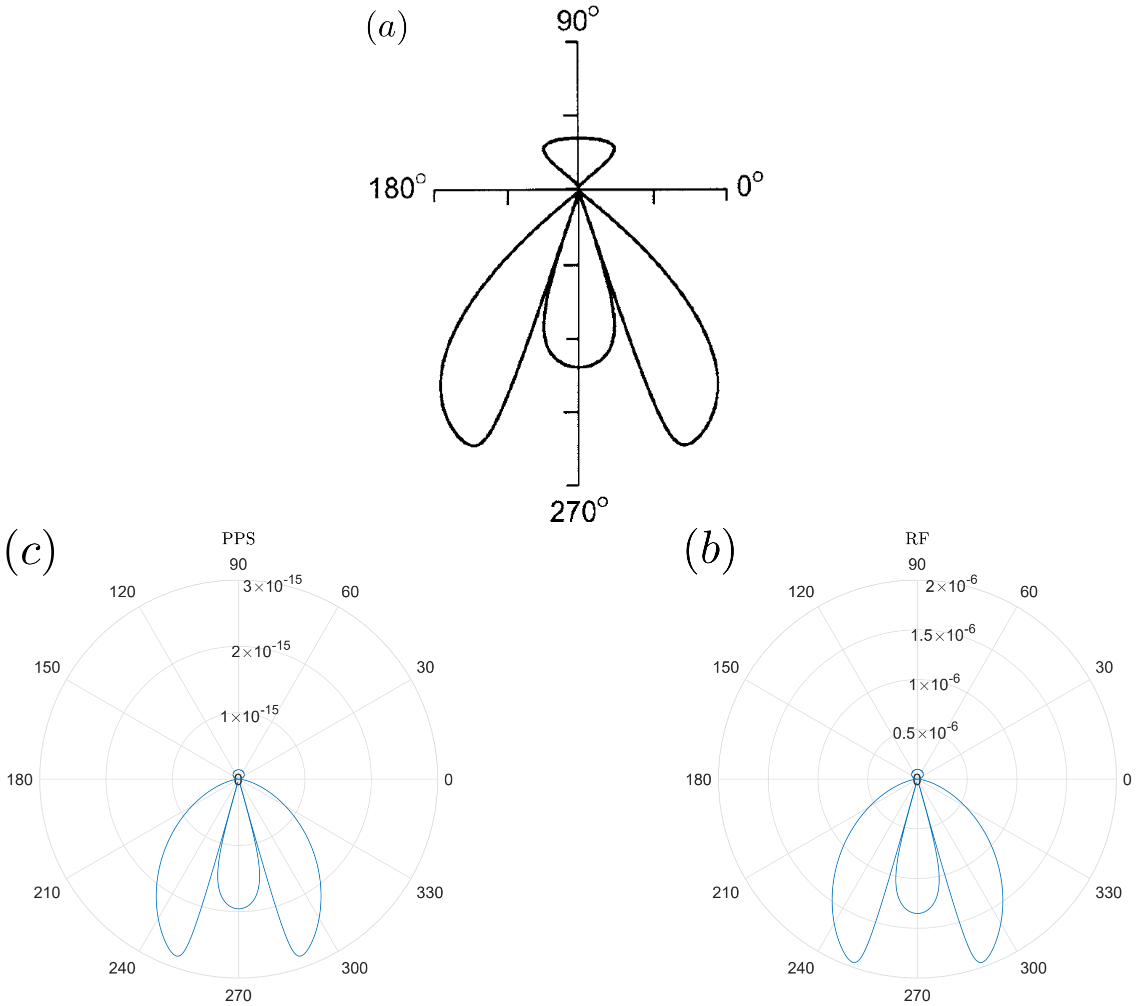

Figure 5.37. Far field radiation pattern calculated for an InGaAs HEMT, with parameters of: gate length w = 400 nm, distance to the 2DEG d = 10 nm, ε = 13.88, m^* = 0.042m_0 and τ = 2.8 $\times$ 10^{-12} s for a frequency of 1.44 THz. The figures are obtained by the following: (a) is obtained from Ref. [34], (b) is calculated by RF using second order basis function and (c) by PPS with point matching.

It is clear that the three calculated by the models in the project ((b),(c)), differ by some amount from the one calculated in Ref. [34], (a), especially the lobe in the ambient layer are significantly larger than those calculated by the derived models. Furthermore, its shape is also quite different the spherical lobes of the other models, with very sharp cutoff angles.

The scattered field in the substrate have over all a more similar shape. The bottom side lobes of (a) begins at about 40°, whereas the other three has power going out at an angle of 30°, measured from the horizontal axis. This is the same as is seen for the upper lobe, where the intensity increases more rapidly with the angle than for the other models. The main lobe of the fields in this project have approximately the same ratio between the power going directly up (90°) and directly down (270°) of 1390% for RF and PPS, more power going directly into the substrate. For the substrate lobes in Fig. 5.37(a), in Ref. [34] they calculate the total reflection angles θ_r in which they obtain $\theta_r = 270^\circ \pm \sin^{-1}((1/\varepsilon_s^{1/2})) \approx 270^\circ \pm 17{,}3^\circ$, which almost corresponds to what is observed in the figure, it is in actuality closer to 20°. However, by using this angle and calculating the used ε_s one obtains ~ 11.3, which differs from what is stated has been used. This change in substrate dielectric could explain the difference of scattering into the ambient medium as the far field calculated in this section has been done with equal substrate and barrier/insulator dielectric $13.88 = \varepsilon_b = \varepsilon_s$, which would limit the reflection from the interface between these two mediums to the term provided by the 2DEG in Eq. (4.60). Calculating the value for r_{bs} for a normal incident wave for both mediums having the same dielectric constant $\varepsilon_{b,s} = 13.88$ gives a reflection coefficient of $0.0011 + 0.0279i$, while using different dielectric constants, $\varepsilon_b = 13.88$ and $\varepsilon_s = 11.3$ one obtains a reflection coefficient of $-0.0502 + 0.0208i$, which is a relatively large difference for the real part. The effect of this modification on the scattering is shown in Fig. 5.38, where it has been calculated using the RF method.

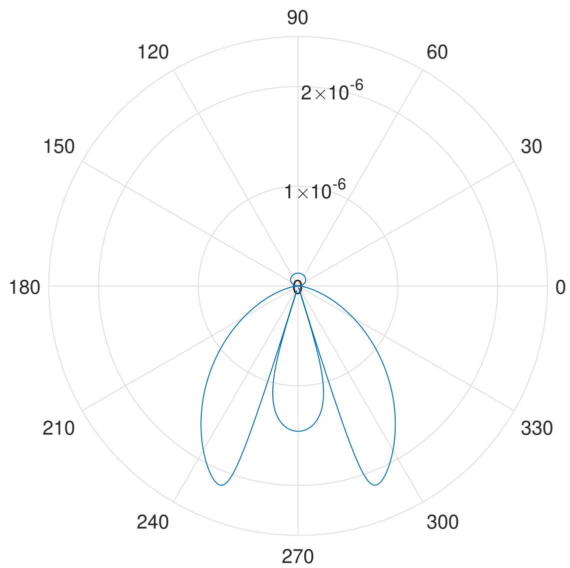

Figure 5.38. Far field radiation pattern using the same parameters as Fig. 5.37 with the substrate dielectric constant change to $\varepsilon_s = 11.3$.

Visually the change is very minor, comparing it with Fig. 5.37(b) one can observe that scattering into the substrate has clearly been reduced. Likewise, the scattering into the ambient can be seen to be slightly larger. This is also confirmed by like before by calculating the relative size of the power scattered directly up and down. In this case the ratio has fallen to 1289% from 1387%. The change in dielectric constant of the substrate does, however, not reproduce the far field pattern from Fig. 5.37(a). But, if the angle of total reflection is calculated for $\varepsilon_s = 13.88$ an angle of $270° \pm 15{,}6°$, which is the angle measured on Fig. 5.37(b) and as stated $\varepsilon_s = 11.3$ corresponds to $\theta_r = 270° \pm 17{,}3°$, which is also what is seen in Fig. 5.38. As stated earlier by actually measuring the angle in Fig. 5.37 one gets an angle of $\sim 20°$, again different from what is stated or used for the calculation, the corresponding substrate dielectric for this angle is 8.55, the calculated far field is shown in Fig. 5.39.

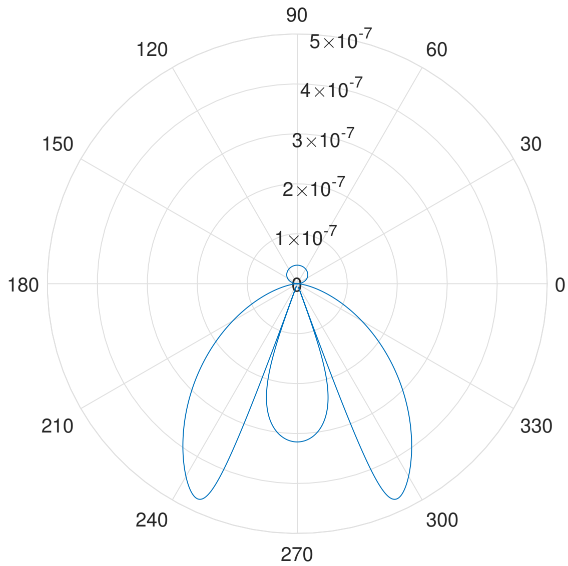

Figure 5.39. Far field radiation pattern using the same parameters as Fig. 5.37 with the substrate dielectric constant change to $\varepsilon_s = 8.55$.

As expected by the previous reasoning this increases the scattering into the ambient medium, by having a greater disparity in dielectric constants between substrate and barrier layer. However, it still does not reproduce the exact far field pattern of Fig. 5.37(a). The answer to this disparity might lie in the fact that the $Z(k_x)$, Eq. (4.46), obtained by the same method as prescribed in Ref. [34] differ from the one shown in the reference.

Likewise for comparison the scattering cross sections has been calculated from Equations (2.15) and (2.16). In Ref. [34] the total scattering cross has been calculated for a 1000 nm gate with varying distances to the 2DEG layer. A corresponding calculation has been done and the results are shown side by side in Fig. 5.37.

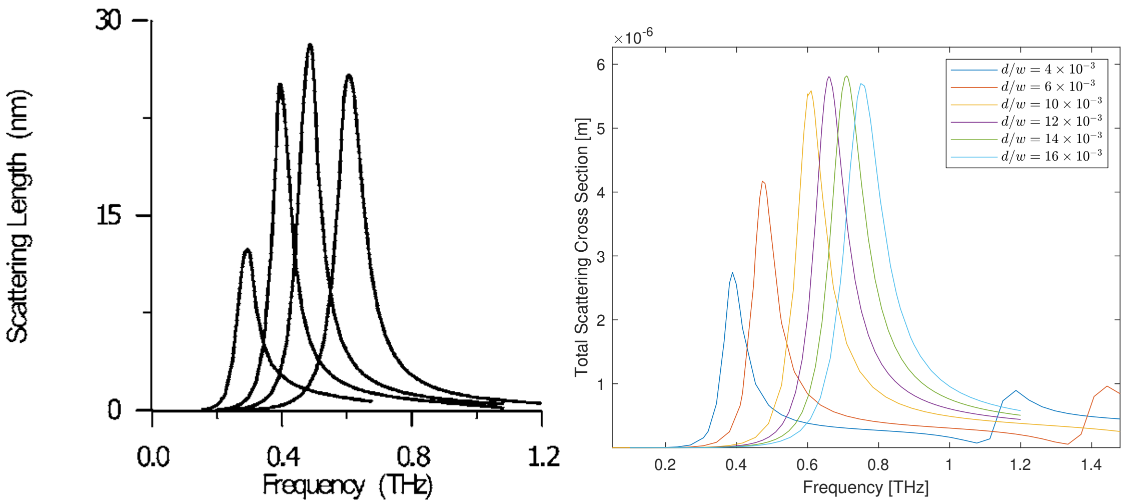

Figure 5.40. Total scattering cross section for a 1000 nm gate as function of frequency with different values of gate to 2DEG layer distances. On the left obtained from Ref. [34], for distances of $d/w(10^{-3})$: 2, 4, 6 and 10. On the right the calculated values using RF for $d/w(10^{-3})$: 4, 6, 10, 12, 14 and 16. The parameters are otherwise those stated in Fig. 5.37.

In terms of position of the resonance frequencies the model in this project are in agreement with the ones of Ref. [34], with the resonance frequencies being 0.4 THz, 0.5 THz and 0.6 THz for $d/w(10^{-3})$= 4, 6 and 10 respectively. For the $d/w(10^{-3})$=4 scattering length calculated begins to increase again at the edge of the calculated interval, this is not seen on the left plot of Fig. 5.40, however, examining the first peak of this figure, the scattering length is not calculated for the entire interval. This might also be true for the corresponding $d/w(10^{-3})$=2 peak and as such this second peak is not shown.

Otherwise, the optimal distance to the 2DEG obtained by RF is different from the reference. For this HEMT, a gate to 2DEG distance of around 12-14 nm is seen to have the largest resonance peak where afterwards the resonance decreases. The increase in peaks from the 6 nm and 10 nm compared to the larger distances is quite large. This is essentially also observed on the left of Fig. 5.40 but at a smaller distance. It would seem that the resonances calculated here are shifted towards higher frequencies compared to those in [34].

Likewise, the values calculated for all gate to 2DEG distances are approximately a factor 100 larger than those of Ref. [34], which far exceeds the geometrical width of the gate. Similarly for the absorption length/total extinction cross section the maximum resonance are shifted to the higher frequencies, as was observed for the scattering cross section. Note that the extinction cross section from Ref. [34] is simply obtained by adding the values of the two left figures of Fig. 5.40 and Fig. 5.41 together.

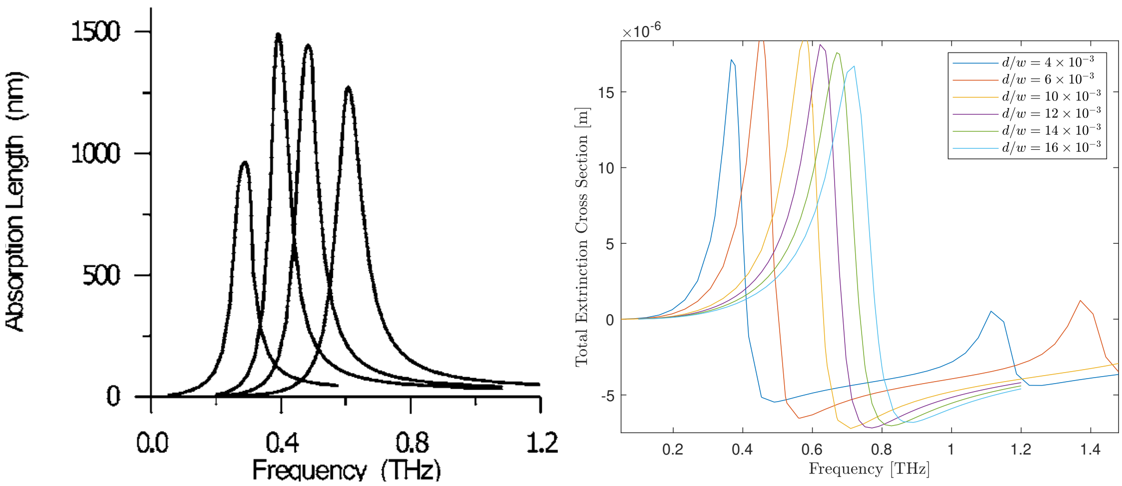

Figure 5.41. Total absorption cross section from Ref. [34] and total extinction cross section calculated by the RF model both for a 1000 nm gate as function of frequency with different values of gate to 2DEG layer distances. On the left obtained from Ref. [34], for distances of $d/w(10^{-3})$: 2, 4, 6 and 10. On the right the calculated values using RF for $d/w(10^{-3})$: 4, 6, 10, 12, 14 and 16. The parameters are otherwise those stated in Fig. 5.37.

As was the case for the scattering cross section, the value rises to a maximum where after it decreases steadily as the gate to 2DEG distance increases. The reason for a maximum gate to 2DEG distance can be understood by considering the mechanism for a resonance in the first place. Resonances in the channel arises due to the excitation of plasma waves oscillations by the scattered fields from the gate. If the 2DEG layer is too far away from the gate the scattered fields are too weak for an effective excitation. In the other limit placing the 2DEG layer too close to the gate, will cause the perfectly conductive gate screen the plasma oscillation in the 2DEG [34]. By this reasoning the two first distances of Fig. 5.40(b) are still being effectively screened by the gate, whereas the gate to 2DEG of 6 nm are closer to the optimum distance, in contrast to Fig. 5.40(a), where the optimum distance is for 4 nm.

A noticeable difference for the total extinction cross section is that its value becomes negative. This is more easily understood by plotting $\sigma_{ext,r}$ and $\sigma_{ext,t}$ separately, Shown in right side of Fig. 5.42.

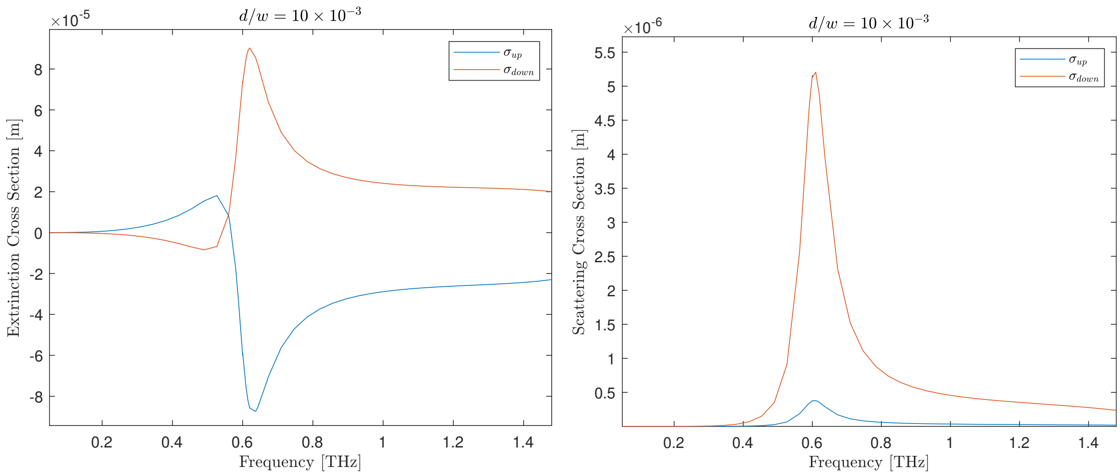

Figure 5.42. Extinction and scattering cross sections for a 1000 nm gate with gate to 2DEG distance of 10 nm.

A negative extinction cross means that the power, either reflected or transmitted depending on which medium is considered, is increased by the presence of the gate. This makes sense as the gate is assumed to be a perfect conductor and should therefore reflect any and all light incident on it. It is clear by comparing the two figure of Fig. 5.42 that the extinction cross section is much larger than the scattering. If this is to be believed it would mean that most of the extinction cross section comes from absorption either by the scatterer (into heat) or by being coupled into the guided modes namely the plasma waves.

As a last comparison of the method the extinction and scattering cross section has been calculated for the same relative gate to 2DEG distance, $d/w = 0.1$, and normalised by the gate length. According to Ref. [34] should the normalised scattering cross section vary proportional with w and should therefore decrease for smaller gate lengths. The calculated normalised values are shown in Fig. 5.43.

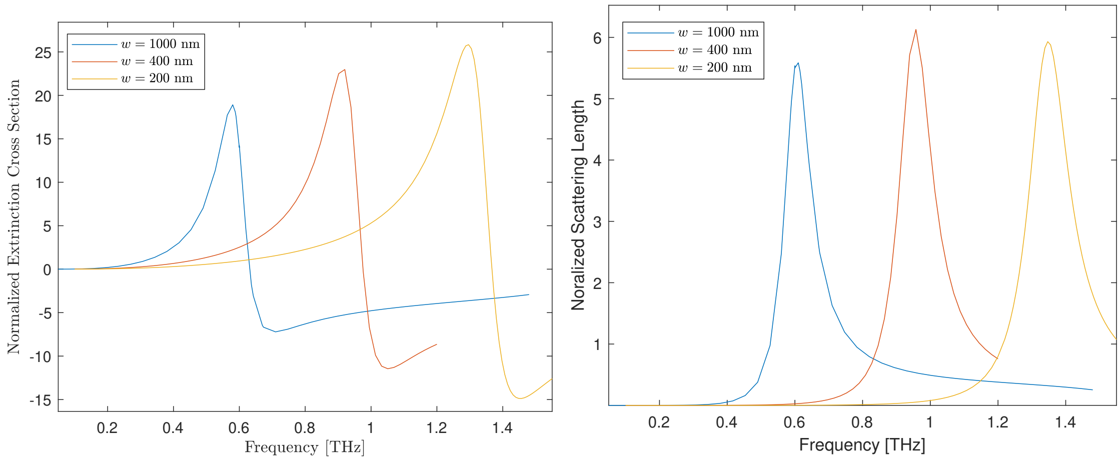

Figure 5.43. Total Extinction and scattering cross sections normalised by per unit gate length. for a 1000, 400 and 200 nm gate with a constant gate to 2DEG ratio of $d/w = 0.01$.

This tendency is not observed here, rather the normalised scattering cross section seems to fairly weakly depend on the gate length and the extinction cross section is seen to increase slightly for smaller gate lengths. This essentially means that a smaller gate should couple better into the guided modes relative to its length.

It would seem that the RF model are able to predict the same resonances as in [34], but the calculated values and which resonance is the maximum does differ. Nevertheless the method will be used on a AlGaN/GaN HEMT in order to examine if the same relation as observed here can be produced and if other correlations can be deduced.

5.5.2 AlGaN/GaN HEMT

For the AlGaN/GaN structure the far fields are plotted similarly to the InGaAs HEMT. First the far field for a 400 nm gate, for an incident wave with a frequency of $f = 1.44$ THz is shown in Fig. 5.44.

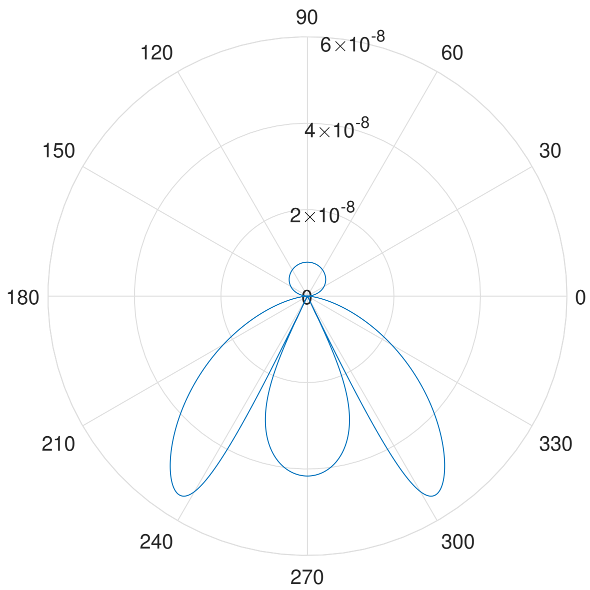

Figure 5.44. The far field plot for a gate length of 400 nm with a incoming wave with frequency of $f = 1.44$ THz.

The far field representation of the structure is very similar to the far field of InGaAs/GaAs. The angle of total reflection was calculated to be 270 ± 25.7. The lobe into the upper half-plane is significantly larger than for the InGaAs HEMT, however there is still significantly more power scattered into the structure. The power going directly up, compared to the power going down has a ratio of 530%, which is a lot lower than for the InGaAs structure. The reason for this bigger reflected power is likely due to the higher difference in the dielectric constants of the materials, compared to the InGaAs structure.

The optimal gate length for the standard structure with a frequency of $f = 1.44$ THz, is assumed to be around 400 nm as stated in Eq. (5.4), using the mode index and antenna theory. The extinction and scattering cross sections for different lengths of gates are shown in Fig. 5.45.

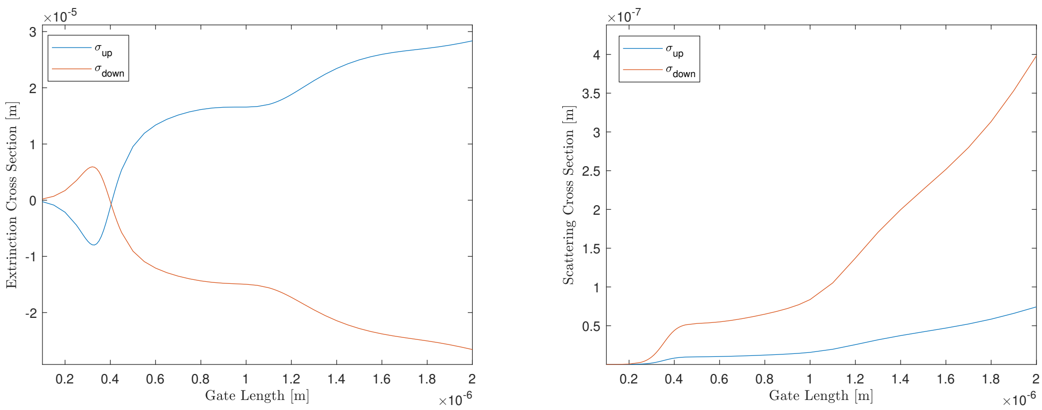

Figure 5.45. The extinction and scattering cross sections for different gate lengths. The frequency is 1.44 THz.

The scattering cross sections increases as the gate length increases, which makes sense as a longer gate can scatter more light. There is a significant increase at a gate length of around 320 nm, where after there is a slight plateu before the scattering cross section increases again. For the extinction cross sections, there is a peak around 320 nm.This differs from the expected peak from the mode index, found in Eq. (5.4). If dipole antenna theory holds for this case, the maximum current in the gate should correspond to the calculated gate length from the mode index. In order to analyse if this corresponds to the optimal gate length, the maximum of the current for the different gate lengths is shown in Fig. 5.46.

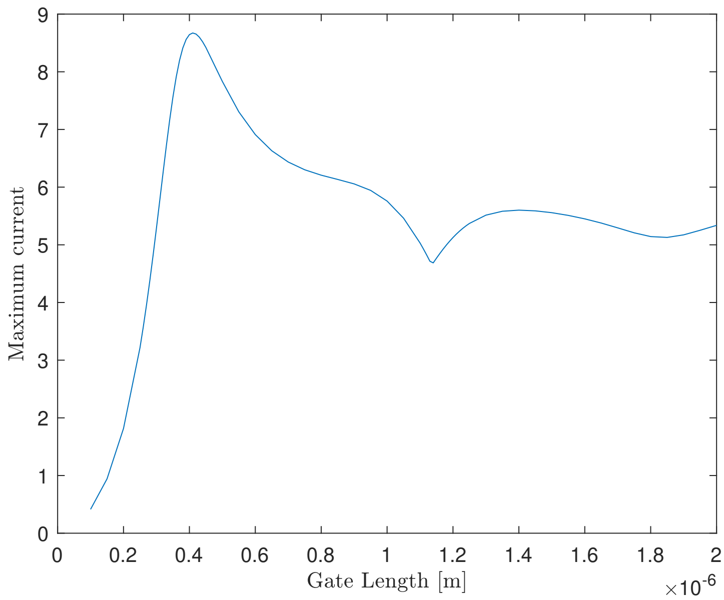

Figure 5.46. The maximum of the absolute value of the current for different gate lengths with parameters as in Fig. 5.45.

The maximum current corresponds to a gate length of around 410 nm, which corresponds perfectly with what was predicted from the mode index. It appears that higher order modes are not as easily excited, as no peaks can be seen for $w \approx 1200$ nm, which is where the next peak was expected from dipole antenna theory. However, there is a dip at around 1140 nm, which might be due to the resonance. In order to examine this, the current for a gate of 1140 nm is compared to gates of similar length in Fig. 5.47.

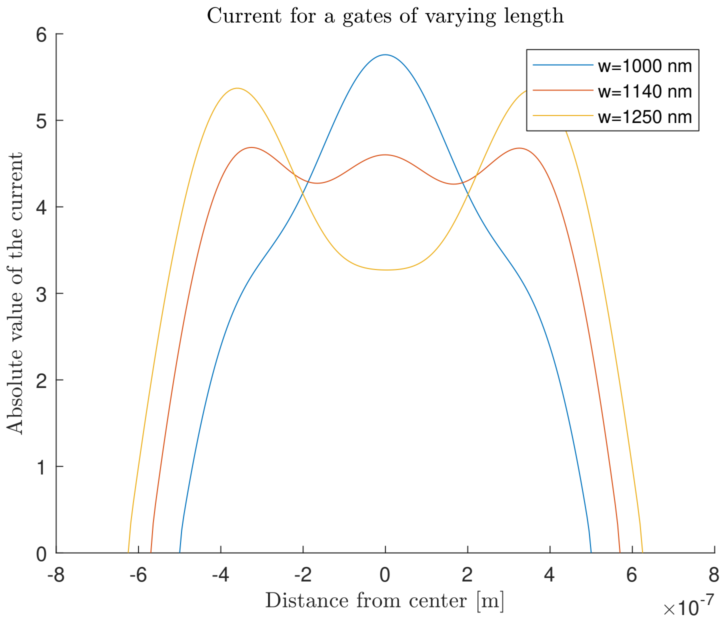

Figure 5.47. The absolute value of the current for a gate length of 1000 nm, 1140 nm and 1250 nm with parameters as in Fig. 5.45.

The current has three separate peaks for $w = 1140$ nm, which is consistent with a antenna with a length of $3/2\lambda$. This indicates that this length corresponds to another plasma mode. However there is a dip in the maximum current. This is due to the current being more evenly distributed in the gate, thus a lower maximum current is found. In order to compare the current in these different gates, the current per length of gate is found by

$$\frac{I_{tot}}{w} = \sum_n I_n \frac{\Delta}{w}, \tag{5.37}$$

where I_{tot} is the total current, and I_n is the current for segment n. Thus the current per length for $w = 1140$ nm is 7.8296, while the current for $w = 1250$ nm is 7.700 and the current per length for $w = 1000$ nm is 7.3351. Thus a gate at the dip in maximum current still have a significant amount of current. The current per length plot is shown in Fig. 5.48.

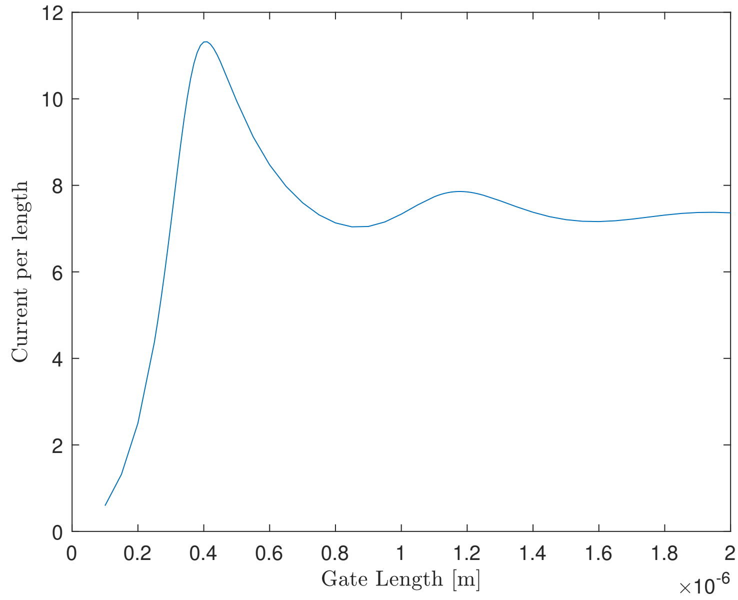

Figure 5.48. The current per length for gates of different lengths, with parameters as in Fig. 5.45.

There is a clear peak for a gate length of around 400 nm, which is to be expected, while a slight peak at the second resonance for a gate length of around 1200 nm can be seen.

The model have also been tested for an incoming wave with the frequency of 10 THz. The optical cross sections are shown in Fig. 5.49.

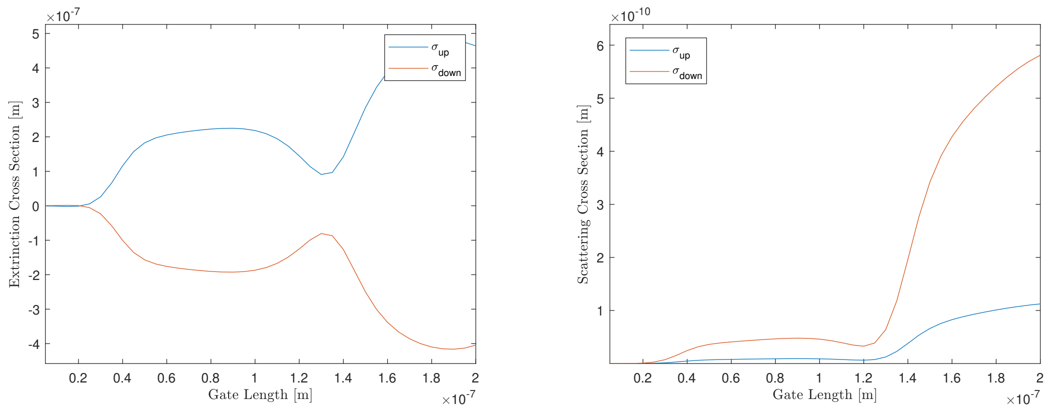

Figure 5.49. The extinction and scattering cross sections for different frequencies. The incoming wave has a frequency of 10 THz.

The scattering cross section for low gate lengths is very low, and after a gate length of around 120 nm, the scattering cross section increases dramatically. The current is shown in Fig. 5.50.

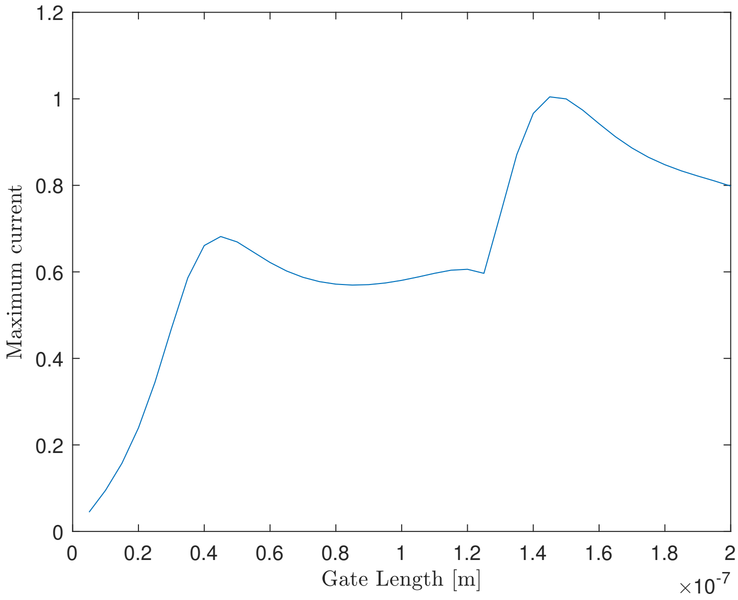

Figure 5.50. The maximum current for gates of different lengths, with parameters as in Fig. 5.49.

The current has a peak for a gate length of around 50 nm, which corresponds with the gate length for the mode index, obtained in Sec. 5.1. There is also a peak for a gate length of around 150 nm, which would correspond to the second resonant mode for the structure. This peak is even higher than the fundamental mode, which might be due to the gate being so small at 50 nm, that the response is severely limited. The current for the gate lengths responding to the resonant modes is shown in Fig. 5.51.

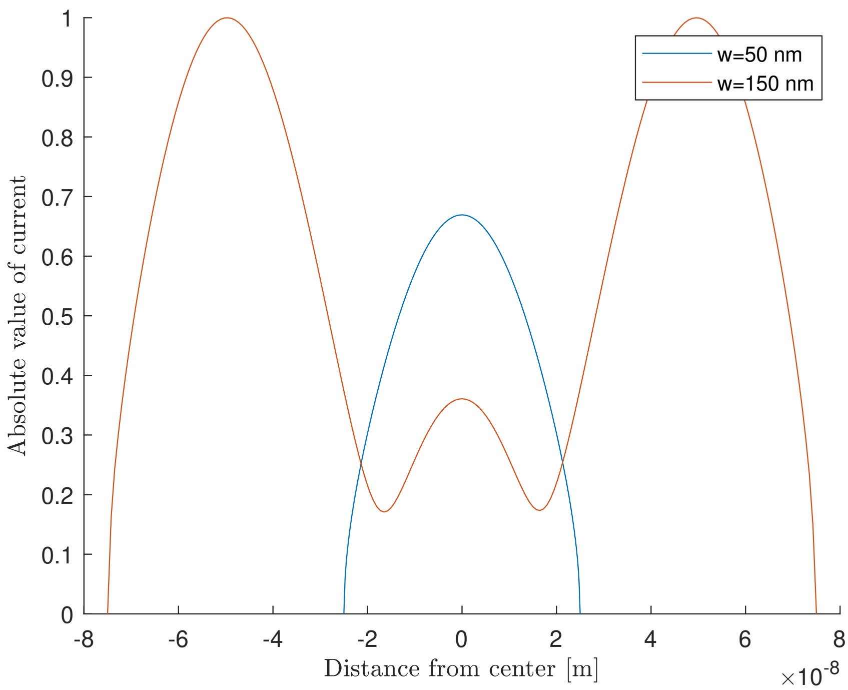

Figure 5.51. The current per length for gates of different lengths, with parameters as in Fig. 5.49.

According to the shapes, these lengths corresponds to the two first resonances.

In order to analyse the frequency dependency, the cross sections for a gate of 400 nm with varying frequency is shown in Fig.5.52.

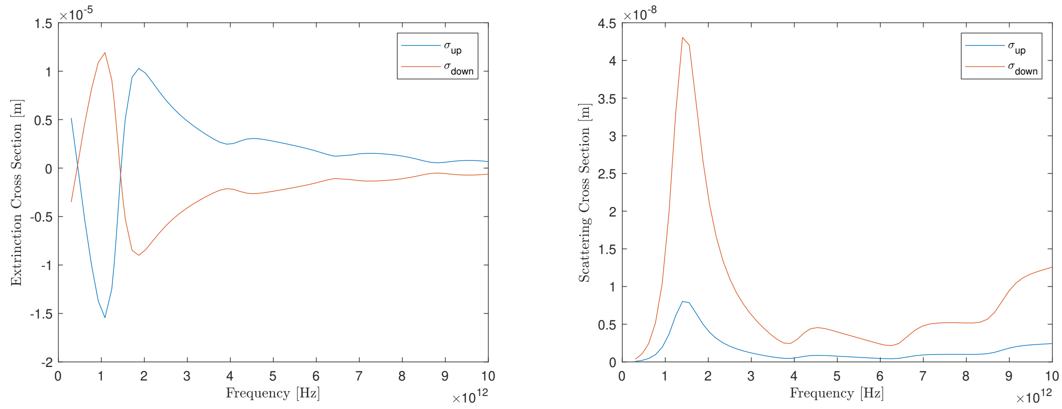

Figure 5.52. The extinction and scattering cross sections for different frequencies. The gate length is 400 nm.

There appear to be periodic dips in both the scattering cross section and the extinction cross section. There appears to be a large peak at around 1.4 THz for the scattering cross sections, which corresponds to the expected plasma resonance. In order to see the effect of these dips, the current for the same setup is shown in Fig. 5.53.

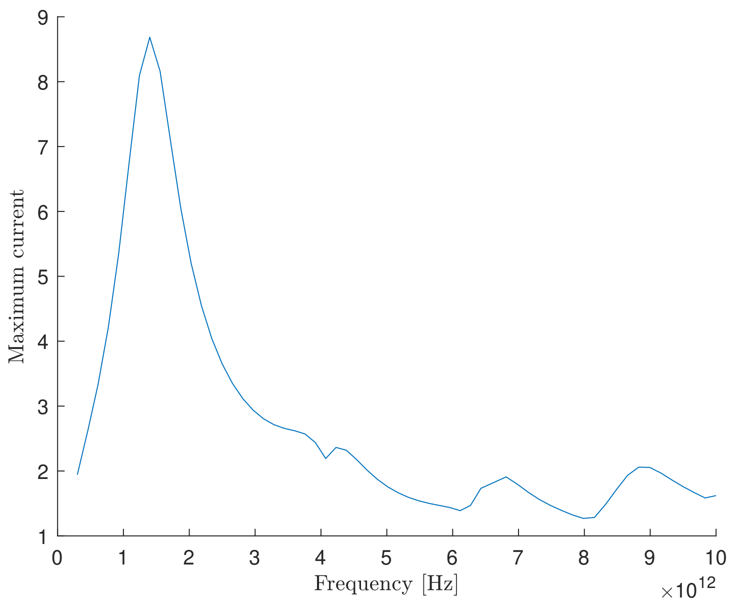

Figure 5.53. The maximum of the absolute value of the current for frequencies with parameters as in Fig. 5.52.

It is clear there is a peak around 1.4 THz, which corresponds to the previous result. The dips in the scattering cross sections seem to correspond to dips in the current for higher frequencies.

The impact of changing distance from the gate to the 2DEG is tested in Fig. 5.54.

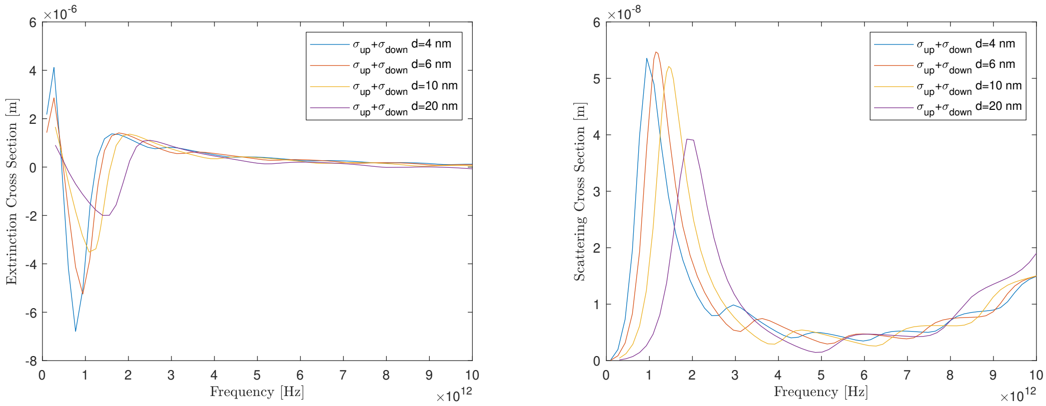

Figure 5.54. The extinction and scattering cross sections for different frequencies. The distance to the 2DEG varies.

It can be seen that as the distance to the 2DEG increases, the peak of the scattering cross section shifts towards higher frequencies. The shapes of the cross sections seem consistent while shifting. The highest scattering is found for $d = 6$ nm, while the scattering decreases if d is either higher or lower. As mentioned earlier, there appear to be an optimal distance for scattering. This optimal scattering length does not correspond to an optimal current, which can be seen in Fig. 5.55.

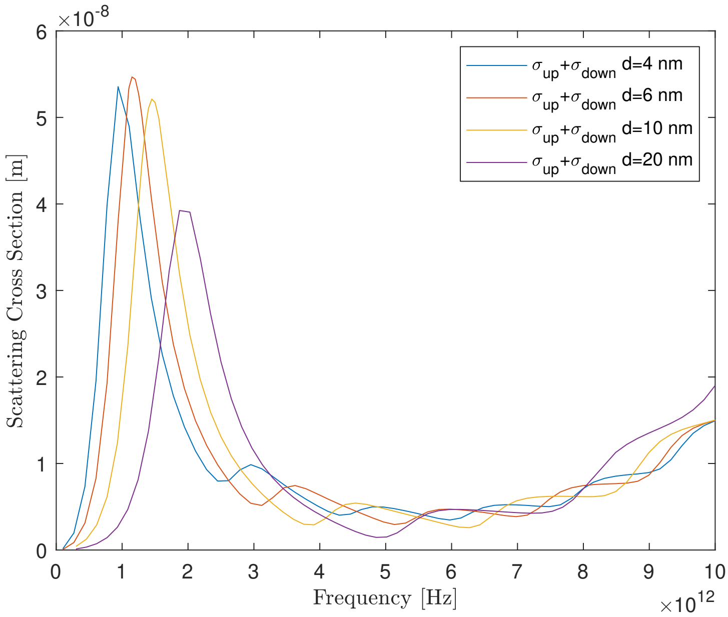

Figure 5.55. The maximum of the absolute value of the current for frequencies and different distances to the 2DEG.

The current decreases as the distance to the 2DEG increases. There are oscillations after the peak, as for the scattering cross section. The shorter the distance to the 2DEG, the better the gate and the 2DEG can interact, which could cause this higher current.

In order to compare how this frequency dependency changes according to the structure, the cross sections for a gate with length 50 nm for different frequencies is shown in Fig. 5.56.

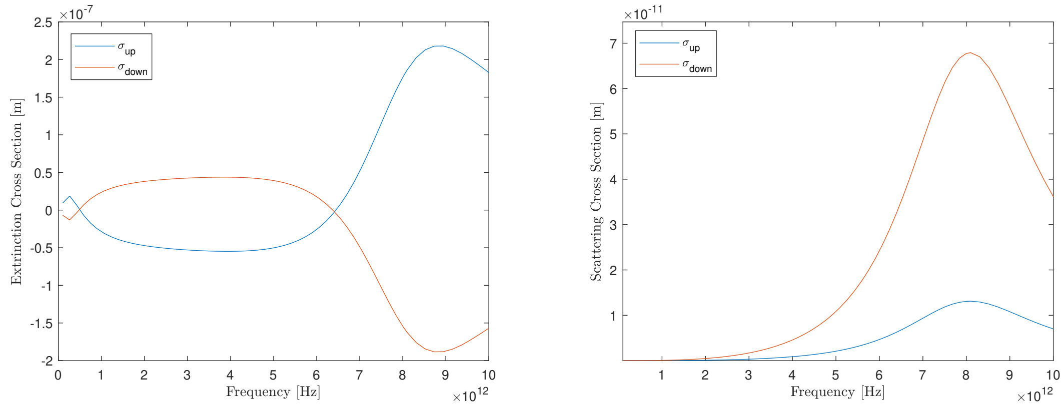

Figure 5.56. The extinction and scattering cross sections for different frequencies. The gate length is 50 nm.

There is a clear peak for both the extinction cross sections and the scattering cross sections, however the peaks are shifted a bit in respect with each other. For the scattering cross sections the peak is around 8.1 THz, where for the extinction cross section the peak is around 8.8 THz. This is a much higher frequency than was found for the 400 nm gate. This is likely due to the gate being shorter, thus shorter wavelengths are necessary to generate a similar response. The currents dependency on the frequency for the same gate is shown in Fig. 5.57.

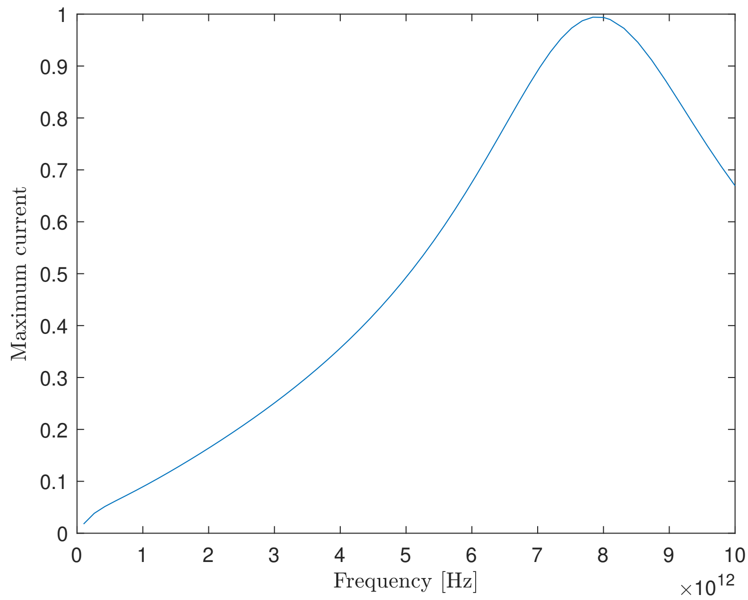

Figure 5.57. The maximum of the absolute value of the current for frequencies with parameters as in Fig. 5.56.

The current exhibits a peak at the same frequency the scattering cross sections peak, at around 8.1 THz. In order to analyse if the results for the gate length of 400 nm are consistent, the effects of increasing the distance to the 2DEG is shown in Fig. 5.58.

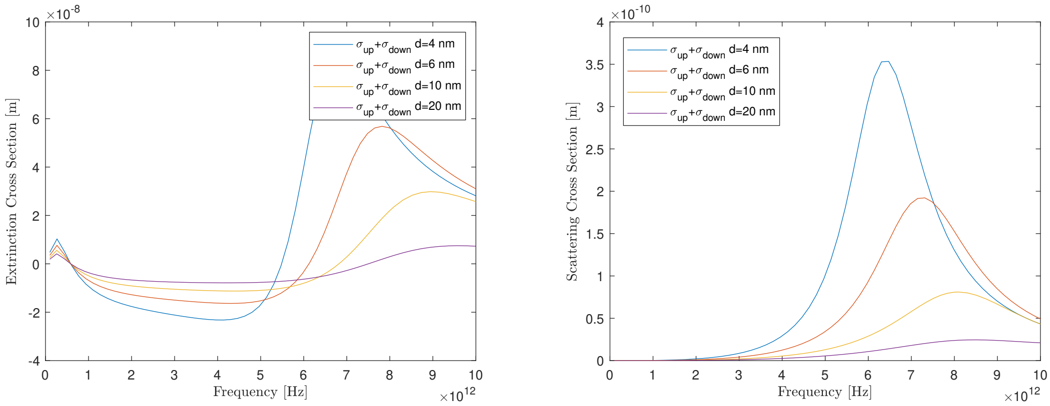

Figure 5.58. The extinction and scattering cross sections for different frequencies. The distance to the 2DEG varies.

It seems the peak shifts and decreases as the distance to the 2DEG increases, which corresponds well with the results for the 400 nm gate. The optimal distance for scattering is not found for the values analysed. It is assumed to be $d < 4$ nm. The peaks are also much broader than they were for the 400 nm gate. This broadening can be explained by the fact that the plots are shown in frequency. As $\lambda \propto 1/f$ the values for low frequencies are more closely spaced in wavelength compared to the higher wavelengths. If the peaks are comparably broad in wavelengths, when shown in frequencies, the shift in packing, will broaden the higher frequency peaks.

Final Words 6

6.1 Discussion

For the mode indices the two models predicted significantly different guided modes. The model assuming the 2DEG having a finite thickness, did not converge towards the higher values of the infinitely thin model as was expected, but converged towards 0. This might be explained due to the method with a finite thickness not incorporating surface currents at the boundaries, and instead assuming that the H field is simply continuous across these boundaries. The mode indices found using the infinitely thin model, did correspond with the results from the numerical models. The overall tendencies for the two methods were the similar when changing the parameters of the 2DEG, though with the infinitely thin model still predicting many times larger mode indices. When changing the sheet carrier density, both models predicted that the mode indices decreased with increasing sheet carrier density, however, by comparing this to the plasma dispersion relation obtained in Sec. 3.2, a positive proportionality was expected. This discrepancy cannot be explained, as it occurs for both models. Initially the current was calculated by the PPS method outlined in [34], however this method had in general numerical problems especially for a large number of segments and only converged when δ was set as a constant value. The convergence was still rather poor and occurred for a small span of constant δs. This is likely due to the model being numerically unstable as $\delta \to 0$ and poorly weighted for larger δ. The PPS method was in general very numerically unstable and the integrals had a limited precision as they could only be done to a certain limit before the integrand was set to zero.To combat this problem, the RF method was established and for this model the integrals showed greater numerical stability, but the convergence was still rather poor as it was still depended on the displacement δ being in a specific range and then still required a high number of segments for convergence.

This method was improved to incorporate second order basis functions, which gave a very stable model. The model had to be handled carefully around $k_x = 0$, however this was accomplished by using a second order Taylor expansion around $k_x = 0$. The model was extremely consistent with regard to this area, as long as the Taylor expansion was done for values < 1. The model did, however, still show problems at the corners of the gate for low values of $delta$, which again likely is due to the distance between the elements becoming too small a part of the distance to the measuring point. The model had a very stable convergence with respect to segmentation for a set δ however the model did not exhibit convergence with respect to δ, and the maximum value of the current decreased as δ decreased. This decrease, however, was vanishingly small, thus the model might be usable, as long as δ was a set value. When computing more complex patterns, the model still needed a relatively high number of segments in order to properly converge. For around 100 segments, the model seemed decently converged for all tried cases. The value of the current in the different models were not comparable, with the point matching being in the order of 10^{-3} and second order being in low integers. This is likely due to the current simply being poorly described by the point matching models.

Far field radiation patterns as well as scattering and extinction cross section was at first calculated for an InGaAs HEMT in order to compare the results with those of Ref. [34]. The radiation pattern obtained was in somewhat disagreement in Ref. [34], shown in Fig. 5.37, especially the scattered lobe into the ambient medium, differed by its general shape, being more triangular, whereas all the far field calculated in this project are circular. The lobes into the substrate were more in agreement, but due to questionable parameters used in the reference, this could not be reproduced exactly. To obtain the larger lobe in the ambient medium, a bigger difference in dielectric constants between the substrate and barrier layer is assumed to be necessary, which was also showed to have the wanted effect, but not to the degree required for reproducing the far field pattern. In terms of the calculated scattering and extinction cross section many of the same correlation was found in this project as those described in Ref. [34]. However, a general tendency was that optimum resonance peaks were shifted towards higher frequencies. This difference might be due to the uncertainty in the parameters used for the InGaAS HEMT. However, the trends for the resonance frequencies for both the extinction and scattering cross sections are consistent for a gate length of 1000 nm and varying distances to the 2DEG. The normalised extinction and scattering cross section has also been analysed and here it was found that the scattering cross section are in largely independent of the gate width, whereas the extinction cross section seem to be weakly, increasing with the decrease in gate length, with a constant gate to 2DEG ratio of $d/w = 0.01$, shown in Fig. 5.43. This is not in agreement with what was found in Ref. [34], where the absorption was also found to be weakly varying with the gate length, but is instead decreasing with shorter gate lengths. And the scattering cross section is seen to have an almost proportional relationship with the gate length.

The far field and scattering plots for the AlGaN HEMT was performed and showed the same tendencies as the InGaAs HEMT. The far field plot showed an even larger lobe in the ambient medium, as was to be expected from the much larger difference in the dielectric constants of the insulator and the substrate. The maximum current corresponded extremely well to the expected resonance peaks from the mode indices for the infinitely thin 2DEG. As both build on the assumption that the 2DEG is infinitely thin, this was to be expected.

Negative extinction cross sections were found for several cases. A negative $\sigma_{ext,up}$ corresponds to a larger amount of field being reflected upwards for the structure with the scatterer, compared to the structure without the scatterer. The scatterer was assumed to be a perfect conductor, thus should reflect all incident light. The extinction cross sections were all much higher values than the scattering cross sections, indicating a large amount of absorption, which could be partly due to guided modes.

6.2 Conclusion

In this project the effects of the 2DEG in a HEMT structure has been used in order to determine the terahertz response. A method for examining the mode indices for the different structures were established. Two models were established, one where the 2DEG had a finite thickness and a model where the 2DEG was assumed to be infinitely thin. The first model predicted guided modes with a mode index of around 30 to 50 in the real part. When letting the finite thickness in the model $d_{2DEG} \rightarrow 0$, the model predicted $n_m \rightarrow 0$. This conflicted with the results for the model with the infinitely thin 2DEG, which predicted guided modes with a mode index of $n_m \approx 250$ for similar cases. The difference is believed to originate in the handling of the boundary conditions for the finite thickness model.

Theory was established using the electromagnetic boundary conditions, in order to establish models comparable to the ones presented in Ref. [34]. First the model was established using point matching and included a small distance, δ, in order to make the model numerically stable. The model depended greatly on δ, where for only for a small interval of values of δ the model seemed to converge. The model also proved numerically unstable when extending the integrals, giving undefined values. The model was rewritten to be more numerically stable and more consistent results were obtained, however the model still depended greatly on δ. For too low values of δ, the model never converges, and for too high values of δ, the model has problems fully determining the current at the corners of the gate. In order to improve the convergence of the model, second order basis functions were introduced to help the model converge. For the second order basis functions, the model still showed a dependency on δ and never converged with respect to δ, however the dependency on δ was very small, thus the model might still be valid as long as δ is kept small. For too high values of δ the model exhibits the same problem at the corners of the gate as point matching. However the model converged with respect to segmentation if δ was kept constant. The model gave consistent shapes as long as δ was kept on an order smaller than the segmentation lengths.

Due to the second order basis functions being singular around $k_x = 0$ a second order Taylor expansion was used In order to avoid numerical problems for $k_x = 0$, this included a distance around $k_x = 0$, δ_k for which the approximation could be justified. The model was found to be very consistent for different values of δ_k, as long as $\delta_k \Delta < 1$. The model converged very well with respect to segmentation, even for relatively long gates and small wavelengths. For reasonably convergent values for the different structures, $N = 100$ was chosen as for all the tested structures, $N = 100$ was decently close to the convergent shape and the differences in values was negligible.

The far field was established using the model with the second order basis functions. It was attempted to replicate the patterns shown in Ref. [34]. The far field patterns created using the established model differed from the patterns from the reference, especially the field scattered into the ambient medium was not possible to be reproduced. In general the far field patterns show a high amount of power being sent into the structure, with only a small part leaving the structure. This can be increased by having a larger contrast in substrate and barrier dielectric due to a increasing the reflection coefficient for this interface. The far field patterns are relatively consistent for different gate lengths and wavelengths, due to the gate being very small compared to the wavelength, therefore the differences barely matter for the far field radiation pattern. It was also attempted to reproduce the optical cross sections presented in Ref. [34]. The cross section for a gate length of 1000 nm as function of gate to 2DEG, was found to be in somewhat agreement with the Ref. [34], with the same tendency observed but with a shift to higher frequencies. The correlation between the normalised cross sections found in Ref. [34] and different gate length was not observed in this project.

Radiation patterns and optical cross sections were tested for an AlGaN HEMT, where peaks corresponding to the mode indices were found, using antenna theory. Especially the fundamental resonant modes were significant, exhibiting huge peaks in the scattering cross sections and in the currents. The dependency on the distance between the gate and the 2DEG were also tested, and as the distance increased, the resonance shifted to a higher frequency and the effect of the resonance produced greatly reduced peaks. The tendencies shown in the InGaAs structure were also seen in the AlGaN structure.

Bibliography

[1] Jiandong Sun. *Field-effect Self-mixing Terahertz Detectors (Springer Theses)*. Springer, 2016.

[2] Christian Jansen, Steffen Wietzke, Ole Peters, Maik Scheller, Nico Vieweg, Mohammed Salhi, Norman Krumbholz, Christian Jördens, Thomas Hochrein, and Martin Koch. Terahertz imaging: applications and perspectives. *Appl. Opt.*, 49(19): E48–E57, Jul 2010.

[3] Ashish Y. Pawar, Deepak D. Sonawane, Kiran B. Erande, and Deelip V. Derle. Terahertz technology and its applications. *Drug Invention Today*, 5(2):157 – 163, 2013.

[4] Calvin Yu, Shuting Fan, Yiwen Sun, and Emma Pickwell-MacPherson. The potential of terahertz imaging for cancer diagnosis: A review of investigations to date. *Quantitative Imaging in Medicine and Surgery*, 2(1), 2012.

[5] Michael Dyakonov and Michael Shur. Shallow water analogy for a ballistic field effect transistor: New mechanism of plasma wave generation by dc current. *Phys. Rev. Lett.*, 71:2465–2468, Oct 1993.

[6] D. Čibiraitė, M. Bauer, A. Rämer, S. Chevtchenko, A. Lisauskas·, J. Matukas, V. Krozer·, W. Heinrich, and H. G. Roskos. Enhanced performance of algan/gan hemt-based thz detectors at room temperature and at low temperature. In *2017 42nd International Conference on Infrared, Millimeter, and Terahertz Waves (IRMMW-THz)*, pages 1–2, Aug 2017.

[7] V. Jakštas, I. Grigelionis, V. Janonis, G. Valušis, I. Kašalynas, G. Seniutinas, S. Juodkazis, P. Prystawko, and M. Leszczyński. Electrically driven terahertz radiation of 2deg plasmons in algan/gan structures at 110 k temperature. *Applied Physics Letters*, 110(20):202101, 2017.

[8] David J. Griffiths. *Introduction to Electrodynamics (3rd Edition)*. Prentice Hall, 1999.

[9] Walton C. Gibson. *The Method of Moments in Electromagnetics*. Chapman and Hall/CRC, 2007.

[10] Thomas M. Søndergaard. *Green's Function Integral Equation Methods in Nano-Optics*. CRC Press, 2019.

[11] Michael Shur. *Physics of Semiconductor Devices*. Pearson, 1990.

[12] Donald A. Neamen. *Semiconductor physics and devices : basic principles*. McGraw-Hill, 2012.

[13] Julius Edgar Lilienfeld. Method and apparatus for controlling electric currents, US Patent 1745175, oct 1926. URL https://patents.google.com/patent/US1745175A/en.

[14] Fikru Adamu-Lema. *Scaling and Intrinsic Parameter Flactuations in Nano-CMOS Devices*. PhD thesis, University of Glasgow, 2005.

[15] Douglas Natelson. *Nanostructures and nanotechnology*. Cambridge University Press, 2015.

[16] Robert W Keyes. Physical limits of silicon transistors and circuits. *Reports on Progress in Physics*, 68(12):2701–2746, sep 2005.

[17] Takashi Mimura, Satoshi Hiyamizu, Toshio Fujii, and Kazuo Nanbu. A new field-effect transistor with selectively doped GaAs/n-AlxGa1-xAs heterojunctions. *Japanese Journal of Applied Physics*, 19(5):L225–L227, may 1980.

[18] R. L. Anderson. Germanium-gallium arsenide heterojunctions [letter to the editor]. *IBM Journal of Research and Development*, 4(3):283–287, July 1960.

[19] O. Ambacher, B. Foutz, J. Smart, J. R. Shealy, N. G. Weimann, K. Chu, M. Murphy, A. J. Sierakowski, W. J. Schaff, L. F. Eastman, R. Dimitrov, A. Mitchell, and M. Stutzmann. Two dimensional electron gases induced by spontaneous and piezoelectric polarization in undoped and doped algan/gan heterostructures. *Journal of Applied Physics*, 87(1):334–344, 2000.

[20] Jesus A. del Alamo. The high-electron mobility transistor at 30: Impressive accomplishments and exciting prospects., 5 2011. 2011 International Conference on Compound Semiconductor Manufacturing Technology, Indian Wells, California.

[21] Takashi Mimura. Development of high electron mobility transistor. *Japanese Journal of Applied Physics*, 44(12):8263–8268, dec 2005.

[22] M. Dyakonov and M. S. Shur. Plasma wave electronics for terahertz applications. In R. E. Miles, P. Harrison, and D. Lippens, editors, *Terahertz Sources and Systems*, pages 187–207. Springer Netherlands, 2001.

[23] Michael S. Shur and Victor Ryzhii. Plasma wave electronics. *International Journal of High Speed Electronics and Systems*, 13(02):575–600, 2003.

[24] S. Das Sarma and E. H. Hwang. Collective modes of the massless dirac plasma. *Phys. Rev. Lett.*, 102:206412, May 2009.

[25] Hartmut Haug and Stephan W Koch. *Quantum Theory of the Optical and Electronic Properties of Semiconductors*. WORLD SCIENTIFIC, 4th edition, 2004.

[26] Thomas Garm Pedersen. *ELECTRIC, OPTICAL & MAGNETIC PROPERTIES OF NANOSTRUCTURES*. Aalborg University, Aalborg, 2019.

[27] M. I. Dyakonov. Boundary instability of a two-dimensional electron fluid. *Semiconductors*, 42(8):984–988, 2008.

[28] M. Dyakonov and M. Shur. Detection, mixing, and frequency multiplication of terahertz radiation by two-dimensional electronic fluid. *IEEE Transactions on Electron Devices*, 43(3):380–387, March 1996.

[29] M. I. Dyakonov and M. S. Shur. Two dimensional electronic flute. *Applied Physics Letters*, 67(8):1137–1139, 1995.

[30] O. Sydoruk, R. R. A. Syms, and L. Solymar. Plasma oscillations and terahertz instability in field-effect transistors with corbino geometry. *Applied Physics Letters*, 97(26):263504, 2010.

[31] L. D. Landau and E. M. Lifshitz. *Fluid mechanics.* Elsevier/Butterworth-Heinemann, 1987.

[32] J. Y. Park, S. Kim, S. Hong, and K. R. Kim. Physical analysis and design of resonant plasma-wave transistors for terahertz emitters. *IEEE Transactions on Terahertz Science and Technology*, 5(2):244–250, March 2015.

[33] M. I. Dyakonov and M. S. Shur. Plasma wave electronics: novel terahertz devices using two dimensional electron fluid. *IEEE Transactions on Electron Devices*, 43 (10):1640–1645, Oct 1996.

[34] V. V. Popov, O. V. Polischuk, and M. S. Shur. Resonant excitation of plasma oscillations in a partially gated two-dimensional electron layer. *Journal of Applied Physics*, 98(3):033510, 2005.

[35] Morgan Hibberd, V. Frey, Ben Spencer, Peter Mitchell, Philip Dawson, M. J. Kappers, R. A. Oliver, C. J. Humphreys, and Darren Graham. Dielectric response of wurtzite gallium nitride in the terahertz frequency range. *Solid State Communications*, 247:68–71, dec 2016.

[36] A. S. Barker and M. Ilegems. Infrared lattice vibrations and free-electron dispersion in gan. *Phys. Rev. B*, 7:743–750, Jan 1973.

[37] A. V. Muravjov, D. B. Veksler, V. V. Popov, O. V. Polischuk, N. Pala, X. Hu, R. Gaska, H. Saxena, R. E. Peale, and M. S. Shur. Temperature dependence of plasmonic terahertz absorption in grating-gate gallium-nitride transistor structures. *Applied Physics Letters*, 96(4):042105, 2010.

[38] Viacheslav (Vyacheslav Popov, G M. Tsymbalov, T V. Teperik, Denis Fateev, and Michael Shur. Terahertz excitation of the higher-order plasmon modes in field-effect transistor arrays with common and separate two-dimensional electron channels. *Bulletin of the Russian Academy of Sciences Physics*, 71:89–92, 01 2007.

[39] Constantine A. Balanis. *Antenna Theory: Analysis and Design.* Wiley, 2016.

[40] Tsuneya Ando, Alan B. Fowler, and Frank Stern. Electronic properties of two-dimensional systems. *Rev. Mod. Phys.*, 54:437–672, Apr 1982.

[41] Charles. Kittel. *Introduction to solid state physics.* Wiley, 2004.

[42] John H. Davies. *The Physics of Low-dimensional Semiconductors: An Introduction.* Cambridge University Press, 1997.

[43] Neil W. Ashcroft and N. David Mermin. *Solid State Physics.* Cengage Learning, 1976.

Appendix A

A.1 Energy Levels of 2DEG

An electron in a bulk semiconductor is subject to a three-dimensional periodic potential due to the lattice. Furthermore, the potential due to the quantum well can be described by a one-dimensional confining potential $V(z)$. The electron can be described by the effective-mass theorem [40]

$$\left[\frac{-\hbar^2}{2}\sum_{i,j}\left(\frac{1}{m_{ij}}\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j}\right)+\mathcal{H}'\right]\Psi=E\Psi, \tag{A.1}$$

where $\mathcal{H}'$ is a potential and m_{ij} is the effective mass tensor. The energy near the band edge can be obtained from the central equation as [11] [41]

$$E(\mathbf{k})=E(\mathbf{k}_0)+\frac{\hbar^2}{2}\sum_{i,j}\left(\frac{1}{m_{ij}}k_ik_j\right), \tag{A.2}$$

where $\mathbf{k}_0$ is the wave vector at the band edge. The effective mass comes as a consequence of the electron moving in a periodic potential and can be seen as the mass the electron seems to have when it is affected by forces. It can be obtained by differentiating the group velocity given as $\mathbf{v}_g=\frac{1}{\hbar}\nabla_{\mathbf{k}}E(\mathbf{k})$, with respect to time

$$\frac{\mathrm{d}\mathbf{v}_g}{\mathrm{d}t}=\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{\hbar}\nabla_{\mathbf{k}}E(\mathbf{k})=\frac{1}{\hbar}\nabla_{\mathbf{k}}\frac{\mathrm{d}E(\mathbf{k})}{\mathrm{d}t}. \tag{A.3}$$

The time derivative of the energy can be rewritten using the chain rule, yielding

$$\frac{\mathrm{d}E(\mathbf{k}(t))}{\mathrm{d}t}=\nabla_{\mathbf{k}}E\cdot\frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t}, \tag{A.4}$$

which by using that the force due to the lattice is given by $\mathbf{F}=\hbar\mathrm{d}\mathbf{k}/\mathrm{d}t$ Eq. (A.3) can be written as

$$\frac{\mathrm{d}\mathbf{v}_g}{\mathrm{d}t}=\frac{1}{\hbar^2}\nabla_{\mathbf{k}}\left(\mathbf{F}\cdot\nabla_{\mathbf{k}}E\right). \tag{A.5}$$

Taking the ith element from Eq. (A.5) yields

$$\frac{\mathrm{d}v_{g,i}}{\mathrm{d}t}=\left(\frac{1}{\hbar^2}\sum_j\frac{\mathrm{d}^2}{\mathrm{d}k_i\mathrm{d}k_j}E(\mathbf{k})\right)F_j, \tag{A.6}$$

where the inverse of the bracket can be identified as a mass, i.e. the effective mass m_{ij}. If $1/m_{ij}$ is a diagonal matrix. Eq. (A.1) will have a solution of the form

$$\psi_{n,k_x,k_y}=e^{ik_xx}e^{ik_yy}f_n(z), \tag{A.7}$$

where $f_n(z)$ is a solution to

$$-\frac{\hbar^2}{2m_{zz}}\frac{\mathrm{d}^2 f_n}{\mathrm{d}z^2} + V(z)f_n = E_{n,z}f_n. \tag{A.8}$$

The total energy will thus be

$$E_n(k_x, ky) = E_{n,z} + \frac{\hbar k_x^2}{2m_{xx}} + \frac{\hbar k_y^2}{2m_{yy}}. \tag{A.9}$$

The energy is essentially the same as for zone boundary in a periodic potential, but with the constant energy term (either E_C or E_V for conduction or valence band) quantised in n subbands.

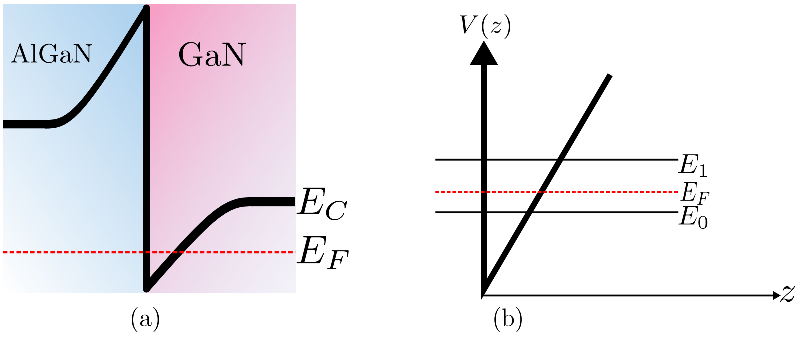

Figure A.1. (a) Conduction band at the AlGaN/GaN interface; (b) The well can approximated as a triangular quantum well with discrete energy levels.

Considering the conduction band near the abrupt junction interface, shown in Fig. A.1(a), the potential well can be approximated by a triangular potential well, as shown in Fig. A.1(b), which has the following boundary conditions for the potential

$$V(z) = \infty \quad \text{for} \quad z \leq 0, \tag{A.10}$$

$$V(z) = eEz \quad \text{for} \quad z > 0. \tag{A.11}$$

Inserting this potential in Eq. (A.8) yields

$$\frac{\mathrm{d}^2 f_n}{\mathrm{d}z^2} + \frac{2m_{zz}}{\hbar^2}(E_{n,z} - eEz)f_n = 0. \tag{A.12}$$

By introducing a new variable as $\zeta = z(\frac{2meE}{\hbar^2})^{1/3} - \frac{2mE_{n,z}}{\hbar^2}(\frac{\hbar^2}{2eE})^{2/3}$, Eq. (A.12) can be written on the form

$$\frac{\mathrm{d}^2 f_n(\zeta)}{\mathrm{d}\zeta^2} - \zeta f_n(\zeta) = 0. \tag{A.13}$$

This can readily be confirmed by noting that from the chain rule:

$$\frac{\mathrm{d}^2 f_n(z)}{\mathrm{d}z^2} = \frac{\mathrm{d}^2 f_n(\zeta)}{\mathrm{d}\zeta^2}\left(\frac{\mathrm{d}^2\zeta}{\mathrm{d}z^2}\right)^2 = \frac{\mathrm{d}^2 f_n(\zeta)}{\mathrm{d}\zeta^2}\left(\frac{2meE}{\hbar^2}\right)^{2/3}. \tag{A.14}$$

The form of Eq. (A.13) is known as the Airy equation, which has solutions called Airy functions, Ai(ζ) and Bi(ζ), they are shown in Fig. A.2.

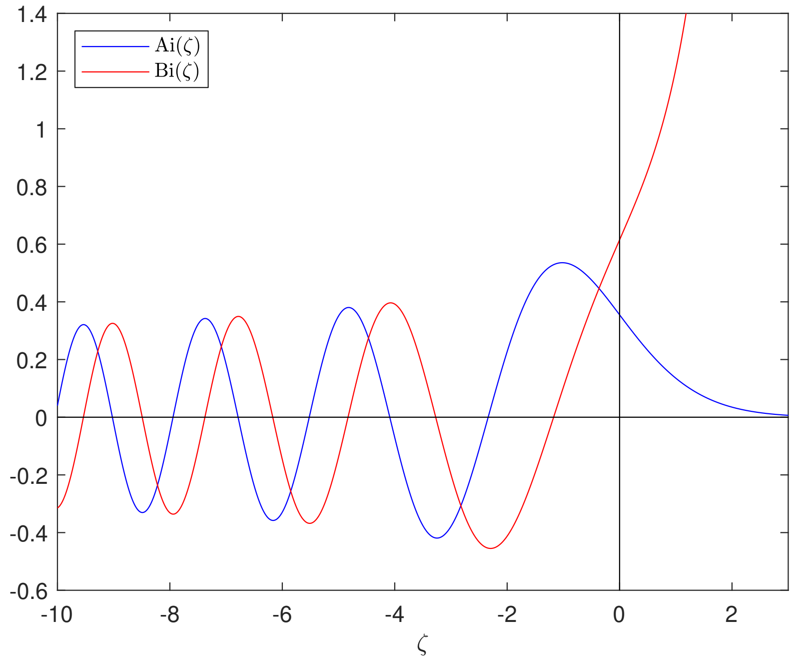

Figure A.2. The two Airy functions $\mathrm{Ai}(\zeta)$ and $\mathrm{Bi}(\zeta)$

The wave function is required to be well behaved for $z \to \infty$, which means the $\mathrm{Bi}(\zeta)$ can be rejected as it diverges for increasing ζ, which is also for increasing z. From the boundary condition at $z = 0$ it is required that $f_n(z = 0) = f_n(\zeta = -\frac{2mE_{n,z}}{\hbar^2}(\frac{\hbar^2}{2eE})^{2/3}) = 0$. For negative values of ζ $\mathrm{Ai}(\zeta)$ is a periodic function oscillating around the axis, as such there is an infinite number of values for ζ where $\mathrm{Ai}(\zeta) = 0$ which is denoted by $-c_n$, where the minus is to remove the sign. Therefore to ensure that the wave function vanish at $z = 0$, $c_n = \frac{2mE_{n,z}}{\hbar^2}(\frac{\hbar^2}{2eE})^{2/3}$ and the allowed energy does becomes

$$E_{n,z} = c_n \left(\frac{\hbar^2 (eE)^2}{2m} \right)^{1/3}, \tag{A.15}$$

where c_n can be approximated with the following formula $c_n \simeq [\frac{3}{2}\pi(n - \frac{1}{3})]^{2/3}$ [42].

A.2 Linear Response Theory

This section is based on reference [26].

A system under the effect of an external field, which varies with time, what is called *time-dependant perturbation theory*. In order to simplify only the linear changes to the wave function are analysed. Starting from a unperturbed system, the system can be described through the time in-dependant Hamiltonian, $\hat{H}_0$ and assuming the solutions to the Schrödinger equations are known, such that

$$\hat{H}_0 \phi_n = E_n \phi_n, \tag{A.16}$$

where ϕ is the stationary wave function. The time-dependence of the perturbation is assumed to be harmonic, i. e. only depending on a single frequency. The Hamiltonian changes due to the perturbations to $\hat{H}_0 + \frac{1}{2}\hat{H}_1 e^{-i\omega t} + \frac{1}{2}\hat{H}_1^\dagger e^{i\omega t}$, where $\hat{H}_1^\dagger$ is the Hermitian

Conjugate. In order to describe this perturbation, the time-dependant Schrödinger equation is used, so that

$$i\hbar\frac{\partial\psi}{\partial t} = \left[\hat{H}_0 + \frac{1}{2}\hat{H}_1 e^{-i\omega t} + \frac{1}{2}\hat{H}_1^\dagger e^{i\omega t}\right]\psi. \tag{A.17}$$

The wave function can be written in the following form, as ϕ_n constitute a complete set,

$$\psi = \sum_n a_n \phi_n e^{-iE_n t/\hbar}, \tag{A.18}$$

where a_n is an unknown, time-dependant coefficient. Inserting this into Eq. (A.17) gives

$$\sum_n \left\{a_n E_n \phi_n + i\hbar\frac{\partial a_n}{\partial t}\phi_n\right\} e^{-iE_n t/\hbar} = \sum_n a_n \left\{\hat{H}_0\phi_n + \frac{1}{2}\hat{H}_1\phi_n e^{-i\omega t} + \frac{1}{2}\hat{H}_1^\dagger\phi_n e^{i\omega t}\right\} e^{-iE_n t/\hbar}. \tag{A.19}$$

From Eq. (A.16) the first part on both sides cancel out, giving

$$\sum_n \frac{\partial a_n}{\partial t}\phi_n e^{-iE_n t/\hbar} = \frac{1}{2i\hbar}\sum_n a_n \left\{\hat{H}_1 e^{-i\omega t} + \hat{H}_1^\dagger e^{i\omega t}\right\}\phi_n e^{-iE_n t/\hbar}. \tag{A.20}$$

Using the orthogonality of the wave functions, multiplying with ϕ_m and integrating gives

$$\frac{\partial a_m}{\partial t} = \frac{1}{2i\hbar}\sum_n a_n \left\{\left\langle \phi_m|\hat{H}_1|\phi_n\right\rangle e^{-i\omega t} + \left\langle \phi_m|\hat{H}_1^\dagger|\phi_n\right\rangle e^{i\omega t}\right\} e^{iE_{mn}t/\hbar}, \tag{A.21}$$

where $E_{mn} = E_m - E_n$ and Dirac notation has been used. Making a Taylor expansion of a_n gives

$$a_n = a_n^{(0)} + a_n^{(1)} + \ldots, \tag{A.22}$$

and combining this with the fact that if for all x, $\sum_p b_p x^p = \sum_p c_p x^p$ then $c_p = b_p$, it can be found that

$$\frac{\partial a_m^{(p)}}{\partial t} = \frac{1}{2i\hbar}\sum_n a_n^{(p-1)} \left\{\left\langle \phi_m|\hat{H}_1|\phi_n\right\rangle e^{-i\omega t} + \left\langle \phi_m|\hat{H}_1^\dagger|\phi_n\right\rangle e^{i\omega t}\right\} e^{iE_{mn}t/\hbar}, \tag{A.23}$$

where the $(p-1)$ on the right side comes from the fact that the right-hand side already contains one power of perturbation. Setting $p = 0$ gives the time dependency of the unperturbed system,

$$\frac{\partial a_m^{(p)}}{\partial t} = 0, \tag{A.24}$$

which is as expected. Taking $p = 1$ gives

$$\frac{\partial a_m^{(1)}}{\partial t} = \frac{1}{2i\hbar}\sum_n a_n^{(0)} \left\{\left\langle \phi_m|\hat{H}_1|\phi_n\right\rangle e^{-i\omega t} + \left\langle \phi_m|\hat{H}_1^\dagger|\phi_n\right\rangle e^{i\omega t}\right\} e^{iE_{mn}t/\hbar}. \tag{A.25}$$

Integrating this to get $a_m^{(1)}$ gives

$$a_m^{(1)} = -\frac{1}{2}\sum_n a_n^{(0)} \left\{\left\langle \phi_m|\hat{H}_1|\phi_n\right\rangle \frac{e^{-i\omega t}e^{iE_{mn}t/\hbar}}{E_{mn} - \hbar\omega} + \left\langle \phi_m|\hat{H}_1^\dagger|\phi_n\right\rangle \frac{e^{i\omega t}e^{iE_{mn}t\hbar}}{E_{mn} + \hbar\omega}\right\}, \tag{A.26}$$

assuming the perturbation was zero for $t = -\infty$. The result in Eq. (A.26) is idealised as dampening effects are not included. A dampening can be included, which yields

$$a_m^{(1)} = -\frac{1}{2}\sum_n a_n^{(0)} \left\{ \left\langle \phi_m|\hat{H}_1|\phi_n \right\rangle \frac{e^{-i\omega t}e^{iE_{mn}t/\hbar}}{E_{mn} - \hbar\omega - i\hbar\Gamma} + \left\langle \phi_m|\hat{H}_1^\dagger|\phi_n \right\rangle \frac{e^{i\omega t}e^{iE_{mn}t\hbar}}{E_{mn} + \hbar\omega - i\hbar\Gamma} \right\}. \tag{A.27}$$

Now the expectation value of a time-independent operator $\hat{X}$, which corresponds to a measurable quantity, is found through $\left\langle \psi|\hat{X}|\psi \right\rangle$. Inserting the wave function from Eq. (A.18) and only keeping the linear contributions gives

$$\left\langle \psi|\hat{X}|\psi \right\rangle \approx \sum_{m,n} \left\{ a_n^{(0)*}a_m^{(0)} + a_n^{(0)*}a_m^{(1)} + a_n^{(1)*}a_m^{(0)} \right\} \left\langle \phi_n|\hat{X}|\phi_m \right\rangle e^{-iE_{mn}t/\hbar}. \tag{A.28}$$

From the normalisation of the wave function, it is given that

$$1 = \langle \psi|\psi \rangle = \sum_n |a_n|^2, \tag{A.29}$$

where the last equality comes from Eq. (A.18). If there is no perturbation then then

$$\sum_n |a_n^{(0)}|^2 = 1. \tag{A.30}$$

As such, the value $|a_n^{(0)}|^2$ is interpreted as the probability that the unperturbed system is in state n. For thermal equilibrium this can be exchanged for the probability distribution function $|a_n^{(0)}|^2 = f(E_n)$. Due to the orthogonality of the wave functions,

$$a_n^{(0)*}a_m^{(0)} = f(E_n)\delta_{m,n}. \tag{A.31}$$

Using this and Eq. (A.27), Eq. (A.28) can be rewritten,

$$\begin{aligned} \left\langle \psi|\hat{X}|\psi \right\rangle \approx \sum_n f(E_n) \left\langle \phi_n|\hat{X}|\phi_n \right\rangle + \sum_{m,n} a_n^{(0)*}a_m^{(1)} \left\langle \phi_n|\hat{X}|\phi_m \right\rangle e^{-iE_{mn}t/\hbar} \\ + \sum_{m,n} a_n^{(1)*}a_m^{(0)} \left\langle \phi_n|\hat{X}|\phi_m \right\rangle e^{-iE_{mn}t/\hbar}. \end{aligned} \tag{A.32}$$

where through Eq. (A.27) the first sum over n and m becomes

$$\begin{aligned} &\sum_{m,n} a_n^{(0)*}a_m^{(1)} \left\langle \phi_n|\hat{X}|\phi_m \right\rangle e^{-iE_{mn}t/\hbar} \\ &= \sum_{m,n} \left\langle \phi_n|\hat{X}|\phi_m \right\rangle e^{-iE_{mn}t/\hbar} \left\{ -\frac{1}{2}\sum_{n'} a_{n'}^{(0)}a_n^{(0)*} \left[\frac{\left\langle \phi_m|\hat{H}_1|\phi_{n'} \right\rangle e^{-i\omega t}e^{iE_{mn'}t/\hbar}}{E_{mn'} - \hbar\omega - i\Gamma\hbar} \right.\right. \\ &\qquad \left.\left. + \frac{\left\langle \phi_m|\hat{H}_1^\dagger|\phi_{n'} \right\rangle e^{-i\omega t}e^{iE_{mn'}t/\hbar}}{E_{mn'} + \hbar\omega - i\Gamma\hbar} \right] \right\}. \end{aligned} \tag{A.33}$$

Here using Eq. (A.28), the sum becomes

$$-\frac{1}{2}\sum_{m,n} \left\langle \phi_n|\hat{X}|\phi_m \right\rangle f(E_n) \left[\frac{\left\langle \phi_m|\hat{H}_1|\phi_n \right\rangle e^{-i\omega t}}{E_{mn} - \hbar\omega - i\Gamma\hbar} + \frac{\left\langle \phi_m|\hat{H}_1^\dagger|\phi_n \right\rangle e^{i\omega t}}{E_{mn} + \hbar\omega - i\Gamma\hbar} \right]. \tag{A.34}$$

The second sum over n and m can be handled similarly, thus the expression in Eq. (A.32) becomes

$$\left\langle \psi|\hat{X}|\psi \right\rangle \approx \sum_n f(E_n) \left\langle \phi_n|\hat{X}|\phi_n \right\rangle \tag{A.35}$$

$$- \frac{1}{2} \sum_{m,n} f(E_n) \left\langle \phi_n|\hat{X}|\phi_m \right\rangle \left[\frac{\left\langle \phi_m|\hat{H}_1|\phi_n \right\rangle e^{-i\omega t}}{E_{mn} - \hbar\omega - i\Gamma\hbar} + \frac{\left\langle \phi_m|\hat{H}_1^\dagger|\phi_n \right\rangle e^{i\omega t}}{E_{mn} + \hbar\omega - i\Gamma\hbar} \right] \tag{A.36}$$

$$- \frac{1}{2} \sum_{m,n} f(E_m) \left\langle \phi_n|\hat{X}|\phi_m \right\rangle \left[\frac{\left\langle \phi_m|\hat{H}_1^\dagger|\phi_n \right\rangle e^{i\omega t}}{E_{nm} - \hbar\omega + i\Gamma\hbar} + \frac{\left\langle \phi_m|\hat{H}_1|\phi_n \right\rangle e^{-i\omega t}}{E_{nm} + \hbar\omega + i\Gamma\hbar} \right]. \tag{A.37}$$

The Fourier decomposition of this expression is [26]

$$\left\langle \psi|\hat{X}|\psi \right\rangle = \sum_n f(E_n) \left\langle \phi_n|\hat{X}|\phi_n \right\rangle + \frac{1}{2} X(\omega) e^{-i\omega t} + \frac{1}{2} X^*(\omega) e^{i\omega t}. \tag{A.38}$$

Combining the parts in Eq. (A.37) which goes as $e^{i\omega t}$ and the parts which goes as $e^{-i\omega t}$ shows that

$$X(\omega) = - \sum_{m,n} \left[f(E_n) - f(E_m) \right] \frac{\left\langle \phi_m|\hat{H}_1|\phi_n \right\rangle \left\langle \phi_n|\hat{X}|\phi_m \right\rangle}{E_{mn} - \hbar\omega - i\Gamma\hbar} \tag{A.39}$$

A.3 Fourier Transform on Coulomb Potential

In order to use the Coulomb potential in k-space, the Fourier transform of the Coulomb potential is necessary. The Coulomb potential is given as

$$V(r) = \frac{e^2}{4\pi\varepsilon r}. \tag{A.40}$$

The Fourier transform of $V(r)$ is found through

$$V(k) = \int V(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}} \mathrm{d}\mathbf{r}. \tag{A.41}$$

The Fourier transform will first be done in 3D, then in 2D.

3 Dimensions

For the 3D Fourier transform, the problem will be solved using the Yukawa potential, given as

$$V_\lambda(r) = \frac{e^2 e^{-\lambda r}}{4\pi\varepsilon r}, \tag{A.42}$$

where λ is a damping term and in the end letting $\lambda \to 0$. Taking the integral and transforming to polar coordinates gives

$$V(k) = \frac{e^2}{4\pi\varepsilon} \int \frac{e^{-\lambda r}}{r} e^{i\mathbf{k}\cdot\mathbf{r}} \mathrm{d}^3 r = \frac{e^2}{4\pi\varepsilon} \int_0^{2\pi} \int_0^{\pi} \int_0^{\infty} \frac{e^{-\lambda r}}{r} e^{ikr\cos(\theta)} r^2 \sin(\theta) \mathrm{d}r \mathrm{d}\theta \mathrm{d}\phi. \tag{A.43}$$

As it the integral does not depend on ϕ the first integral gives 2π. In order to solve the second integral, substitution is used, with $u = \cos(\theta)$, giving

$$V(k) = \frac{e^2}{2\varepsilon}\int_{-1}^{1} re^{-\lambda r}e^{ikru}\mathrm{d}r\mathrm{d}u = \frac{e^2}{2\varepsilon}\int_0^{\infty} re^{-\lambda r}\left[\frac{e^{ikr}}{ikr} - \frac{e^{-ikr}}{ikr}\right]\mathrm{d}r. \tag{A.44}$$

The last integral can be solved analytically as long as it is the Yukawa potential, giving

$$V(k) = \frac{e^2}{2ik\varepsilon}\int_0^{\infty} e^{(ik-\lambda)r} - e^{-(ik+\lambda)r}\mathrm{d}r = \frac{e^2}{2ik\varepsilon}\left[\frac{e^{(ik-\lambda)r}}{ik-\lambda} + \frac{e^{-(ik+\lambda)r}}{ik+\lambda}\right]_0^{\infty}. \tag{A.45}$$

Giving the exponentials the same denominator and inserting the limits, gives

$$V(k) = \frac{e^2}{2ik\varepsilon}\left[\frac{ik-\lambda+ik+\lambda}{k^2+\lambda^2}\right] = \frac{e^2}{k^2+\lambda^2\varepsilon}. \tag{A.46}$$

Letting $\lambda \to 0$ gives the Fourier transformed Coulomb potential as

$$V(k) = \frac{e^2}{k^2\varepsilon} \tag{A.47}$$

2 Dimensions

For 2D situations, the Fourier transform of Coulomb potential can be calculated directly. Starting from Eq. (A.42), the Fourier transform becomes

$$V(p) = \frac{e^2}{4\pi\varepsilon}\int \frac{e^{i\mathbf{p}\cdot\mathbf{r}}}{r}\mathrm{d}^2r. \tag{A.48}$$

Transforming into polar coordinates gives

$$V(p) = \frac{e^2}{4\pi\varepsilon}\int_0^{\infty}\int_0^{2\pi} e^{ikr\cos(\theta)}\mathrm{d}r\mathrm{d}\theta = \frac{e^2}{2\varepsilon}\int_0^{\infty} J_0(kr)\mathrm{d}r, \tag{A.49}$$

where $J_0(kr)$ is the zeroth order Bessel function. The integral of the Bessel function from 0 to ∞ is unity, thus

$$V(p) = \frac{e^2}{2\varepsilon p}\int_0^{\infty} J_0(p')\mathrm{d}p' = \frac{e^2}{2\varepsilon p}. \tag{A.50}$$

A.4 Fluid Dynamics

This section is based on reference [31].

It turns out that the governing equations for a 2DEG coincides with those of shallow water waves. As such an understanding of how these equations emerge in fluid dynamics is useful. First the continuity equation is derived.

The continuity equation in fluid dynamics represents the conservation of matter. Consider some volume, V_0, the mass of fluid flowing out through the surface enveloping this volume per unit time is

$$\oint \rho\mathbf{v}\cdot\hat{\mathbf{n}}\mathrm{d}S, \tag{A.51}$$

where, ρ is the fluid density. The mass inside the volume will decrease per unit time as the derivative of the mass inside the volume, given as

$$-\frac{\partial}{\partial t}\int \rho \mathrm{d}V. \tag{A.52}$$

As the mass passing through the surface of the volume must be equal to the change of mass in the volume, Eq. (A.51) must be equal to Eq. (A.52), thus

$$\oint \rho \mathbf{v} \cdot \hat{\mathbf{n}} \mathrm{d}S = -\frac{\partial}{\partial t}\int \rho \mathrm{d}V. \tag{A.53}$$

The left-hand side can be transformed into a volume integral through the divergence theorem,

$$\oiint g(\mathbf{F} \cdot \hat{\mathbf{n}}) \mathrm{d}S = \iiint \left[\mathbf{F} \cdot (\nabla g) + g(\nabla \cdot \mathbf{F})\right] \mathrm{d}V, \tag{A.54}$$

where using $[\mathbf{F} \cdot (\nabla g) + g(\nabla \cdot \mathbf{F})] = \nabla \cdot (g\mathbf{F})$, leads to

$$\int \left[\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v})\right] \mathrm{d}V = 0. \tag{A.55}$$

As this equation must hold for every arbitrary volume, the integrand must vanish, yielding

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \tag{A.56}$$

This equation is known as the continuity equation.

To derive the equation known as the Euler equation, consider again some volume, V_0, in a liquid. The force acting on this volume is equal to

$$-\oint P\hat{\mathbf{n}} \mathrm{d}S = -\int \nabla P \mathrm{d}V, \tag{A.57}$$

where the first integral is of the pressure on the surface and the second comes from the divergence theorem. Thus the fluid surrounding an element with volume $\mathrm{d}V$ exerts the force $-\nabla P \mathrm{d}V$ on the volume. The force per unit volume is therefore $-\nabla P$. From Newton's second law, this force can be written as

$$\rho \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\nabla P. \tag{A.58}$$

As the fluid changes position in time the time derivative has to take into account that the position $\mathbf{r} = \mathbf{r}(t)$ is a function of time as well. The change in velocity $\mathrm{d}\mathbf{v}$ during the time $\mathrm{d}t$ can be split into two parts, the change in velocity at a fixed position and the difference in the velocity of two points $\mathrm{d}\mathbf{r}$ apart, where $\mathrm{d}\mathbf{r}$ is the distance the fluid particle have moved in $\mathrm{d}t$. The change at a fixed position in time $\mathrm{d}t$ is $\frac{\partial \mathbf{v}}{\partial t}\mathrm{d}t$. The second part can be written as

$$\mathrm{d}x\frac{\partial \mathbf{v}}{\partial x} + \mathrm{d}y\frac{\partial \mathbf{v}}{\partial y} + \mathrm{d}z\frac{\partial \mathbf{v}}{\partial z} = (\mathrm{d}\mathbf{r} \cdot \nabla)\mathbf{v}. \tag{A.59}$$

Combining these gives

$$\mathrm{d}\mathbf{v} = \frac{\partial \mathbf{v}}{\partial t}\mathrm{d}t + (\mathrm{d}\mathbf{r} \cdot \nabla)\mathbf{v}. \tag{A.60}$$

Dividing both sides with dt to obtain an expression for $\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t}$ yields

$$\frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v}. \tag{A.61}$$

Inserting this into Eq. (A.58) gives

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{1}{\rho}\nabla P. \tag{A.62}$$

This is known as the Euler equation and is the equation of the motion of a fluid. This equation does not take into account the effects of energy dissipation, e.g. through internal friction due to viscosity. Any forces that act upon the fluid is added to the right-hand side, e.g. if the fluid is affected by a gravitational force $\rho\mathbf{g}$, Eq. (A.62) becomes

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{1}{\rho}\nabla P + \mathbf{g}. \tag{A.63}$$

A.5 Mode Index for an Ungated Structure

This section is based on reference [10].

A layered structure without a gate, illustrated in Fig. A.3, is analysed.

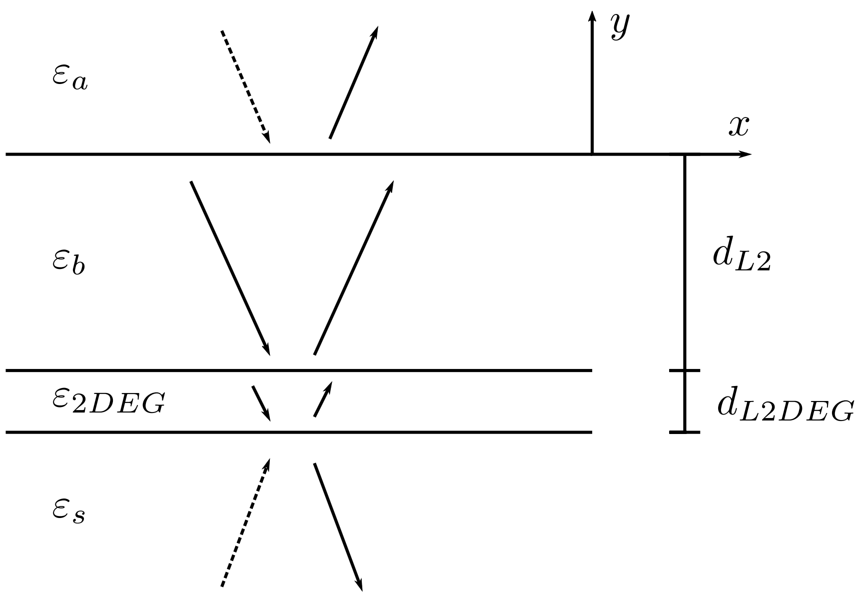

Figure A.3. The HEMT structure without a gate. The dashed lines are the fields which are zero for guided modes.

The structure can be split into four section, where for each section the field can be described as

$$H(\mathbf{r}) = e^{ik_x x}\left(H_\alpha^+ e^{-ik_{y,\alpha}(y-y_\alpha)} + H_\alpha^- e^{ik_{y,\alpha}(y-y_\alpha)}\right), \tag{A.64}$$

where α represents one of the regions in Fig. A.3, H_α^+ is the coefficient for the field propagating downwards and H_α^- is for the field propagating upwards. In order to determine the guided modes in the 2DEG, solutions are sought for which the emitted waves perpendicular to the surface becomes evanescent. Defining $k_{y,\alpha} = \sqrt{k_0^2\varepsilon_\alpha - k_x^2}$, where $k_{y,\alpha}$ is the wavevector for region α, gives the condition that $\mathrm{Im}\{k_{y,\alpha}\} \geq 0$, for the waves perpendicular to the surface being evanescent. In order to find the mode index of these guided modes, the propagation throughout the structure has to analysed, which entails the transmission through each interface and the propagation through the individual layers. As stated in reference [10] transmission through a single interface can be obtained using the standard electromagnetic boundary conditions, which gives the matrix equation for the fields on either side of a boundary:

$$\begin{bmatrix} \tilde{H}_\alpha^+ \\ \tilde{H}_\alpha^- \end{bmatrix} = \frac{1}{t_{\alpha,\alpha+1}} \begin{bmatrix} 1 & r_{\alpha,\alpha+1} \\ r_{\alpha,\alpha+1} & 1 \end{bmatrix} \begin{bmatrix} H_{\alpha+1}^+ \\ H_{\alpha+1}^- \end{bmatrix} = \mathbf{H}_{\alpha,\alpha+1} \begin{bmatrix} H_{\alpha+1}^+ \\ H_{\alpha+1}^- \end{bmatrix}, \tag{A.65}$$

where $\tilde{H}_\alpha$ represent the field at the bottom of layer α, $r_{\alpha,\alpha+1}$ is found as in Eq. (4.57) and $t_{\alpha,\alpha+1}$ is found as in Eq. (4.55). In order to handle the propagation of the field in the given layers, a propagation matrix is used, which according to reference [10] is given as,

$$\begin{bmatrix} H_\alpha^+ \\ H_\alpha^- \end{bmatrix} = \begin{bmatrix} e^{-ik_{y,\alpha}d_\alpha} & 0 \\ 0 & e^{ik_{y,\alpha}d_\alpha} \end{bmatrix} \begin{bmatrix} \tilde{H}_\alpha^+ \\ \tilde{H}_\alpha^- \end{bmatrix} = \mathbf{L}_\alpha \begin{bmatrix} \tilde{H}_\alpha^+ \\ \tilde{H}_\alpha^- \end{bmatrix}, \tag{A.66}$$

where d_α is the thickness of layer α. By applying the matrices from Eq. (A.65) and (A.66) from the first layer throughout the structure, the following relation between the fields at the beginning of the structure and the end of the structure is obtained:

$$\begin{bmatrix} H_1^+ \\ H_1^- \end{bmatrix} = \mathbf{H}_{1,2}\mathbf{L}_2\mathbf{H}_{2,3}\mathbf{L}_3\mathbf{H}_{3,4} \begin{bmatrix} H_4^+ \\ H_4^- \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} H_4^+ \\ H_4^- \end{bmatrix}. \tag{A.67}$$

For guided modes it is required that $H_4^- = H_1^+ = 0$, which means $M_{11} = 0$. The mode index, n_m, is related to the x-component of the wavevector as $k_x = k_0 n_m$. In order to find these mode indices, the real and imaginary part of $M_{11}(n_m)$ must both be zero. This can be checked using a contour plot, illustrating the points where either the imaginary and real part of M_{11} is zero. An example of a contour plot is illustrated in Fig. A.4.

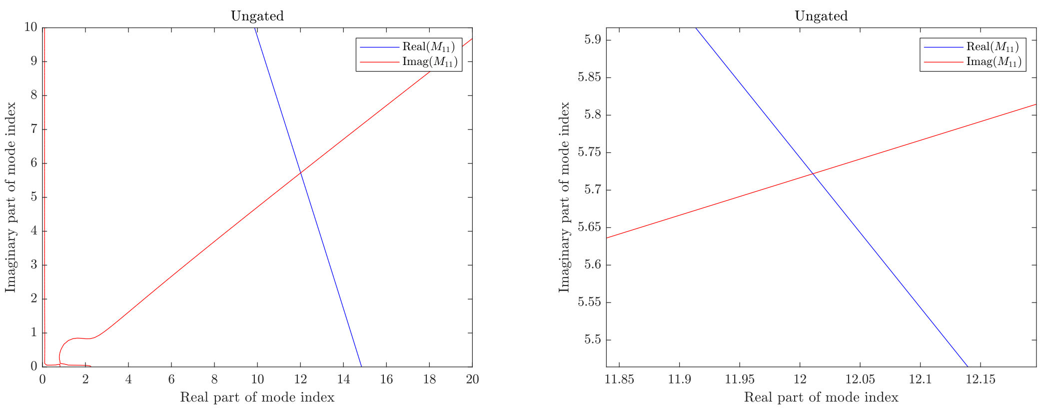

Figure A.4. The contour plot of $M_{11}(n_m) = 0$ for a HEMT structure with a 2DEG, with parameters as given in Ch. 5, where $d_{L2} = 6$ nm and $d_{L2DEG} = 4$ nm. The right picture is focused on the guided mode.

From the contour plot, the mode index can be read off as $n_m \approx 12 + 5.7i$. Another way of finding mode indices is the Newton-Raphson algorithm. The Newton-Raphson algorithm starts from an estimated value, n_{m1} and improves the solution into by assuming the function $M_{11}(n_m)$ is linear, thus giving

$$M_{11}(n_m) = M_{11}(n_{m1}) + (n_m - n_{m1})\frac{\mathrm{d}M_{11}(n_{m1})}{\mathrm{d}n_m}. \tag{A.68}$$

Using the algorithm, the next guess for the mode index is

$$n_{m,i+1} = n_{m,i} - \frac{M_{11}(n_{m,i})}{\frac{\mathrm{d}M_{11}(n_{m,i})}{\mathrm{d}n_m}}. \tag{A.69}$$

This can be done iteratively until a desired precision is attained. It should be noted that the function only converges if the initial estimate is close enough to the solution and the function is approximately linear at the initial estimate. In order to find the initial estimates, the complex plane, where the solutions are sought after, is split into sections, and the phase shift from the corners of each section is checked in order to find possible mode indices. If the phase shifts more than π across the section, an initial estimate in the center of the section is used, as this indicates a change in sign on either the real or the imaginary part of the mode index, e.i. an area where the function has a zero. Using the Newton-Raphson algorithm on the structure described in Fig. A.4 gives the mode index as $n_m = 12.01 + 5.722i$, which corresponds with the value found from the contour plot.

In order to compare with the models where the 2DEG is assumed to be infinitely thin, the mode index for a structure with an infinitely thin 2DEG has also been established. The method is similar, although only a three layer structure is considered, where the 2DEG is incorporated into the boundary between the layer with ε_b and the layer with ε_s. The reflection coefficient for the 2DEG boundary is given in Eq. (4.60) and the transmission coefficient is given in Eq. (4.61). The contour plot is illustrated in Fig. A.5.

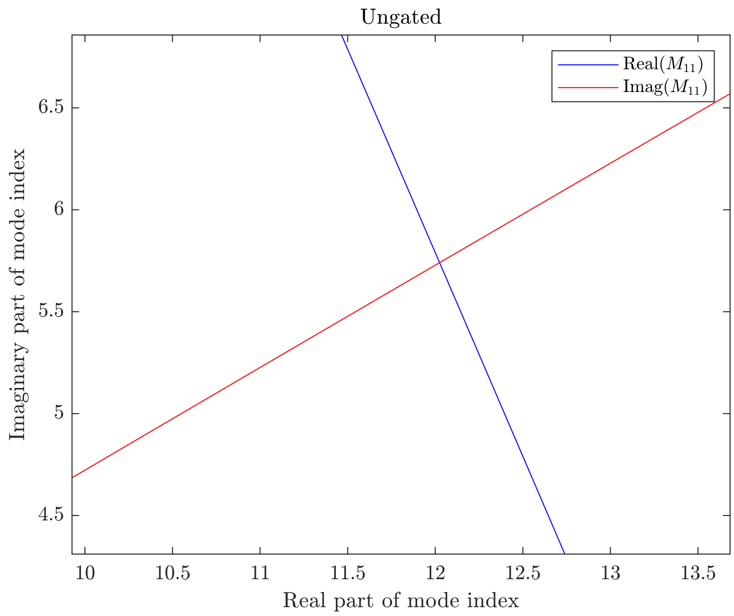

Figure A.5. The contour plot of $f(n_m) = 0$ for a HEMT structure with a infinitely thin 2DEG, where the distance from the gate to the 2DEG is 10 nm and the rest of the parameters are as in Ch. 5.

The mode index corresponds to $n_m \approx 12 + 5.75i$, which is very similar to the model with a thick 2DEG. Using the Newton-Raphson algorithm, the mode index is found to be $n_m = 12.025 + 5.741i$, which is very similar to the thick 2DEG.

A.6 Drude Model

This section is based on reference [43].

In order to analyse the 2DEG, the behaviour of the electrons is considered using the Drude model. The Drude model is a simple model which describes electrons as spherical particles and their momentum through collisions. The relation between an electric field and the current density it introduces is given as

$$\mathbf{E} = \frac{\mathbf{J}}{\sigma}, \tag{A.70}$$

where σ is the conductivity of the material. For a n electrons, the current density can also be described as their movement, yielding

$$\mathbf{J} = -ne\mathbf{v}, \tag{A.71}$$

where $\mathbf{v}$ is the velocity of the electrons. In the presence of a constant electric field, the electrons will experience a force from the electric field, causing the average velocity of the electrons to be

$$\mathbf{v}_{avg} = -\frac{e\mathbf{E}\tau}{m}, \tag{A.72}$$

where, τ is the relaxation time, the average time between collisions. Combining Eq. (A.70), (A.71) and (A.72), gives the conductivity as

$$\sigma = \frac{e^2 n\tau}{m}. \tag{A.73}$$

For a time-varying electric field, the average velocity can be described through the momentum, making the current density

$$\mathbf{J} = -\frac{ne\mathbf{p}}{m}. \tag{A.74}$$

In order to determine $\mathbf{p}(t)$, examine $\mathbf{p}(t + dt)$, where dt is a small time step. The electrons which experiences collision in the time dt, will have a negligible contribution to the momentum, and can thus be ignored. The probability the electron experiences a collision in dt, is dt/τ, thus making the non colliding electrons, $1 - dt/\tau$. The electrons will experience a force from the electric field as $-e\mathbf{E}dt$, giving the momentum of the non colliding electrons as

$$\mathbf{p}(t + dt) = \left(1 - \frac{dt}{\tau}\right)\left[\mathbf{p}(t) - e\mathbf{E}dt\right]. \tag{A.75}$$

Rearranging and dividing with dt gives

$$\frac{\mathbf{p}(t + dt) - \mathbf{p}(t)}{dt} = -eE - \frac{\mathbf{p}(t)}{\tau} + e\mathbf{E}\frac{dt}{\tau}. \tag{A.76}$$

Taking the limit where τ goes to zero, gives

$$\frac{\mathrm{d}\mathbf{p}(t)}{\mathrm{d}t} = -\frac{\mathbf{p}(t)}{\tau} - e\mathbf{E}. \tag{A.77}$$

A solution is given as $\mathbf{p}(t) = \mathbf{p}(\omega)e^{-i\omega t}$. Inserting this solution and the time dependency of $\mathbf{E}$ into Eq. (A.77) yields

$$-i\omega\mathbf{p}(\omega) = -\frac{\mathbf{p}(\omega)}{\tau} - e\mathbf{E}(\omega). \tag{A.78}$$

Isolating $-\mathbf{p}(\omega)$ gives

$$\mathbf{p}(\omega) = -\frac{eE\tau}{1 - i\omega\tau}, \tag{A.79}$$

which inserted into Eq. (A.74) gives

$$\mathbf{J} = -ne\frac{\mathbf{p}}{m} = \frac{ne^2\tau}{m(1 - i\omega\tau)}. \tag{A.80}$$

A.7 Fourier Transform of Second Order Basis Functions

Here the Fourier transforms of the two last second order basis functions from Sec. 5.4 are derived. The first is

$$f_n^{(1)}(k_x) = 4A_n \int_0^1 \left(-z^2 + z\right) e^{-ik_x\Delta z}\mathrm{d}z = 4A_n \left(\int_0^1 ze^{-ik_x\Delta z}\mathrm{d}z - \int_0^1 z^2e^{-ik_x\Delta z}\mathrm{d}z\right). \tag{A.81}$$

First integral, by integration by parts

$$\int_0^1 ze^{-ik_x\Delta z}\mathrm{d}z = \left[-\frac{ze^{-ik_x\Delta z}}{ik_x\Delta}\right]_0^1 + \int_0^1 \frac{e^{-ik_x\Delta z}}{ik_x\Delta}\mathrm{d}z = -\frac{e^{-ik_x\Delta}}{ik_x\Delta} + \frac{e^{-ik_x\Delta} - 1}{k_x^2\Delta^2}. \tag{A.82}$$

The second integral by integration by parts

$$\begin{aligned}\int_0^1 z^2e^{-ik_x\Delta z}\mathrm{d}z &= \left[-\frac{z^2e^{-ik_x\Delta z}}{ik_x\Delta}\right]_0^1 + \int_0^1 2z\frac{e^{-ik_x\Delta z}}{ik_x\Delta}\mathrm{d}z \\ &= -\frac{e^{-ik_x\Delta}}{ik_x\Delta} + \left[-\frac{2ze^{-ik_x\Delta z}}{k_x^2\Delta^2}\right]_0^1 - \int_0^1 2\frac{e^{-ik_x\Delta z}}{k_x^2\Delta^2}\mathrm{d}z \\ &= -\frac{e^{-ik_x\Delta}}{ik_x\Delta} + \frac{2e^{-ik_x\Delta}}{k_x^2\Delta^2} + \frac{2e^{-ik_x\Delta} - 2}{ik_x^3\Delta^3}.\end{aligned} \tag{A.83}$$

Combined this gives

$$\begin{aligned}f_n^{(1)}(k_x) &= 4A_n\left(-\frac{e^{-ik_x\Delta}}{ik_x\Delta} + \frac{e^{-ik_x\Delta} - 1}{k_x^2\Delta^2} + \frac{e^{-ik_x\Delta}}{ik_x\Delta} - \frac{2e^{-ik_x\Delta}}{k_x^2\Delta^2} - \frac{2e^{-ik_x\Delta} - 2}{ik_x^3\Delta^3}\right) \\ &= 4A_n\left(-\frac{e^{-ik_x\Delta} + 1}{k_x^2\Delta^2} - \frac{2e^{-ik_x\Delta} - 2}{ik_x^3\Delta^3}\right).\end{aligned} \tag{A.84}$$

The last basis function is

$$f_n^{(2)}(k_x) = A_n \int_0^1 \left(2z^2 - z\right) e^{-ik_x\Delta z}\mathrm{d}z = A_n \left(\int_0^1 2z^2 e^{-ik_x\Delta z}\mathrm{d}z - \int_0^1 z e^{-ik_x\Delta z}\mathrm{d}z\right). \tag{A.85}$$

The first integral is handled through integration by parts, yielding the same result as Eq. (A.83) times two, while the last integral gives the same result as Eq. (A.82), making the basis function

$$\begin{aligned} f_n^{(2)}(k_x) = A_n &\left(-\frac{2e^{-ik_x\Delta}}{ik_x\Delta} + \frac{4e^{-ik_x\Delta}}{k_x^2\Delta^2} + \frac{4e^{-ik_x\Delta} - 4}{ik_x^3\Delta^3} + \frac{e^{-ik_x\Delta}}{ik_x\Delta} - \frac{e^{-ik_x\Delta} - 1}{k_x^2\Delta^2}\right) \\ &= \left(-\frac{e^{-ik_x\Delta}}{ik_x\Delta} + \frac{3e^{-ik_x\Delta} + 1}{k_x^2\Delta^2} + \frac{4e^{-ik_x\Delta} - 4}{ik_x^3\Delta^3}\right). \end{aligned} \tag{A.86}$$