

Master Thesis



**LUNDS UNIVERSITET**  
Lunds Tekniska Högskola

**Simulating the Effect of Spatial  
Potential Variation in Nanodevices  
with the NEGF Method**

Author  
**Markus Berthilsson**

Supervisor  
**Erik Lind**

First draft:  
*October 27, 2022*  
Second draft:  
*November 16, 2022*

---

## Abstract

In this master thesis nanowires have been simulated to study the effect of spatial imperfections in the nanowires. This spatial imperfection took the form of potential fluctuations in the conduction band across the wire. To achieve this, scripts were written in MATLAB to simulate the nanodevices by applying the non-equilibrium Green's function (NEGF) method. To verify the simulations, multiple quantum phenomena were reproduced in simulations. To reduce the data-size of the simulation results, a compression algorithm based on the Quite OK Image Format (QOI) was also developed.

The simulations were able to qualitatively reproduce quantum tunneling and quantum point conductance. The simulations showed that increasing the amplitude of the applied potential noise to the nanowire had a bigger impact on the conductance of a nanowire than the correlation length of the noise. The noise had a smaller impact on the quantum hall effect than the transmission. Finally the compression algorithm was able to reduce the data size of the results with a worst case scenario of 84.9% and best case scenario reaching 97.8% while having minimal effect on the result.

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Quantum Physics</b>	<b>3</b>
2.1	History . . . . .	3
2.2	Quantum Phenomena . . . . .	4
<b>3</b>	<b>The NEGF Method</b>	<b>6</b>
3.1	The NEGF Equations . . . . .	6
3.2	Hamiltonian . . . . .	8
3.3	Contact Self Energies . . . . .	12
<b>4</b>	<b>The Code</b>	<b>14</b>
<b>5</b>	<b>Results and Discussion</b>	<b>17</b>
5.1	Quantum Tunneling . . . . .	17
5.2	Quantum Point Conductance . . . . .	18
5.3	Transmission through 2D device with spatially varying potential	19
5.4	Quantum hall effect in 2D device with spatially varying potential . . . . .	21
5.5	Compression algorithm . . . . .	22
<b>6</b>	<b>Reflections</b>	<b>24</b>

# 1 Introduction

The transistor is the cornerstone of information technology today and was first developed in 1947. It acts as an electrical amplifier with three terminals: the source, drain and gate. The current between the source and the drain is proportional to the bias applied to the gate and an applied signal to the gate will induce an amplified signal between the source and the drain[1]. Since the discovery of the transistor, its performance has been improved every year by scaling the the transistor down, reducing the channel length distance between the source and the drain[2]. Decreasing the channel length has allowed transistors to work at higher frequencies while also consuming less power and increasing the density of integrated devices. Higher frequency electronics allows for better performance in computation and many other applications. This improvement of the transistor has been exponential as dimensions have shrunk but in later years the frequency improvements have stagnated as channel lengths of less than 20 nm have been reached. Recent development has therefor focused on material optimization and more radical changes are through geometrical structure changes. The finFET transistor allowed the development of transistors to reach a feature size of 7 nm with the drawback that they exhibit a large leakage current. Nanowires are of great interest as they allow gate-all-around geometry which have better theoretical scalability than the classical planar geometry and finFET while also reducing leakage currents[3, 4]. The goal of this master thesis was to study how potential variations and imperfections in nanowires effect their behaviour. This was carried out through computational simulations that use the non-Equilibrium Green's function (NEGF) method.

The NEGF method was developed in the 1960's and is a method of treating quantum transport in nanodevices. Due to the small dimensions of modern devices quantum effects are becoming prominent and needs to be taken into account to accurately predict the behaviour of devices of this scale. Some emerging devices that rely on quantum physics, like the resonant tunneling diodes that use a double barrier to achieve negative differential resistance require the use of quantum mechanics to predict the devices behaviour. The code for the simulations was written in MATLAB. To verify the functionality of the simulations, classical quantum phenomena that have been observed experimentally during the 20th century was replicated through simulations.

Section 2 includes a short qualitative description of some quantum phenomena and history. The following section 3 explains the basics of the NEGF method. This is followed by section 4 which gives a short summery of the script that was written. Finally, results and analysis is in section 5 with reflections of the thesis in section 6.

## 2 Quantum Physics

To verify the functionality of the simulation, experimental results in quantum physics from the past century are going to be replicated through simulations. This following section will give a short description of the history behind quantum physics and the phenomena that are going to be replicated.

### 2.1 History

At the end of the 19th century classical physics was facing a problem regarding the radiation of a black box. Experiments had shown that the emitted spectrum of a black body increases with shorter wavelengths until reaching a peak dependent on temperature before falling off. The problem was that classical physics could not predict this emission spectrum, instead the emission would increase with the frequency of the light towards infinity. It was in 1901 that the physicist Max Planck was able to solve this problem by introducing a constrain that the emitted light could only have quantised energy levels according to  $E = nhf$ , where  $h$  is the Planck constant,  $n$  is an integer and  $f$  is the frequency of the emitted light. At the time this quantization was unexpected as light was only considered to be a wave and should be able to carry any energy in a continuous spectrum.

A few years later in 1905, Albert Einstein proposed the idea of the photon to explain the photoelectric effect where light was able to eject electrons off a metal plate. Photons would be small packets of energy that would act as particles where the energy of the photon was  $E = hf$ . This was in contrast to the wave-like behaviour that light had displayed until that point and was the first case of the wave-particle duality property that light has. Later Louis de Broglie proposed in 1924 that this wave-particle duality was a property of all particles which was related to their momentum according to  $\lambda = h/p$ . This was observed in experiments in the following years which gave Loise the nobel prize in physics 1929. The Austrian-Irish physicist Erwin Schrödinger would then in 1925 set the groundwork to all of quantum physics when he proposed the Schrödinger wave equation for particles.

$$i\hbar \frac{\partial}{\partial t} \Psi(\mathbf{r}, t) = -\frac{\hbar^2}{2m} \Delta^2 \Psi(\mathbf{r}, t) + V(\mathbf{r}) \Psi(\mathbf{r}, t)$$

Where  $\Psi$  is the wave function and  $\hbar = h/2\pi$ . The physical interpretation of the wave function of a particle is still not well established but  $|\Psi(\mathbf{r}, t)|^2$  gives the probability to find a particle at position  $\mathbf{r}$  at time  $t$ . This wave behaviour leads to a series of unexpected behavior for particles that is more well known in wave physics such as audio and optics. For example, a particle that is bound in a small region form standing waves, where the probability of finding a particle is lower at the nodes of the wave function and higher at the antinodes. This means that there are positions in the bounded region where the particle has a 0% chance to be found.

## 2.2 Quantum Phenomena

One famous phenomena in quantum physics is quantum tunneling. This occurs when a particle collides with a barrier that is too great for the particle to cross. The wave function however has a non-zero transmission into the barrier and will have a non-zero value past the barrier. This means that there is a chance for the particle to appear on the other side of the barrier even though it should not have enough energy to cross it according to classical physics. Quantum tunneling is one of the problems with scaling down transistors today as leakage currents are growing large and the difference between the ON and OFF-states are not distinguishable enough. Quantum tunnelling also plays a fundamental part in flash memory. There electrons are trapped in a memory cell by tunneling over a barrier.

Quantum tunneling is similar to normal optics where part of an incident wave will transmit through a material while part of the incident wave is reflected. Anti-reflective coating is used in optics to improve the transmission close to 100% for the right wavelengths by the use of multiple coats of different materials. This creates a series of reflections and transmissions which causes constructive interference for the transmitted wave and destructive interference for the reflected wave if the wavelength matches the coat thicknesses. The same idea can be applied in quantum physics where a particle can have a higher transmission through two barriers compared to a single barrier for certain energies. In electronics this would mean that the resistance of two resistors in series would be lower than the resistance of a single resistor. One emerging application of resonant tunneling are compact oscillators in the tera-hertz range for high speed wireless communication[5].

When an electron travels through a device it will collide with other particles exchanging energy and losing momentum. However if the device that electron passes through is small enough there will be practically no collisions and the electron will only lose energy when being transferred from the contact to the channel. The electron is then in the ballistic regime and will experience a conductance of

$$G_0 = \frac{e^2}{\pi\hbar},$$

known as the quantum conductance. In 1988 it was discovered experimentally that as the width of a 2D ballistic device increased[6], the conductance would take on values of  $G = G_0 \cdot n$  as seen in figure 1. An explanation to this is that as the width of the device increases, more standing waves can be created in the device along the width, allowing more channels that carries current to form. The same behavior is observed as the voltage across the device increases as this reduces the wavelength of the wave function.

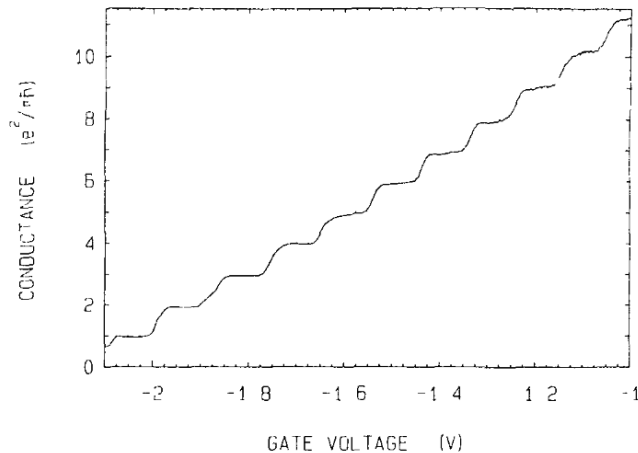


Figure 1: The conductance of the device plateaus at regular voltage levels. The applied gate voltage restricted the movement for charges in the channel effectively adjusting it's width. Figure taken from [6].

When a current flows through a metal strip and a strong magnetic field is applied over the strip, a voltage difference will appear across the width of the strip perpendicular to the current flow. This is known as the hall effect and is due to how magnetic fields create a force on charges perpendicular to their direction of momentum. The hall resistance ( $R_{xy}$ ) is defined as  $R_{xy} = V_H/I$  where  $V_H$  is the hall voltage and  $I$  is the current over the channel. Classically the hall resistance is continuous but in very small two-dimensional systems at low temperatures this behavior is quantised and changes in steps. It was shown experimentally in 1980 by the physicist Klaus von Klitzing that this quantum hall resistance increases in steps according to

$$R_{xy} = \frac{h}{e^2 n}. \quad (1)$$

Where  $n$  is an integer dependent on the applied magnetic field. This quantum hall resistance is very accurate in measurements and only deviates in parts of a million from eq (1). Von Klitzing was awarded the Nobel prize in physics 1985 for this discovery[7]. Due to the accuracy of the quantum hall effect, the National Metrology Institutes uses the quantum hall effect as one method to accurately recreate the SI unit of resistance where the von Klitzing constant is exactly defined as  $R_k = h/e^2$ [8].

### 3 The NEGF Method

The non-equilibrium Green's function (NEGF) method was first developed in the 1960's by the work of Martin, Schwinger, Kadanoff, Baym, Keldysh and others building on many-body perturbation theory. It is an efficient method for computers to solve the Schrödinger equation in discrete space and can simulate the quantum transport in large systems with thousands of lattice points. The theory and rigorous derivation based on many-body perturbation theory is far beyond the scope of this thesis work and would not be possible to learn in the time span of the work. However simpler reasoning starting from the Schrödinger equation gives the same equations as the original work. The following section 3.1 gives a summary of the NEGF equations that are used to simulate systems and sections 3.2-3.3 describes the quantities used in the NEGF method further.

#### 3.1 The NEGF Equations

The NEGF method works in the discrete space where not every single point of a continuous system are represented but instead discrete lattice points separated in space. This is similar to digital electronics where digital systems measure a continuous voltage and convert it into a set of discrete points separated in time. If the resolution is high enough the discrete representation of the continuous system can be sufficient for accurate calculations. This allows computers to make numerical computations on discrete values instead of having to solve continuous equations.

The simulated system can be broken up to 3 different components as seen in figure 2, the channel described by the Hamiltonian ( $\mathbf{H}$ ) and the interactions with the contacts that are described by the self energies ( $\Sigma_n$ ).

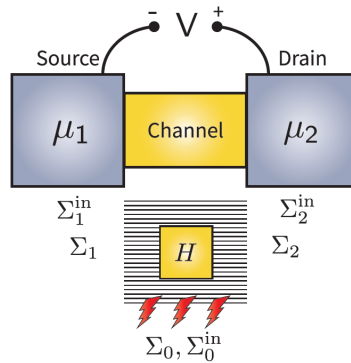


Figure 2: A simple representation of a device. Two contacts are connected to a channel where the interaction is described by  $\Sigma_n, \Sigma_n^{in}$ . The channel is described by  $H$  and  $\Sigma_0, \Sigma_0^{in}$ . Figure taken from [9].

Both these quantities are represented by matrices and will be further described in sections 3.2 and 3.3.  $\mathbf{H}$  and  $\Sigma_n$  arises in the Schrödinger equation when boundary conditions are applied for the system in figure 2,

$$E\Psi = \mathbf{H}\Psi + \underbrace{\Sigma\Psi}_{\text{outflow}} + \underbrace{\mathbf{s}}_{\text{inflow}}. \quad (2)$$

Where  $\mathbf{s}$  is the matrix representation of the inflow of charge to the device from the contacts,  $\Sigma$  is the sum of  $\Sigma_N$  from the contacts and  $\Psi$  is the wave equation describing the charges in the system.  $\Sigma$  and  $\mathbf{s}$  arises from the Schrödinger equation when the boundary conditions imposed by the contacts are applied. Rearranging this gives the retarded green's function ( $\mathbf{G}^R$ ) which describes the relation between  $\mathbf{s}$  and  $\Psi$ ,

$$\mathbf{G}^R(E) = [E\mathbf{I} - \mathbf{H} - \Sigma]^{-1}, \quad \Psi = \mathbf{G}^R\mathbf{s} \quad (3)$$

Where  $\mathbf{I}$  is the identity matrix and  $E$  is the energy in the system. For example if a bias of 1 V is applied to the device, the system will have an energy of 1 electron volt. The anti-retarded greens function ( $\mathbf{G}^A$ ) is defined as the complex conjugate of  $\mathbf{G}^R$

$$\mathbf{G}^A = [\mathbf{G}^R]^*.$$

From  $\mathbf{G}^R$  and  $\mathbf{G}^A$  the spectral function ( $\mathbf{A}$ ) can be calculated as,

$$\mathbf{A} = i[\mathbf{G}^R - \mathbf{G}^A]. \quad (4)$$

The diagonal elements of  $\mathbf{A}$  corresponds to the local density of states (multiplied by  $2\pi$ ) for every lattice point that is being simulated.

The self energies  $\Sigma$  relate to the the outflow of charge in the channel and there is a corresponding quantity which represents the inflow of charge ( $\Sigma^{in}$ ).  $\Sigma^{in}$  is given by

$$\Sigma_n^{in} = \Gamma_n \cdot f(E) \quad (5)$$

$$\Gamma_n = i[\Sigma_n - \Sigma_n^*]. \quad (6)$$

Where  $f(E)$  is the Fermi level at the contact.  $\Sigma_n^{in}$  represents the inflow of electrons to the device from channel  $n$ . The electron density in the channel is then given by

$$\mathbf{G}^n = \Psi\Psi^* = \mathbf{G}^R\mathbf{s}[\mathbf{G}^R\mathbf{s}]^* = \mathbf{G}^R \underbrace{\mathbf{s}\mathbf{s}^*}_{\Sigma^{in}} \mathbf{G}^A = \mathbf{G}^R\Sigma^{in}\mathbf{G}^A. \quad (7)$$

Where  $\Sigma^{in}$  is the sum of all  $\Sigma_n^{in}$ . The diagonal elements of  $\mathbf{G}^n$  corresponds to the electron density (multiplied by  $2\pi$ ) at different lattice points of the simulation. The Fermi level in the channel is given by

$$f_m = \mathbf{G}_{mm}^n / \mathbf{A}_{mm} \quad (8)$$



where  $\mathbf{G}_{mm}^n$  is the  $m$ 'th diagonal element of  $\mathbf{G}^n$ .

Finally the conductance at contact  $n$  per unit energy is given by

$$\hat{I}_n = \frac{q}{h} \text{Trace}[\boldsymbol{\Sigma}_n^{in} \mathbf{A} - \boldsymbol{\Gamma}_n \mathbf{G}^n] \quad (9)$$

$$\text{T} = \text{Trace}[\boldsymbol{\Sigma}_n^{in} \mathbf{A} - \boldsymbol{\Gamma}_n \mathbf{G}^n] \quad (10)$$

and T is the transmittance of an incoming wave to the channel.

These equations represent the ideal case where there is no phonon interactions in the channel and the electrons will only interact with the channel itself. In realistic scenarios the electron continuously interacts with other electrons flowing through the channel and experiences a rapidly changing potential in the channel. This causes the electrons to lose the phase and at higher temperatures phonon interaction will also lead to a loss of momentum of the electrons. This scattering phase relaxation can be represented by a self energy, in this work referenced as  $\boldsymbol{\Sigma}_0$ , which is given by

$$\boldsymbol{\Sigma}_0 = \mathbf{D} \times \mathbf{G}^R \quad (11)$$

$$\boldsymbol{\Sigma}_0^{in} = \mathbf{D} \times \mathbf{G}^n \quad (12)$$

In these equations the  $\times$  symbol denotes element wise multiplication. Depending on how the matrix  $\mathbf{D}$  is constructed, different types of scattering can be simulated. An element  $D_{ij}$  in the  $\mathbf{D}$  matrix represents the correlation between the random potential at point  $i$  to point  $j$ . There are two special cases that are of interest, the first is

$$D_{ij} = D_0. \quad (13)$$

This means that all elements in  $\mathbf{D}$  are equal to  $D_0$ . This uniform matrix will only introduce phase relaxation in the channel and no momentum loss. This phase relaxation will cause quantum phenomena like resonance in the channel to be less pronounced. The other case is when both momentum relaxation and phase relaxation is present which occurs when

$$\mathbf{D} = D_0 \cdot \mathbf{I} \quad (14)$$

where  $\mathbf{I}$  is the identity matrix.

Since  $\boldsymbol{\Sigma}_0$  is dependent on  $\mathbf{G}^R$  and  $\mathbf{G}^R$  is dependent on  $\boldsymbol{\Sigma}_0$  according to Eq. (3),  $\boldsymbol{\Sigma}_0$  has to be solved iteratively where an initial guess of  $\boldsymbol{\Sigma}_0$  is used to calculate a new value for  $\boldsymbol{\Sigma}_0$  according to Eq. (11). The same reasoning applies to  $\boldsymbol{\Sigma}_0^{in}$  where  $\mathbf{G}^n$  depends on  $\boldsymbol{\Sigma}_0^{in}$  and vice versa.

### 3.2 Hamiltonian

The Hamiltonian is a central part in describing a quantum system used in the time independent Schrödinger equation,

$$E\Psi(\mathbf{r}) = \mathbf{H}\Psi(\mathbf{r}) \quad (15)$$

where  $\mathbf{H}$  is the Hamiltonian operator which corresponds to the potential and kinetic energy of the system in real space representation,

$$\mathbf{H} = -\frac{\hbar^2}{2m}\nabla^2 + U(\mathbf{r}). \quad (16)$$

This is sufficient to solve simple systems analytically and can be used to solve the wave function for electrons in potential wells or describe quantum tunneling. These approaches are continuous however and not practical for numerical calculations with computers. For numerical operations the system is divided up into lattice points transforming the continuous space of the system into a discrete one. The wave function is then represented as a vector where every vector component equates to the value of the wave function at a defined lattice point. The Hamiltonian changes to a matrix where the elements correspond to how strongly the lattice points interact with the wave function over the entire system.

To show how the Hamiltonian in matrix form is constructed, consider the following simple case of a uniform one-dimensional wire with infinite length as illustrated in figure 3.

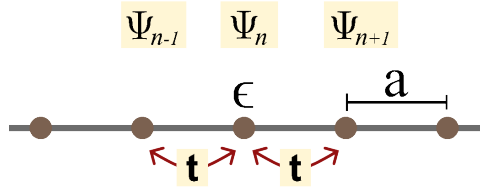


Figure 3: A one dimensional wire with lattice points separated by the distance  $a$ . The interaction between lattice points is described by  $t$  and the interaction between the lattice point and the wave equation is described by  $\epsilon$ .

At every lattice point along the wire the energy contribution to the wave function at that point is  $\epsilon$  while the neighboring lattice point has an interaction with magnitude  $t$  to the wave function. The resulting equation for every point is then

$$E\Psi_n = t\Psi_{n-1} + \epsilon\Psi_n + t\Psi_{n+1}. \quad (17)$$

Assuming that the wire is homogeneous and infinite or with periodic boundary condition, a solution to (17) is of the form

$$\Psi_n = \Psi_0 e^{ikna} \quad (18)$$

gives the  $E(k)$  relation

$$E = \epsilon + 2t \cos ka. \quad (19)$$

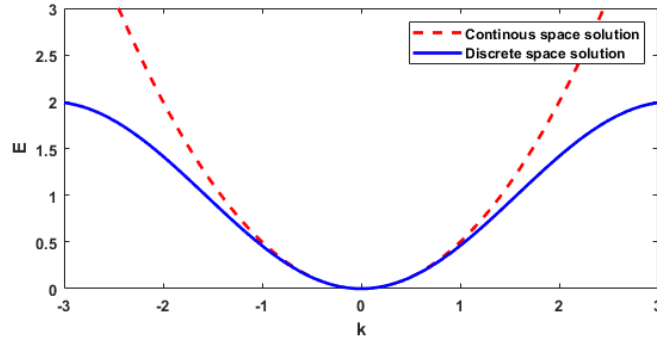


Figure 4: Comparing  $E(k)$  relation of Eq. (19) and Eq. (20).

Approaching the same wire with the Schrödinger equation gives

$$E\Psi = -\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\Psi + E_c\Psi$$

where  $E_c$  is the conduction band level of the material. Assuming the same solution as in Eq. (18) gives the  $E(k)$  relation

$$E = E_c + \frac{\hbar^2 k^2}{2m}. \quad (20)$$

Taylor expanding Eq. (19) and comparing with the solution in Eq. (20) gives the values

$$\epsilon = E_c - 2t \quad \text{and} \quad t = -\frac{\hbar^2}{2ma^2}. \quad (21)$$

The solutions of the two methods differ, but for smaller values of  $E$  the result is similar as seen in figure 4. The discrete approach can give a good approximation of a quantum system by using appropriate values for the Hamiltonian, however the energies have to remain relatively low. The values for  $\epsilon$  and  $t$  does not have to be derived by theoretical means but can instead be derived by studying smaller samples to then be used to simulate more complex systems. The expressions in Eq. (21) also indicates that  $t$  is related to how strongly the lattice are bonded to each other since it depends on the distance between lattice while  $\epsilon$  is related to the conduction band level.

Constructing the Hamiltonian is done using the relation in Eq. (17). Assuming the same wire as earlier but now with a limited length of 3 lattice points, the system can be described as

$$E \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \end{bmatrix} = \begin{bmatrix} \epsilon & t & 0 \\ t & \epsilon & t \\ 0 & t & \epsilon \end{bmatrix} \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \Psi_3 \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} \epsilon & t & 0 \\ t & \epsilon & t \\ 0 & t & \epsilon \end{bmatrix}. \quad (22)$$

Thus the matrix version of the Hamiltonian has been constructed. Since  $\epsilon$  is related to the conduction band level at each point, it is easy to insert a barrier in the material by changing the value of  $\epsilon$  related to the barriers position.

This is only a simple one-dimensional case but the construction for higher dimensional problems can be done in a similar manner. For a two-dimensional rectangular system of length  $M$  and width  $N$ , every column of  $N$  units can be represented as a unit cell as illustrated in figure 5. Every unit cell can then be described as a  $N \times N$  matrix ( $\alpha$ ) which would correlate to the  $\epsilon$  value in Eq. (22). In figure 5 the values of  $t$  has separate values in the  $x$  and  $y$  direction but this does not need to be the case. In the one-dimensional case the  $t$  values describe the interaction between the lattice points in the system, so for the two-dimensional case a  $N \times N$  matrix ( $\beta$ ) will describe the interaction between columns. In figure 5,  $\beta$  has the value  $t_x$  along the diagonal and the rest of the elements being 0.

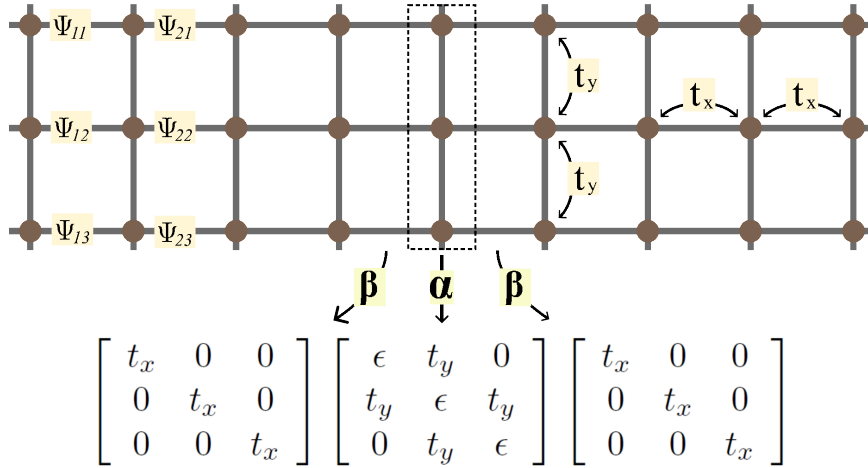


Figure 5: A two dimensional wire where the columns are represented by a unit cell marked by the dotted box. The wire has a width of 3 lattice points and length of 9 lattice points, so the resulting Hamiltonian will be of the size  $27 \times 27$ .

The Hamiltonian for the two-dimensional system is constructed similar to the one-dimensional case in Eq. (22) but replacing the values  $\epsilon$  and  $t$  with the matrices  $\alpha$  and  $\beta$ . Note that every single lattice point contributes with one diagonal element to the Hamiltonian as these diagonal elements all are the  $\epsilon$  values of the lattice point. This means that a device with  $M$  by  $N$  lattice points will have an Hamiltonian that is a square matrix with a side of  $M \cdot N$  elements. It is possible to construct Hamiltonian for more complex systems which are not rectangular as in the examples mentioned. As long as the relation between the lattice points and the wave equation is established

at every point, a Hamiltonian can be constructed that fulfills this relation in matrix form. For the scenario depicted in figure 5 and described in this section, the resulting Hamiltonian would have the following form,

$$E\Psi = \overbrace{\begin{bmatrix} \alpha & \beta & & & \\ \beta & \alpha & \beta & & \\ & \beta & \alpha & \beta & \\ & & & \ddots & \ddots \end{bmatrix}}^{\mathbf{H}} \overbrace{\begin{pmatrix} \Psi_{11} \\ \Psi_{12} \\ \vdots \\ \Psi_{21} \\ \Psi_{22} \\ \vdots \end{pmatrix}}^{\Psi} \quad (23)$$

Once a magnetic field is present over the channel, an additional phase factor will be introduced to the coupling between lattice points. Coupling to the right as described by Eq. (17) will gain a positive phase change while coupling to the left will have a negative phase change. Same applies to "above" and "under" in the 2-dimensional case.

$$t_x(r) = te^{\pm i\phi_x(r)}, \quad \phi_x(r) = \frac{qA_x(r)a}{\hbar} \quad (24)$$

$$t_y(r) = te^{\pm i\phi_y(r)}, \quad \phi_y(r) = \frac{qA_y(r)a}{\hbar} \quad (25)$$

Where  $q$  is the charge of the particle and  $\mathbf{A}$  is the magnetic vector potential at the lattice point.  $\mathbf{A}$  is not unique for a field  $\mathbf{B}$ , but must follow the relation

$$\nabla \times \mathbf{A} = \mathbf{B}. \quad (26)$$

### 3.3 Contact Self Energies

In section 3.2 systems have only been described in isolation. The Hamiltonian describes the channel of the device that will be simulated but to simulate a connected system there needs to be contacts attached to the device. These contacts supply an inflow and outflow of charge and imposes boundary conditions which add terms to Eq. (15),

$$E\psi = \mathbf{H}\psi + \underbrace{\Sigma\psi}_{\text{outflow}} + \underbrace{\mathbf{s}}_{\text{inflow}}. \quad (27)$$

In this equation a source term  $\mathbf{s}$  has been added as well as the self energies  $\Sigma$  which represents the outflow and inflow of electrons. Rearranging this gives the retarded green function as mentioned in section 3.1.

$$\mathbf{G}^R = [E\mathbf{I} - \mathbf{H} - \Sigma]^{-1}, \quad \psi = \mathbf{G}^R\mathbf{s} \quad (28)$$

Once  $\mathbf{G}^R$  has been solved, most of the desired characteristics of the system can be derived simply through the equations described in section 3.1. The

Hamiltonian was described in the previous section 3.2 but the self energies from the contacts also have to be solved before calculating  $\mathbf{G}^R$ . The self energy of a contact is given by<sup>1</sup>

$$\Sigma_m = \tau_m \hat{\mathbf{g}}_m \tau_m^* \quad (29)$$

$$\hat{\mathbf{g}}_m = [(E + i\eta)\mathbf{I} - \mathbf{H}_m]^{-1}. \quad (30)$$

Where  $\mathbf{H}_m$  is the Hamiltonian describing the contact and  $\tau_m$  is describing how strong the interaction between the contact and the channel is.  $\eta$  is chosen to be a small value. Since the contacts tend to be very large and  $\mathbf{H}_m$  is of the size  $M \times M$ , where  $M$  is the total amount of lattice points in the contact, the resulting matrix  $\hat{\mathbf{g}}_m$  in Eq. (30) is enormous. Calculating  $\hat{\mathbf{g}}_m$  is not practical and another recursive method is used instead to find the surface greens function ( $\mathbf{g}_m$ ) for a contact. This method assumes the contact to be infinite and homogeneous.  $\mathbf{g}_m$  is effectively the smaller part of  $\hat{\mathbf{g}}_m$  that interacts with the channel. The recursive method uses unit cells as in 3.2 to reduce higher dimensional problems to one dimension, the setup can be visualized in figure 6.

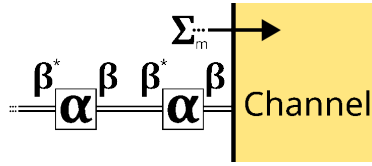


Figure 6: The contact represented as a one dimensional system where the columns of the contacts are described by  $\alpha$  interacting with the next column described by  $\beta$ .

Starting with the first column of the contact and repeatedly using Eq. (30) gives,

$$\mathbf{g}_m^N = [(E + i\eta)\mathbf{I} - \alpha - \beta^* \mathbf{g}_m^{N-1} \beta]^{-1}$$

Which then converges towards the solution. Since the contacts to the channel are very large this gives when  $N \rightarrow \infty$ ,

$$\mathbf{g}_m = [(E + i\eta)\mathbf{I} - \alpha - \beta^* \mathbf{g}_m \beta]^{-1}. \quad (31)$$

So finally the self energy is given by using the final result from Eq. (31) in Eq. (29).

The dephasing effects are also implemented as a virtual contact that interacts with the entire channel of the device. This virtual contact is denoted by the self energies  $\Sigma_0$  and  $\Sigma_0^{in}$  which are calculated iteratively as explained at the end of section 3.1.

<sup>1</sup>The full derivation for the self energies in systems with multiple dimensions is rather lengthy and is found at [10].

## 4 The Code

The scripts for applying the NEGF equation was developed over the course of 7 months. Analysis of a sample is carried out in 3 steps. The first step is to construct a sample that is going to be simulated and specify energy values. The next step is to perform the computation to solve  $\mathbf{G}^R$  and  $\Sigma_0$ . This is done by constructing  $\mathbf{H}$  as described in section 3.2. The self energies for the contact are computed next by methods described in section 3.3. Once this is done  $\mathbf{G}^R$  and  $\Sigma_0$  are solved with Eqs. (3) and (11) from section 3.1. Final step is to extract the data from the result with equations summarized in section 3.1. The first step, setting up the sample specifying the parameters for the simulations are done by using the following objects:

- `Sample.m` which is an object that represents the sample containing information about the lattice points in the channel and contacts attached. The `Sample` class has the functions `append` to let the user append a matrix to the channel; `applyNoise` to add a noise level to the lattice points channel to replicate the effect of imperfections. This noise has a Gaussian height distribution with a specified correlation length; `addContact` to add a `Contact.m` to a desired location in the channel. The contacts Fermi level also has to be specified if the electron density and current levels are to be extracted.
- `NEGF_param.m` which is a data class to fill the parameters for the NEGF simulation. Here error rates and convergence rates can be specified for the NEGF simulations. The `NEGF_param` object needs a `Sample` and energy range ( $\mathbf{E}$ ) to be constructed. A magnetic range ( $\mathbf{B}$ ) can also be specified if desired. `NEGF_param` creates a copy of the `Sample` and should be constructed just before calling `NEGF_map`.

The next step is to perform the computation to extract the result. This is done by calling `NEGF_map` or `NEGF`. The result is then returned as a `NEGF_result` object:

- `NEGF_map.m` is a function that repeatedly calls `NEGF` for the entire range of  $\mathbf{E}$  and  $\mathbf{B}$  specified by `NEGF_param`. `NEGF.m` only computes the simulations for a single value of  $\mathbf{E}$  and  $\mathbf{B}$  and `NEGF_map.m` is more general. `NEGF_map` uses results from completed NEGF simulations to speed up the process of simulating other values for  $\mathbf{E}$  and  $\mathbf{B}$ . If the system is very large or if the resolution is high, the simulations may take a very long time and the progress can be periodically saved specified by an interval set in `NEGF_param`. A flowchart for `NEGF_map.m` and `NEGF.m` is found in figure 7 at the end of this section.
- `NEGF_result.m` is the data class that is returned from `NEGF`. since `NEGF_map.m` saves a series of results, these are embedded into a

structure. The result contains data such as the calculated self energies,  $\Sigma_0$  and  $\Sigma_0^{in}$ . The result also contains the function *compress* which does a lossy compression of the data. This is done by first reducing the resolution of the data from 64-bit to 8-bit by transforming the data to values between 0 and 255 then converting the data-type. This is followed by a lossless compression called QOI which is an open source compression algorithm developed 2021. This compression algorithm was chosen due to its excellent speed offering 20-50 times the speed performance compared to a much more popular image compression format PNG while achieving similar compression rates[11, 12]. The compression algorithm is also very simple so it was more feasible to implement in MATLAB code for the project. Performance of this compression is studied in section 5.5.

Once the results from the simulations have been calculated, analysis are simplified by the following two functions:

- [NEGF\\_transmission.m](#) which takes a [NEGF\\_result](#) as input and calculates the transmission by using Eq. (10).
- [NEGF\\_result\\_remap.m](#) remaps desired data from a [NEGF\\_result](#) to a 2D matrix. Supported data types as of writing this report is "*Electrons*" which extracts the electron density in  $\mathbf{G}^n$ , "*Fermi*" which returns the Fermi level for every point in the channel and "*A*" which returns the diagonal values of  $\mathbf{A}$  which corresponds to the density of states in the channel.



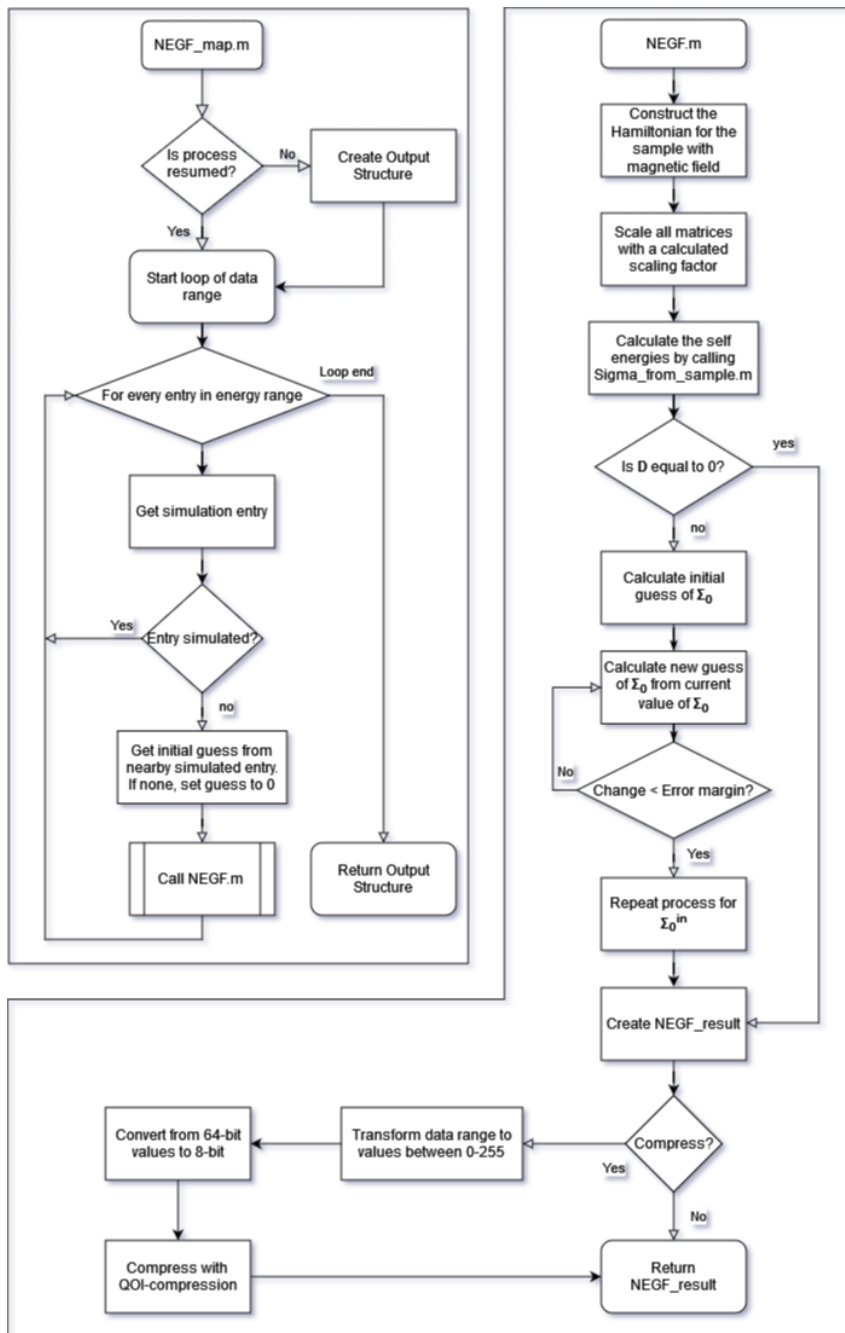


Figure 7: Flow chart representing the process of `NEGF_map.m` and `NEGF.m`

## 5 Results and Discussion

In the following section, the results of the simulations and script performance are summarized. Classical quantum phenomena are replicated in sections 5.1 and 5.2. Simulations of nanowires and the effects of imperfections are then be studied in sections 5.3 where the transmission is studied and 5.4 where the quantum hall effect is studied. The section finishes off in 5.5 by analysing performance of the compression algorithm that was made for the simulation results.

### 5.1 Quantum Tunneling

Quantum tunneling was simulated by creating a one-dimensional wire featuring two barriers. This one dimensional wire was set up with a length of 42 lattice points and with barriers placed at lattice point 18 and 24. The effect of momentum and phase relaxation was also applied according to Eq. (13,14). The conduction band level was set to  $E_c = 1.12$  eV and then the distance between lattice points was set to  $a = 2\text{\AA}$ . From this the values of  $\epsilon$  and  $t$  was given from Eq. (21). The result from the simulations is illustrated in figure 8 and 9.

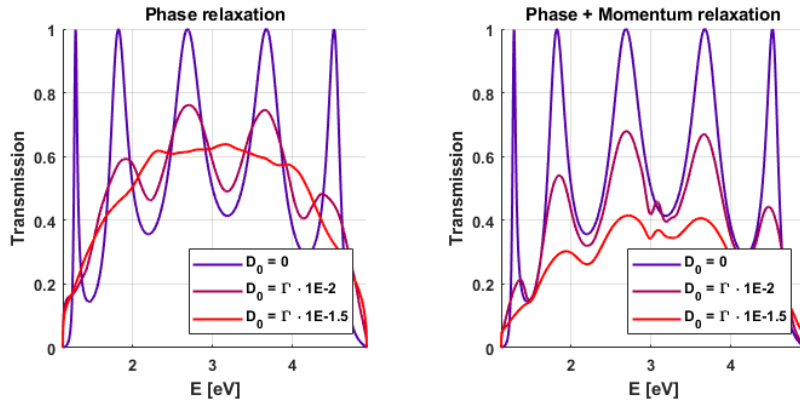


Figure 8: The transmission as a function of energy for a wire with two barriers.  $\Gamma$  is defined as  $\epsilon \cdot 1$  eV.

In figure 8 the effect of resonance is seen in where the transmission reaches 1 at certain energy values. When the value of  $D_0$  grows the resonance effect become less prominent and the transmission through the two barriers becomes smaller. Comparing the effect of only phase relaxation to the effect of both phase and momentum relaxations shows that momentum and phase relaxation has a greater impact on the behaviour of the device. The transmission reducing to 0 when  $E$  approaches 5 is an error of the simulation since the simulation is only valid for smaller values of  $E$ . This was

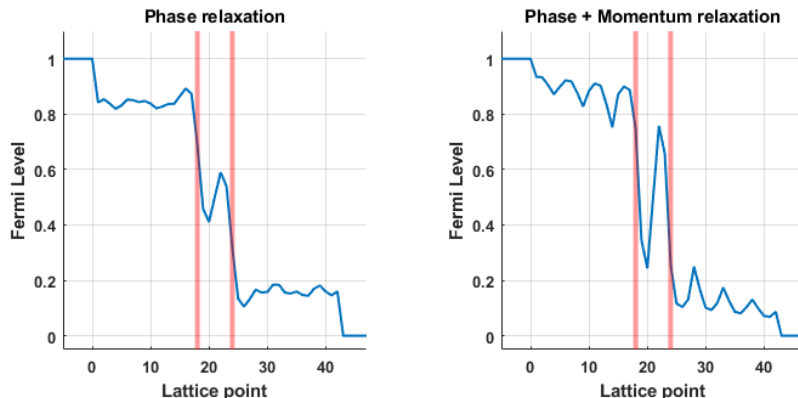


Figure 9: Fermi level across the one-dimensional wire with 2 barrier's marked in red.  $D_0 = \Gamma \cdot 1e - 2$ .

shown in section 3.2 where figure 4 shows that the simulations are inaccurate for larger values of  $E$ . Figure 9 shows how the momentum relaxation adds a slope of energy loss across the channel while simply introducing phase relaxation only reduces the effect of resonance.

It's worth noting that the simulation time of only phase relaxation is longer than the simulation of phase and momentum relaxation. On average, over 30 full run times to calculate the results in figure 8, phase relaxation took 34.4 seconds to complete simulation while phase and momentum relaxation only took 12.7 seconds. When only phase relaxation is added  $\mathbf{D}$  is a uniform matrix with the value  $D_0$  for every element in the matrix while phase and momentum relaxation only includes values of  $D_0$  along the diagonal. When  $\mathbf{D}$  is diagonal,  $[\mathbf{G}^R]^{-1}$  is mainly going contain elements close to the diagonal with a lot of elements being equal to 0, allowing faster algorithms for inverting the matrix.

## 5.2 Quantum Point Conductance

To investigate the quantum point conductance, a channel of length 5 and varying width was simulated. The distance between lattice points was chosen to 1 nm,  $E_c$  was set to 1.12 eV and the effective mass of the electrons was chosen to be  $0.5 m_e$ . The width was then varied from 1 lattice point to 60 and the transmittance is plotted in figure 10.

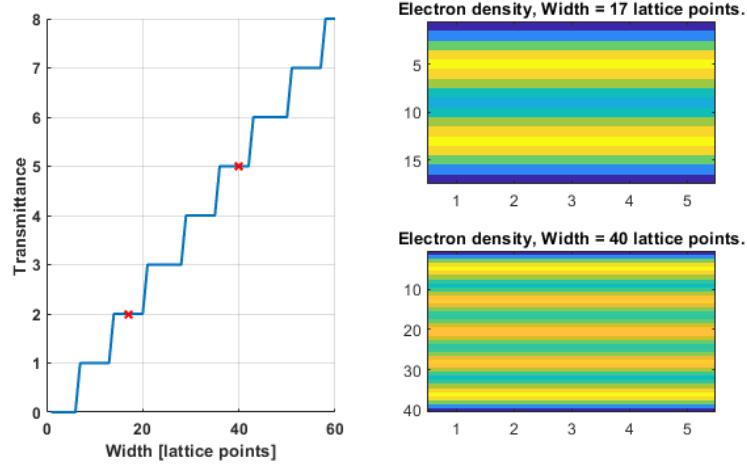


Figure 10: Conductance of a 2 dimensional wire when increasing the width of the wire. To the right are the resulting electrons densities extracted with [NEGF\\_result\\_remap.m](#), yellow indicates a higher concentration. The two crosses in the transmittance curve correspond to the plotted electron concentrations to the right.

The conductance can be derived when applying Eqs. (9,10). Using the differential resistance gives  $R = \frac{dU}{dI} = \frac{dU/dE}{dI/dE}$ .  $dI/dE = \frac{q}{h}T$  is given from directly from Eqs. (9,10).  $dU/dE = 1/q$  since one joule of energy corresponds to  $1/q$  eV. This gives  $R = \frac{h}{q^2T} \rightarrow G = \frac{q^2T}{h}$  which is the result found in 1988 study[6] with a difference of factor 2. This factor difference is due to spin degeneracy which the model currently does not include.

### 5.3 Transmission through 2D device with spatially varying potential

The nanowire that was simulated had a width of 22 lattice points and length of 40 lattice point with the lattice points spaced 1 nm from each other. The effective mass of the electron was set to  $0.1 m_e$  and  $E_c = 1.12$  eV. Increasing the bias voltage applied over the sample gave similar transmission curve as observed in section 5.2. The reason to this can be seen in the first image of the electron density in the same figure. Since the frequency of the wave increases according to Eq. (20) with  $E$  more standing waves can be created transversal to the current, creating more conduction channels in the device. When the noise becomes greater the conductance of the device decreases. When the noise is introduced it becomes harder to form standing waves in the device. When the amplitude of this noise becomes larger, the restriction on the formed waves become stronger and the electrons have to take more convoluted paths as seen in the later images of figure 11. Effectively the

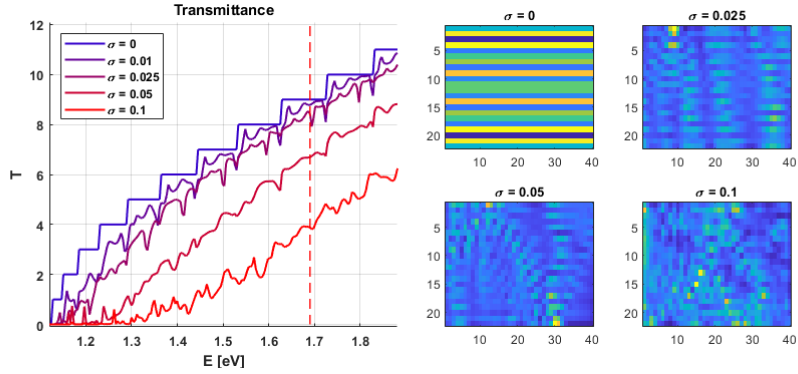


Figure 11: Multiple simulations of varying noise amplitude.  $\sigma$  is the standard deviation for the amplitude of the noise in relation to  $\epsilon$  of the device. The images to the right are the electron density in the device at  $E = 1.69$  eV as indicated by the red line in the transmission curves. The correlation length is set to  $C_l = 4.4$ .

current stops becoming ballistic as the electrons cannot travel straight across the channel. Varying the correlation length has a smaller impact as seen in figure 12. Since the path of the electrons is not as restricted as in figure 11 the loss of transmittance is less prominent.

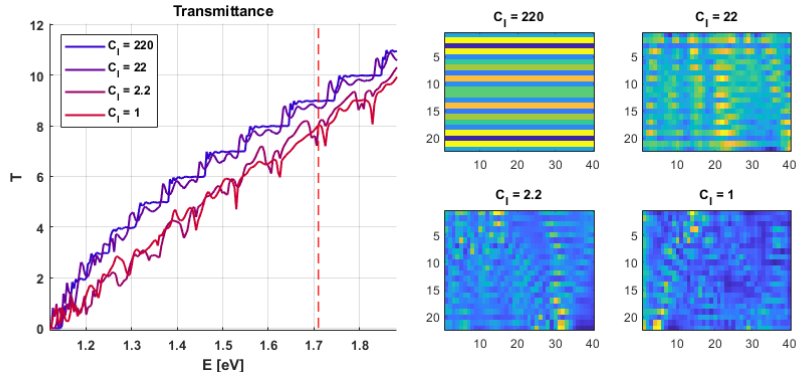


Figure 12: Multiple simulations of varying correlation length.  $A_l$  is the correlation length in lattice points. The images to the right are the electron density in the device at  $E = 1.71$  eV as indicated by the red line in the transmission curves. Noise amplitude is set to  $\sigma = 0.025$

## 5.4 Quantum hall effect in 2D device with spatially varying potential

For the quantum hall simulations, the same device as in section 5.3 was used with the difference of  $E_c$  was set to 0 eV. The reason for this was because the bias was kept low above the conduction band at  $E = E_c + 0.015$  eV. Since the noise addition to the device is relative to  $\epsilon = E_c + 4t_0$  this noise would be much greater than the bias if  $E_c$  was large. In similar fashion of section 5.3, the amplitude of the noise has a bigger impact than the correlation length as seen in figure 13 and 14. Since the quantum hall resistance depends on both the voltage across the device and the transmission the fluctuation may increase much more depending on how the voltage is measured. For figures 13 and 14 the average value across the entire length of the channel along the edges were used.

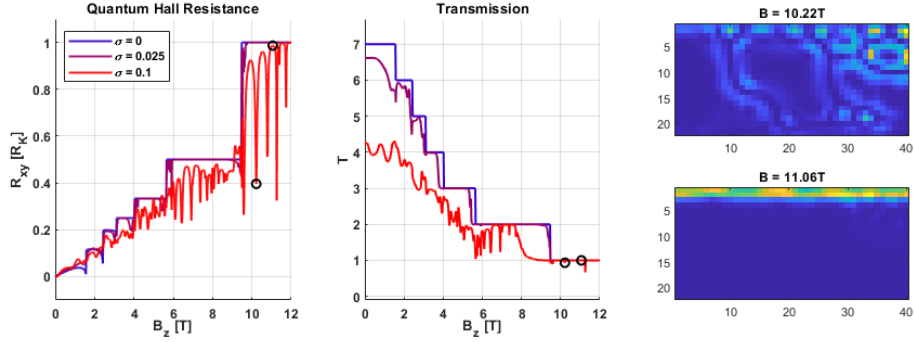


Figure 13: Multiple simulations of varying noise amplitude.  $\sigma$  is the standard deviation for the amplitude of the noise in relation to  $\epsilon$  of the device. The images to the right are the electron density in the device with  $\sigma = 0.1$ . The correlation length is set to  $C_l = 4.4$ .

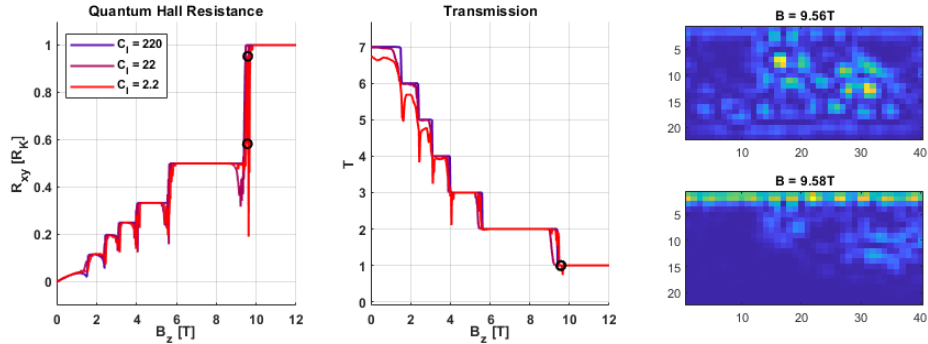


Figure 14: Multiple simulations of varying correlation length.  $C_l$  is the correlation length in lattice points. The images to the right are the electron density in the device with  $C_l = 2.2$ . Noise amplitude is set to  $\sigma = 0.025$ .

Unlike the transmission measurements in section 5.3 the quantum hall resistance is less effected by the noise in the device and the plateaus remain the same values as in the noise-free device. The fluctuations seen in both figures are probably due to the voltage measurement varying as the distribution of electrons in the device fluctuate with varying B field. Images supporting this idea are present in figures 13 and 14 to the right where the images representing a B-field with low hall resistance features smaller local regions separated from the walls where electrons accumulate. The noise may introduce potential wells at these regions where electrons get trapped which influences the voltage measurement and thus the hall resistance. As in section 5.3 the amplitude of the noise has a larger effect on the performance on the device than the correlation length.

## 5.5 Compression algorithm

Beside the physical results, there was also effort put into making compression algorithms to reduce the size of the data of the results. This was done since the resulting matrices can reach sizes of hundreds of gigabytes once bigger systems are simulated. Naively using larger and larger data storage to handle results of these sizes would not work as the systems would scale up. Therefor a compression algorithm was implemented for the results. This algorithm has 2 steps, first reducing the resolution of the data from 64-bit to 8-bit, this is followed by a lossless compression called QOI. To test the performance and error induced by the lossy compression, a 2D wire with different length, barriers and applied noise were simulated and then the result was compressed. 3 different wires with different sizes were simulated with 1000 simulation points each. Each wire had a width of 20 lattice points but the length was set to 10,20 and 30 points respectively where the nanowire with a length of 20 lattice points had a single barrier and the nanowire with a length of 30 points had a double barrier. This was followed by calculating the transmission with the compressed data and comparing the result with non-compressed data. The error in transmission calculated from the compressed data was less then 0.06 for all simulations and with an average error of less than  $4.2E-5$ . This was achieved while reducing the necessary storage by 84% in the worst case and often over 90% as seen in figure 15.

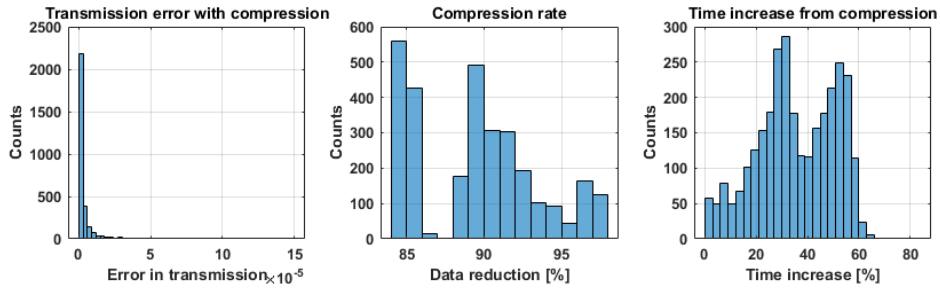


Figure 15: Performance of the compression method tested on 3000 simulation results of different 2D nanowires.

The compression and decompression needed to calculate the transmission however caused slowdown to the simulations, with compression the results and then analysing the data increasing total calculation time by 66% in the worst case and 35.4% on average. This does not sound great, however, the simulations were rather simple ones with small values for  $\mathbf{D}$  which means that the results were fast to extract. The simulations will take more time to compute in more complicated scenarios where external electrostatics are applied and the simulations have to account for the poisson equation while the compression and decompression time would remain the same, thus improving the relative performance greatly.



## 6 Reflections

There's a lot to take away from this thesis work and it has shown how versatile the NEGF method is and simulations like this could assist as a great tool to help understand nanodevices and the impact quantum mechanics has on them. The results in sections 5.1, 5.2 and 5.4 shows that the simulations are able to recreate quantum phenomena accurately. A great benefit to simulating systems like this compared to analytically studying them is that more complex results can be extracted easily such as the electron density in the devices which helps giving an understanding to the inner workings of these devices. The simulations has shown that the potential fluctuation introduced into the device will reduce the transmittance and I propose that this is due to the electrons not functioning in the ballistic regime anymore which would result in a lower transmittance. The simulations also shows that the quantum hall effect is very resilient to applied noise and the hall resistance plateaus would remain consistent even when  $\sigma = 0.1$  in figure 13. A lot of the fluctuations in  $R_{xy}$  seems to be due to how the voltage was extracted from the simulation examples and real measurement would likely not exhibit fluctuations this aggressive. Moreover, the simulations in sections 5.3 and 5.4 had no scattering present since including this in simulations would be to computationally heavy.

With the work finishing there's many things that I wish I could improve or study further. There's still great expandability of the code and performance improvements to be made. The first is applying the Poisson equation for electrostatics which has to be considered if external voltages are applied. In the case of a transistor where external voltage is applied through a gate, the applied voltage will affect the NEGF equations. The result of the NEGF equations will then in turn effect the Poisson equations and finding a solution that fulfill both equations has to be done iteratively. This will increase computation requirements greatly and as it stands the current scripts need to be improved before this can be done. Performing the simulations in section 5.3 took 6 minutes for the plots in figures 11 and 12 each, and performing the simulations for figures 13 and 14 took 16 minutes and 25 minutes respectively. These systems were still fairly small and simulating larger systems that would then have to be done iteratively could take upwards hours or days. Since the size of  $\mathbf{H}$  grows by the number of lattice points ( $n$ ) and the inversion of a matrix is of  $O(n^3)$ , a larger system will require much more computation to simulate. One potential way to improve simulation times would be through GPU acceleration. There's also faster methods of finding the diagonal elements of  $\mathbf{G}_n$  and  $\mathbf{G}$  which uses block wise inversion [13, 14] which was considered to be implemented for the scripts but due to lack of time it was never realised. Except for performance improvements it would be interesting to simulate real systems and compare

the results to measurements on real samples of nanowires. This would be done in 2 steps, first measuring a small sample which a model will replicate. Through the results from the measurement on a nanowire parameters in the model could be adjusted until the simulated results fit the real world data. The second step would then be to tweak this model and use the extracted parameters to simulate another sample to the compare and validate the simulations. In this thesis only qualitative simulations were done to study the general behaviour of nanowires but if real devices can be simulated of great accuracy this would open up possibilities to assist in the growth and study of nanowires.

## References

- [1] I.M. Ross. “The invention of the transistor”. In: *Proceedings of the IEEE* 86.1 (1998), pp. 7–28. DOI: 10.1109/5.658752.
- [2] R.R. Schaller. “Moore’s law: past, present and future”. In: *IEEE Spectrum* 34.6 (1997), pp. 52–59. DOI: 10.1109/6.591665.
- [3] Erik Lind. “High frequency III–V nanowire MOSFETs”. In: *Semiconductor Science and Technology* 31.9 (Aug. 2016), p. 093005. DOI: 10.1088/0268-1242/31/9/093005.
- [4] J. Ajayan et al. “Nanosheet field effect transistors-A next generation device to keep Moore’s law alive: An intensive study”. In: *Microelectronics Journal* 114 (2021), p. 105141. ISSN: 0026-2692. DOI: <https://doi.org/10.1016/j.mejo.2021.105141>.
- [5] Yusei Suzuki et al. “Phase Control of Terahertz Waves Using Injection-Locked Resonant Tunneling Diode Oscillator”. In: *IEEE Transactions on Terahertz Science and Technology* 12.5 (2022), pp. 481–488.
- [6] B. J. van Wees et al. “Quantized conductance of point contacts in a two-dimensional electron gas”. In: *Phys. Rev. Lett.* 60 (9 Feb. 1988), pp. 848–850. DOI: 10.1103/PhysRevLett.60.848.
- [7] *The nobel prize in physics 1985*. 2020. DOI: 10.48550/ARXIV.2008.01275.
- [8] T J Quinn. “News from the BIPM”. In: *Metrologia* 26.1 (Jan. 1989), pp. 69–74. DOI: 10.1088/0026-1394/26/1/006.
- [9] Kerem Y. Camsari, Shuvro Chowdhury, and Supriyo Datta. *The Non-Equilibrium Green Function (NEGF) Method*. 2020. DOI: 10.48550/ARXIV.2008.01275.
- [10] Supriyo Datta. *Courses: Nanohub-U: Fundamentals of nanoelectronics - part B: Quantum Transport, 2nd Edition*. URL: <https://nanohub.org/courses/FON2>.
- [11] Dominic Szablewski. *Lossless image compression in  $O(N)$  time*. Nov. 2021. URL: <https://phoboslab.org/log/2021/11/qoi-fast-lossless-image-compression>.
- [12] Dominic Szablewski. *The quite ok image format for fast, lossless compression*. URL: <https://qoiformat.org/>.
- [13] S. Li, Shaikh Ahmed, and Eric Darve. “Fast inverse using nested dissection for NEGF”. In: *Journal of Computational Electronics* 6 (Jan. 2007), pp. 187–190. DOI: 10.1007/s10825-006-0112-8.

- [14] S. Li, W. Wu, and E. Darve. “A fast algorithm for sparse matrix computations related to inversion”. In: *Journal of Computational Physics* 242 (2013), pp. 915–945. ISSN: 0021-9991. DOI: <https://doi.org/10.1016/j.jcp.2013.01.036>.