Quantum Transport Through Monolayers Of Organophosphonate Molecules On Silicon/Silicon Dioxide Substrates

A Master’s Dissertation Submitted In Partial Fulfilment For The Degree Of MSc. In Microelectronics

By
Mohammed Ahmed Fouad Ibrahim Darwish
June 29, 2015

To
Institute for Nanoelectronics
Technische Universität München (TUM)
And
School of Electronics & Electrical Engineering
Nanyang Technological University (NTU)
Abstract

Nanoelectronics is becoming ever so important as more and more research is done in this field, not only due to the fact that it shows a lot of promise regarding the future of electronic devices but also the ability to combine different sciences together hence expanding to areas which the field can be utilized. One intriguing device application as a result of this combination is the “self-assembled monolayer field effect transistor (SAMFET)”. SAMFETs are interesting due to the ability to control the size and length of a channel region, which is chemically synthesized, on an atomic scale. There are several aspects still unknown concerning the functioning of such devices which are related to how charges are injected inside the SAM layer from the contacts. This is a single molecule charge transfer that requires a quantum mechanical machinery to be correctly addressed, hence the motivation for the work done in this thesis.

Two molecules used as test cases for SAMFET experiments are under investigation in this work. One molecule is an alkane chain, more specifically a decyl-(C10) phosphonic acid, and the other molecule is an anthracene chain more specifically a 2,6-diphosphonoanthracene. The first part of this investigation focuses on understanding the characteristics of the chemical bonding between the molecules and the silicon/silicon dioxide substrate. In the second part of this investigation, following the bonding of the molecules to the silicon/silicon dioxide substrate (which acts as one of the contacts), gold surfaces are connected to the other end of the molecules acting as the second contact, and transport simulations are then carried out to investigate how charge is transported through the molecular junctions. This work aims at understanding the charge injection inside SAMFET devices.
## Contents

1 Introduction ........................................................................................................................................ 1

1.1 From Micro to Nano ..................................................................................................................... 1

1.2 Nanotechnology .......................................................................................................................... 2

1.2.1 Single-Electron Devices .......................................................................................................... 4

1.2.2 Novel Memory Devices .......................................................................................................... 7

1.2.3 Carbon Nanotubes ................................................................................................................ 10

1.2.4 Organic Electronics .............................................................................................................. 13

1.2.5 Molecular Electronics ........................................................................................................... 18

1.2.5.1 Molecular Diodes ............................................................................................................. 18

1.2.5.2 Molecular Storage Elements ............................................................................................. 19

1.2.5.3 Three Terminal Devices .................................................................................................... 21

1.2.5.4 Self-Assembled Monolayers ............................................................................................. 22

2 Electronic Structure Theory ........................................................................................................... 26

2.1 Electronic Structure Problem ..................................................................................................... 26

2.1.1 Born-Oppenheimer Approximation ....................................................................................... 27

2.1.2 Hartree-Fock Approximation ............................................................................................... 28

2.2 Basis Sets ................................................................................................................................... 32

2.2.1 Slater-Type Orbitals ............................................................................................................... 32

2.2.2 Gaussian-Type Orbitals ......................................................................................................... 33

2.3 Density Functional Theory ......................................................................................................... 36

2.3.1 Thomas-Fermi Model ........................................................................................................... 37

2.3.2 Hohenberg-Kohn Theorem .................................................................................................. 38

2.3.3 Kohn-Sham Equations ......................................................................................................... 40

2.3.4 Exchange Correlation ........................................................................................................... 42

2.3.4.1 Local Density Approximation ......................................................................................... 42

2.3.4.2 Generalized Gradient Approximation ............................................................................. 43
Contents

3 Electronic Transport ........................................................................................................45
  3.1 The Quantum of Conductance ..................................................................................45
  3.2 Landauer Formalism ...............................................................................................46

4 Results and Discussion ..................................................................................................52
  4.1 Building and optimizing the system .........................................................................52
    4.1.1 Si/SiO₂ substrate and the decyl-(C₁₀) phosphonic acid ....................................57
    4.1.2 Si/SiO₂ substrate and the 2,6-diphosphonoanthracene .......................................58
  4.2 Transport Analysis .....................................................................................................61
    4.2.1 Transport through Si/SiO₂/decyl-(C₁₀) phosphonic acid/Au .............................63
    4.2.1 Transport through Si/SiO₂/2,6-diphosphonoanthracene/Au ..............................67

5 Conclusion .....................................................................................................................72

Bibliography ......................................................................................................................74

Acknowledgments .............................................................................................................83
List of Figures

1.1 Exponential decrease of minimum feature size versus time................................. 1
1.2 Effect of electronic confinement on the density of states..................................... 3
1.3 Circuit diagram of a SET and its energy band diagram......................................... 5
1.4 Illustration of a QCA with both logic states....................................................... 5
1.5 Different QCA Boolean functions......................................................................... 6
1.6 Structure of a magnetic tunneling junction MRAM based memory cell............... 8
1.7 Illustration of the “Millipede” concept............................................................... 9
1.8 40nm bits formed with various pitch sizes......................................................... 9
1.9 Illustration of Graphene honeycomb lattice, and the different types of CNTs......... 11
1.10 Bandstructure of Graphene................................................................................ 11
1.11 Separation of CNTs in a centrifugation tube....................................................... 12
1.12 Different fabrication techniques of organic devices............................................ 13
1.13 Top and bottom contact OTFT architectures..................................................... 14
1.14 Structure and band diagram of an OLED............................................................ 16
1.15 TPD and Alq3 chemical structures...................................................................... 16
1.16 Different types of OLED configurations............................................................. 17
1.17 Aviram-Ratner single molecular rectifier........................................................... 19
1.18 Illustration of a molecular storage element......................................................... 20
1.19 I-V characteristics of a molecular storage element............................................. 21
1.20 Back gate and electrochemical gate configurations............................................. 22
1.21 Self-assembled monolayers formed by solution based deposition...................... 23
1.22 Different architectures of SAMFETs................................................................. 24

2.1 Comparison between contracted Gaussian functions and Slater functions........... 34
2.2 Comparison between normal basis sets and split valence basis sets...................... 35

3.1 Broadening of a channel between source and drain contacts............................... 46
3.2 Incoming, transmitted, and reflected modes in a two probe system....................... 47

4.1 Construction of Si substrate form a unit cell....................................................... 52
4.2 Construction of amorphous oxide....................................................................... 53
4.3 Radial distribution function of an amorphous oxide............................................ 54
List of Figures

4.4  Alkane and anthracene chains before and after relaxation.......................... 55
4.5  HOMO/LUMO for the alkane chain................................................................. 55
4.6  HOMO/LUMO for the anthracene chain.......................................................... 56
4.7  Relaxation of the alkane chain on Si/SiO$_2$....................................................... 57
4.8  HOMO/LUMO of the alkane chain after bonding.............................................. 58
4.9  Relaxation of the anthracene chain on Si/SiO$_2$................................................. 59
4.10 HOMO/LUMO of the anthracene chain after bonding...................................... 59
4.11 Construction of a gold contact from a unit cell............................................... 61
4.12 Final arrangement of both molecular configurations divided into specified parts.... 61
4.13 Doping of the Si substrate.................................................................................. 62
4.14 Alkane chain I-V characteristics and transmission spectra.............................. 63
4.15 DOS at 0V and 0.3V......................................................................................... 64
4.16 DOS at 0.6V and 0.9V..................................................................................... 65
4.17 Log scale of the molecule density of states....................................................... 66
4.18 Anthracene chain I-V characteristics and transmission spectra....................... 67
4.19 DOS at 0V and 0.3V......................................................................................... 68
4.20 DOS at 0.6V and 0.9V..................................................................................... 69
4.21 Log scale of the molecule density of states....................................................... 70
List of Tables

4.1 Simulation parameters and results................................................................. 56
4.2 Summary of the geometry optimization (relaxation) process............................. 60
Chapter 1: Introduction

1.1 From Micro to Nano

The electronics industry has come a long way since the invention of the first transistor [1]. Almost all integrated circuits consist of a Metal Oxide Semiconductor Field Effect Transistor (MOSFET) as the basic element. Over the years the dimensions of these transistors have been constantly scaled down from the micrometer to nanometer range and the device density increased, an observation known as Moore’s Law [2], to achieve higher performance. Computer processors that used to have thousands of transistors, now are made of well over a billion of transistors. This is evident from the huge improvement we can see today. Since the beginning of the microelectronic era, the minimum feature size has been reduced by a rate of 13% each year [3]. Fig. 1.1 shows what the industry has been able to achieve in terms of scaling down the minimum feature size.

![Fig. 1.1: Exponential decrease of minimum feature size versus time [3]](image)
But over the past few years this rate has been slowing down, because of the challenges that arise in trying to reach the next technology node. Such challenges are a consequence of the MOSFET technology reaching its limits [4]. One fundamental limit is the 60 mV/dec subthreshold swing [5]. With scaling down the device dimension the supply voltage must also be scaled down to minimize power dissipation, but it can’t go lower than 0.5V otherwise we will lose control over the device operating conditions (i.e. switching between on and off states), hence the subthreshold swing limit. Another limit is material dependent, and one example is the gate oxide integrity. With the constant thinning down of the gate oxide, this has led to the rise in the leakage current through the gate of a MOS transistor. That is why the industry has turned to high-k materials, so that the thickness of the gate can be increased and keeping the capacitance constant. Even though this approach was successful and has been used for the 45nm technology [6], it did not solve the problem entirely, rather suppress it for the time being.

Furthermore, fabrication processes are becoming more and more difficult due to the limitation of the current techniques such as the wavelength of the lithography processes. The same thing goes for measurements and defect inspection. With future technologies well inside the nanometer regime, the electronic properties of these devices are extremely sensitive to even the smallest changes in their physical properties and this requires precise characterization techniques [7].

Conventional MOSFETs which have provided all the advances in the electronic industry will eventually run into severe problems, and that is why a lot of effort has now turned to nanotechnology in order to find alternative solutions if Moore’s law continues to be valid.

1.2 Nanotechnology

Nanotechnology is defined as the technology of devices with a diameter of less than 100nm [8], which effectively means that since the year 2002 with the introduction of the 90nm MOSFET, we have entered the nanoelectronics era. It is also important to point out that the science used to explain the nature of how devices in the micron and sub-micron regime operate was not going to be sufficient for those in the nanoscale regime. Attention was shifted from classical theory to quantum theory. Theories developed by the great minds of the 20th century such as Schrödinger, Bohr, De Broglie, and Heisenberg (to name a few!) were of great importance to understand how these nanodevices operate.

There are two major effects that come into play when dealing with structures with very small dimensions.
The first effect is quantum confinement [9], and it consists of 4 different cases:

1. **0-Dimensional Confinement** - Electrons are free to move in the x, y, and z directions (Bulk, Fig. 1.2(a)).
2. **1-Dimensional Confinement** - Electrons have two free directions of motion (Quantum Well, Fig. 1.2(b)).
3. **2-Dimensional Confinement** - Electrons have one free direction of motion (Quantum Wire, Fig. 1.2(c)).
4. **3-Dimensional Confinement** - Electrons have no free motion in any direction. (Quantum Dot, Fig. 1.2(d)).

![Fig. 1.2: Effect of electronic confinement on the density of states for (a) bulk (b) quantum well (c) quantum wire (d) quantum dot](image)

The second effect is quantum mechanical tunneling. According to classical physics, if an electron energy is lower compared to that of the barrier it will not be able to penetrate the barrier and is reflected back (the region inside the barrier is classically forbidden) [10]. But the wave function associated with the electron must be continuous at the barrier and will show an exponential decay inside the barrier. Furthermore the wave function must also be continuous on the other side of the barrier, hence there is a finite probability that if the barrier is thin enough an electron may be able to tunnel through. Both quantum confinement and quantum mechanical tunneling are important to understand how conduction occurs in nanoscale devices. Of course these effects are most relevant in quantum mechanics under the assumption that there is no correlation among electrons, but such effects generate more and more complex behavior taking into consideration a many-body system.
consisting of highly correlated states. Nevertheless these two effects provide the basis or starting point from which one can go deeper into understanding the nature of nanoscale devices.

Even though MOSFETs were and still are the source of all the advancements in the electronic industry, this has not stopped researchers within the same time frame to explore other alternatives keeping in mind that eventually everything must come to an end. Since it became evident that they were right and that MOSFETs are slowly reaching the end of the line with challenges arising at a rate faster than that of which they can be solved, the need for something new was ever so important. For that reason a lot of time, effort, and money have been poured into research to come up with novel devices that would eventually replace MOSFETs and be integrated into the different areas of the electronics industry. Furthermore, what made nanotechnology even more exciting is that in order to investigate structures on an atomistic scale, a good knowledge in chemistry and how these structures are chemically synthesized was also required. This requirement bridged the gap between two different sciences that opened the door to new possibilities in which MOSFETs could not be utilized. Not only would this be beneficial for the electronics industry, but also in areas such as mechanical engineering and medicine. Therefore, I will introduce some of the achievements that have been done in the field of nanoelectronics based on the previously stated quantum effects and that make use of the different ways these materials are chemically synthesized.

### 1.2.1 Single-Electron Devices

Single-electron devices are ones that control charge flow with one electron at a time. They are basically composed of a quantum dot and two electrodes. One electrode is connected to the quantum dot via a tunneling junction acting as a barrier and the other electrode (i.e. gate electrode) is coupled to it via an insulating material hence forming a capacitor. The placement and removal of a single electron to/from the quantum dot depends on the voltage applied to the gate and the charging energy of the quantum dot. If the charging energy required for an electron to tunnel through to the quantum dot is larger than the thermal energy, hence only at a specified applied bias tunneling occurs. Moreover, once the quantum dot is filled by an electron, the injection of another electron is forbidden by Coulomb repulsion. The same procedure is done if an electron is to be removed, of course this is done with a reverse polarity to the gate. This phenomena of injection or ejection is called the “Coulomb blockade” effect.

An equivalent circuit schematic for a single-electron transistor (SET) based on the Coulomb blockade is shown in Fig. 1.3(a). The quantum dot acting as a conducting island is capacitively coupled to gate. At first the system is biased off resonance and a gap exists between the states in the island and the Fermi levels of both the source and drain (Fig. 1.3(b)). Resonance occurs when
the states due to Coulomb charging lie between the Fermi levels on the left and right sides in the linear response regime (Fig. 1.3(c)). The gate voltage allows the tuning between stable regimes, thus adding or subtracting an electron to and from the island. Such SETs were first investigated by Fulton and Dolan [11] and Kuzmin and Likharev [12].

A very interesting structure that employed the idea of a single-electron device was proposed by Tougaw, Lent, and Porod called the “Quantum Cellular Automata (QCA)” [13] that can be used for logic applications. The basic cell of a QCA is made up of four quantum dots in a form of a square closely placed to each other for tunneling to occur. If the cell is filled with two electron, due to the Coulomb repulsion between the electrons they will occupy the two quantum dots furthest from each other, hence the dots diagonally opposite of each other. And because there are two diagonals which will have equal probability of being occupied and having opposite polarizations, one of them can resemble a logic “0” while other a logic “1” (See Fig. 1.4).
Now consider an array of cells placed next to each other, the electrons in each cell will occupy the respective dots in which they are electrostatically favorable. This arrangement can act as wire and can be used to transfer a logic state. Usually in such a configuration the first cell in the wire is called the *driver cell* while the subsequent cells are the *driven cells*. According to the input to the first cell, all other following cells will rearrange their electrons to minimize the electrostatic energy. From that point onwards, it was clear that more logic functions could be implemented. One of them was the ”*NOT gate*” while another was the ”*majority gate*”. The majority gate takes 3 inputs with the output being equal to the majority of the inputs (2 out of the 3). The ”*AND*” and ”*OR*” functions were implemented by the majority gate and with that alongside the NOT gate, all Boolean functions were realized by the QCA. Fig. 1.5 shows the different orientation of the QCA cells used to realize what has been stated earlier.

![Diagram showing QCA Boolean functions](image)

**Fig. 1.5: Different QCA Boolean functions**
Since these structures look a lot similar to dominos, sometimes QCAs are referred to as domino cells. Other than electrostatic interactions controlling the placement of the electrons, some studies investigated magnetostatic interactions [14] or mechanical interactions [15].

Single-electron devices offer numerous advantages including low power consumption, high scalability, and high integration density. Nevertheless, the major challenge is making these devices operate normally at room temperatures. Because of their nanometer sizes, the operating temperature has to be lowered below 300K in order to investigate and control the quantum behavior of the electrons.

### 1.2.2 Novel Memory Devices

Data storage plays an integral part in the development modern computers and other electronic devices such as smartphones. The most famous and widely used data storage element in today’s electronics is the dynamic random access memory (DRAM). DRAMs are regarded as fast devices because they have data access times of typically 10ns, but their main drawback is that they are volatile storage elements which require that the stored data must be constantly refreshed. A single DRAM cell is composed of a capacitor which stores the data and a transistor driving this capacitive element. Fabricating such a device in order to comply with Moore’s law is becoming increasingly difficult with each technology node. With DRAMs now being in the sub-nanometer regime, leakage currents are becoming more and more prominent leading to the reduction in bit retention time. Furthermore, the need to incorporate large number of devices per chip is taking its toll on the lithography techniques in the fabrication process, and even if this was successfully handled the large packing densities are giving a rise to unwanted interference between neighboring cells.

Magnetic random access memories (MRAMs) which are based on the effect of magnetoresistance are being investigated as an alternative because they have the attribute of being non-volatile memories. Magnetic tunneling junctions (MTJs), which are MRAM based devices were developed in 1995 [16]. They consist of two ferromagnetic layers separated by an insulator, as shown in Fig. 1.6. Orthogonal conductors are placed above and below the ferromagnetic layers, with a transistor establishing electrical connection from the top contact to ground. The magnetic field of the bottom layer is fixed beforehand in one direction, while the top layer is left free to interchange directions under an applied field. If the magnetizations are in a parallel orientation, it is more likely that electrons will tunnel through the insulating layer than if they are in an anti-parallel orientation. Consequently, the MTJ can be switched between two states of electrical resistance, one with low and one with high resistance. Since such materials experience hysteresis, once a change to the direction of the orientation is achieved and the applied field is removed, the bit will remain in that
same state until a sufficient field of opposite polarity is applied, hence the non-volatility of this memory device. The resistance in a MTJ is dependent on the thickness of the tunneling barrier, and the resistance change $\Delta R/R$ has been reported to be in the range of 30%-50% [17, 18]. A more recent study proposed a double barrier MTJ with a ratio of 74% being reported [19].

Not only does the non-volatile character of the MRAMs make them an attractive candidate to compete with the DRAMs, but also the low voltage and power requirements since they rely on the direction of the magnetic moments. Other advantages include high density fabrication, high speed operations, and unlimited read and write operations without noticeable device degradation.

Fig. 1.6: Structure of a magnetic tunneling junction MRAM based memory cell

In early 2000, IBM introduced whole new concept for storing data at high speed and with ultrahigh density. Rather than the conventional approach used in DRAMs or the tunneling mechanism in MRAMs, this new concept was built around the idea of thermo-mechanical deformation of materials acting as storage media [20]. It was called the “Millipede”, and its concept structure is illustrated in Fig. 1.7. A polymer is placed on top of a silicon substrate with a photoresist sandwiched between them. This configuration is then topped by a 2D array chip consisting of many tiny atomic force microscope (AFM) cantilever tips. Tips can work simultaneously together or can be controlled individually. Digital information is represented by either the presence or absence of a pit on the polymer surface. For the former this was equivalent to a “1” while for the latter it meant a “0”. In the writing process, the AFM tip is heated to 400°C and then impacted with the polymer.
The polymer melts as the tip is pressed at which the temperature is then reduced, effectively forming a pit. Bits of 40nm in diameter were successfully created with a pitch of 120nm [21] (i.e. distance between two bits), as shown in Fig. 1.8. This implies that storage densities can reach 400 GB/inch².

As for the reading process, the tip is heated at 350°C. At this temperature the polymer does not melt, instead a temperature gradient is established between the tip and the polymer. Whenever a tip falls into a pit, some increased heat transfer occurs as both the tip and surface experience a greater contact area. The decrease in the temperature of the tip can be translated as a decrease in
its electrical resistance, hence indicating a “1” is read. The same thing goes for reading a stored “0” since the change in temperature and hence the resistance is dependent of the distance between the tip and the underlying surface. Finally, if the temperature is raised again to that of the writing process the polymer will start to flow again and will retain its original shape (which is also indicative of erasing and re-writing).

The Millipede does have some drawbacks, with the main concern being the durability of both the polymer and the tip. With the constant deformation of the polymer it will wear out over time and the same thing applies for the tip with the constant impaction and removal alongside the heating process. Furthermore, the complexity of fabricating such a device on a large scale for commercial use took away from its competitive edge when compared with other storage elements, of which during that time were still faring well. Nevertheless, the Millipede represents an enticing new technology with the promise of both rapid read/write rates and incredibly high bit densities giving it a bright future.

### 1.2.3 Carbon Nanotubes

In the search for solutions that would overcome the stumbling blocks of silicon technology in the nanometer regime, this has led to the discovery of carbon nanotubes (CNT) by Iijima in 1991 [22]. Carbon nanotubes are made up of graphene sheets rolled into cylinders. These cylinders come in 3 different shapes,

- **Armchair** ($n = m$), Fig. 1.9(b)
- **Zig-Zag** ($m = 0$), Fig. 1.9(c)
- **Chiral** ($n \neq m$), Fig 1.9(d)

$n$ and $m$ are integers denoting the number of unit vectors along the two directions of the honeycomb lattice of a graphene sheet as depicted in Fig. 1.9(a). An interesting property of graphene is that the conduction and valence bands touch along the Fermi level at 6 points (illustrated in Fig. 1.10) leading to CNTs being either metallic or semiconducting. From the 3 shapes stated previously, the armchair CNT shows a metallic behavior because the allowed modes (states) are present at the points where the conduction and valence bands touch. As for the zig-zag CNT, it acts as a semiconductor since the allowed modes are not present at these points hence a band gap appears
for such a structure. For chiral CNTs it can be either semiconducting or metallic depending on the values of $n$ and $m$. 

Fig. 1.9: Illustration of the unit vectors in the graphene honeycomb lattice, and the different types of CNTs

Fig. 1.10: Bandstructure of Graphene
There are several methods in synthesizing and fabricating CNTs. One of them is by *arc discharge* [22, 23]. In this method, two graphite electrodes are placed a millimeter apart with typical temperatures around 3000K. An arc is generated across the electrodes by a large dc current in a helium atmosphere. During this process, a deposited carbon rod builds up at the cathode of which the nanotubes are formed in a bundle called the “bucky bundle” made up of single wall nanotubes (SWNT) and multi wall nanotubes (MWNTs). The disadvantage of this method is there is also a formation of undesired amorphous carbon making it an unclean process. Another method which avoids the high electric fields involved in the arc discharge process, is the *laser ablation* [24] method. A pulsed laser vaporizes a graphite target in a high temperature reactor while an inert gas is bled into the chamber. Nanotubes develop on a water cooled surface as the vaporized carbon condenses. This method produces primarily SWNTs with good purity but is rather expensive. SWNTs were also grown by chemical vapor deposition (CVD) [25]. Nanotubes were grown on silicon wafers having islands of catalytic nanoparticles. The CVD process involves the dissociation of hydrocarbons (main carbon precursor) molecules on the catalyst. Methane gas was used as a source for carbon but it was also found that other carbon precursors including carbon monoxide, ethylene and benzene could be used [26].

Electronic applications CNTs rely on selective production of semiconducting or metallic nanotubes. It was crucial to find a way in separating the semiconducting and metallic nanotubes since during their growth both types are produced. One technique of separation relies on density gradient centrifugation [27]. By using structure discriminating surfactants nanotubes with different densities were separated. For example, nanotubes with smaller diameters are more buoyant and settle at higher points of the centrifugation tube while nanotubes with larger diameters settle at lower points of the centrifugation tube. This separation is shown in Fig. 1.11.

![Figure 1.11: Separation of CNTs in a centrifugation tube, taken from [27]](image-url)
Chapter 1: Introduction

1.2 Nanotechnology

Carbon nanotubes exhibit a lot of very good physical and electrical properties. Several studies have shown a Young’s modulus for both SWNTs and MWNTs ranging between 1.25 TPa to 1.47 TPa [28], while tensile strength ranged from 13 to 53 GPa [29] making carbon nanotubes one of the strongest and stiffest materials. Furthermore, CNTs we proposed as interconnects since today’s widely used copper is suffering from dimension scaling with noticeable increase in resistivity [30] and electromigration [31]. Ballistic transport in CNTs have been reported with current densities up to $10^{10}$ A/cm² [32, 33] compared to copper which fails at current densities of approximately $10^7$ A/cm². CNT transistors (CNTFETs) we also explored with the nanotube acting as the channel between the source and drain regions. Both-bottom gate [34] and top-gate [35] transistors were investigated exhibiting very promising I-V characteristics similar if not better to their silicon counterparts.

1.2.4 Organic Electronics

Organics have long been attractive for use in electronics because of their lightweight and flexibility, but the main advantage is the low cost and far simpler fabrication techniques. Unlike conventional silicon technology which undergoes high temperature, high vacuum deposition processes and complicated lithography techniques, organic material fabrication processes involve spin coating, spray coating, ink-jet printing, or screen printing.

Fig. 1.12: Different fabrication techniques of organic devices
An example of organic devices is the *organic thin film transistor (OTFT)*. OTFTs are similar to their silicon counterparts, but instead of using amorphous silicon they employ molecules or polymers which can come in either p-type [36, 37] or n-type [38, 39]. Conduction in silicon TFTs rely on minority carriers to create an inversion layer for current to pass, but in OTFTs conduction relies on majority carriers, hence no inversion layer is created. Furthermore, in order for organics to compete with amorphous silicon, their mobility should be higher than 0.1-1 cm²/V.s and also exhibit high on/off current ratios. Indeed that has been the case with mobilities between 1-10 cm²/V.s [40] and current ratios as high as 10⁶ [41] have been reported.

OTFTs come in two different types of device configuration and they look quite similar to the MOSFET architecture (a three terminal device). Both are common in that they have a substrate/gate/dielectric stack. The difference comes in the source/drain patterning and organic material deposition. In the top contact architecture (Fig. 1.13(a)) the organic material is deposited first and then the source and drain are patterned on top. As for the bottom contact (Fig. 1.13(b)) architecture the contacts are defined first and then the organic layer is deposited at the final stage. Although both configurations are being used, it has been suggested that top contacts have higher performance measures due to the smaller area of contact between the source/drain electrodes and the organic material leading to lower contact resistance and higher current drive due to improved carrier transport [42].

OTFTs have been demonstrated in variety of applications. Such applications include electronic paper [43, 44], radio frequency identification cards (RFIDs) [45] and even as sensors [46, 47]. Furthermore, what is really appealing about these OTFTs is that they can be fabricated on a variety of substrates other than silicon. Glass or polyimide (which is flexible) are such examples. But the main market in which thin film transistors are widely used, are as driving transistors controlling other devices. In that specific area organic TFTs lack the competitiveness as compared to their
silicon counterparts due to their lower switching speeds. Nevertheless with all the interest being taken in OTFTs this might change in the near future.

Alongside the OTFT is another device built on organic material that has actually reached maturity and is being employed commercially on a large scale for quite some time. This is the organic light emitting diode (OLED). OLEDs are now used in the car industry, for example as headlights or tail lights. More commonly they are used in today’s smartphones as part of the display. One example is Samsung’s famous active matrix organic light emitting diodes (AMOLEDs). These devices rely on electron-hole recombination. These recombinations occur either through transitions of singlet excitons or triplet excitons to the ground state respectively. Singlet excitons are spin anti-symmetric which are quantum mechanically allowed and therefore are fast (1 - 10ns). Triplet excitons on the other hand do not conserve spin and therefore are not allowed and have slower transition times (100μs - 10s). Because the resulting spin is of a random nature, from a statistical point for every singlet transition there are 3 triplet transitions, leading to an internal quantum efficiency of the OLED to be 25%. But this issue can be solved by incorporating phosphorescent guest molecules into the organic material and achieve an efficiency close to unity [48, 49]. Such molecules exhibit spin-orbital coupling facilitating intersystem crossing between singlet and triplet states, hence contributing to radiative decay for both singlet and triplet excitons.

The OLED’s efficiency also depends on its structure. The basic structure of an OLED consists of an anode (for hole injection) and a cathode (for electron injection) and an organic material in between. But this simple configuration does not provide the required efficiency since there is a difference in electron and hole mobilities, making the active region at which the excitons are formed to be closer to the electrodes rather than in the middle. Furthermore, carriers with the higher mobility can reach the opposite electrode without being captured. More layers had to be incorporated to solve these problems. With the addition of so-called “Hole Transport Layers (HTL)” and “Electron Transport Layers (ETL)”, this improved the OLED’s performance. Both layers have large HOMO-LUMO gaps than the “Light Emitting Layer (EML)” (situated in the middle of the device). They maintain the flow of holes and electrons into the EML layer without putting a huge emphasis on differences in mobility and also act as barriers to prevent injection of either a hole or an electron into the opposite electrode. This configuration makes the OLED a multi-heterojunction structure. Typical materials used for these layers are listed below and the structure along with the band diagram is shown in Fig. 1.14.

- **Cathode**: Metal Alloy (Typically Mg-Ag or Li-Al)
- **Anode**: Indium Tin Oxide (ITO)
• **Hole transport layer (HTL):** N,N’-diphenyl-N,N’-bis(3-methylphenyl)-1-1’ biphenyl-4,4’ diamine (known as TPD) (Fig. 1.15(a))

• **Light emitting layer (EML):** Depends on the desired color of emitted light. For green, tris (8-hydroxyquinoline) aluminum (Alq3) (Fig. 1.15(b))

• **Electron transport layer (ETL):** Different materials but can also be Alq3 (Fig. 1.15(b))
OLEDs can be either bottom emission or top emission devices. Bottom or top does not refer to the orientation of the OLED, rather to the direction of the emitted light. For active matrix displays top emission is preferred because it makes them easier to integrate with backplane TFTs which are typical not transparent [50]. Moreover, there are 4 different schemes in which OLEDs can emit light (all shown in Fig. 1.16) [51]. The basic one (Fig. 1.16(a)) includes 3 separate structures each emitting different colors. Another scheme (Fig. 1.16(b)) involves the filtering of white light with the downside of lower efficiency due to the large amount of light absorbed by the filters. Using blue OLEDs (Fig. 1.16(c)) instead of white light and prepatterned color changing media (CCMs) improves the overall efficiency. As for the final scheme (Fig. 1.16(d)) it uses stacked OLEDs with each one emitting different colors at the same time. Stacked OLEDs that were demonstrated by [52, 53] offers a more compact configuration and color depth compared to the side by side OLEDs.

Fig. 1.16: Different types of OLED configurations
1.2.5 Molecular Electronics

Molecular devices, which have been synthesized and investigated in the last decades, are described as an organic or inorganic molecule (or even multiple thin layers of them), bridging two contacts [54]. This field of research is very promising due to the fact that it involves structures on a nanometer scale, which of course is the main aim for the pursuit of lower critical dimensions. Furthermore, the idea of tuning the electronic properties of a device by changing the chemical structure, is what makes this field more and more appealing.

Of course the main target was to gain an understanding of charge transport through these molecules, which is dependent on several factors. Such factors include the type of molecule itself. One is an insulating chain consisting of σ-bonds, in which atoms are strongly bonded together and the molecule is in a stable state. Another is a conductive chain consisting of an alternating sequence of single and double bonds leading to a formation of a π-system that is characterized by weak orbital overlap. The electrons involved are not strongly tight to a specific atom and can be considered delocalized, hence contribute to an electric current once a voltage is applied. Another factor that must be taken into consideration are the electrodes used to contact the molecule, because the electronic properties of a molecule connected to an electrode is different from a molecule in solution. Electrodes can be metallic, usually gold, where the molecules are bonded with the aid of a thiol endgroup. This gives mechanical stability to the molecular chain, but breaks the electron conjugation which is formed by the overlap of the inner conducting orbitals of the molecule and the outer metallic electronic states. Semiconducting substrates can also be used as electrodes, mainly with molecules consisting of hydroxyl endgroups. The advantage of this approach is the use of materials from conventional microelectronics and this could form a bridge between the fields [55].

Even though this field is still in its research phase, different experimental setups have been carried out to implement various functions. Some of these functions will be discussed briefly in the following sub-sections.

1.2.5.1 Molecular Diodes

The first approach to molecular electronics was taken by Aviram and Ratner in 1974 [56]. They proposed a π-system made up of a donor and acceptor linked together by an insulating spacer, which is contacted on both ends by a metal (See Fig. 1.17(a)). Creating the donor and acceptor is done by adding substituent groups that either donate or withdraw electrons, hence raising or lowering the energy level respectively.
At zero bias the electrons in the metal contacts are filled up to the Fermi level. Once a positive voltage is applied (forward bias - Fig. 1.17(b)), the potential on the left is raised while on the right side is lowered. At that point current will start to flow from the left electrode to the LUMO of the acceptor, then tunnel through the barrier towards the HOMO of the donor and further into the right electrode. If a negative voltage is applied (reverse bias - Fig. 1.17(c)) both the donor and acceptor levels are separated further apart, that eventually no current will flow expect at much higher voltages. This is a behavior of a diode that favors current flow in one direction.

![Diagram](Fig. 1.17: (a) Aviram-Ratner single molecular rectifier, (b) device under forward bias, (c) device under reverse bias)

### 1.2.5.2 Molecular Storage Elements

There are some molecules that can be stable in two different states. These so-called metastable molecules will have differences in physical properties, such as conductance. If the difference in conductance is high to the extent that a distinction can be observed for a “0” state and a “1” state, these molecules can be used as storage elements. Experimental analysis reported in [57], show that molecules can exhibit the same function of a random access memory. In this study, one of the molecules used consisting of nitro and amine functionalities, was sandwiched between two gold...
contacts. The configuration of the devices along with the direction of current flow for the write and erase operations are depicted in Fig. 1.18.

![Illustration of a molecular storage element, taken from [57]](image)

At first the molecule is at an initial state of low conductivity. Under successive positive voltage sweeps, the conductivity state changes from low to high (indicative of a write operation), hence a bit is stored. With each successive sweep the I-V characteristics are identical to the previous values (i.e. read operation). On the other hand once the device is swept under a reverse voltage the characteristics revert back to the initial state (indicative of an erase operation).

Fig. 1.19 illustrates the I-V characteristics for either a stored “0” or “1” and their difference as a function of temperature. Moreover, it was also reported that the operations were repeatable, no device degradation was observed, and bit retention time of approximately 800s at 260K. All of these are promising signs towards the ability of using molecular devices as memory storage elements.
Chapter 1: Introduction

1.2 Nanotechnology

1.2.5 Three Terminal Devices

Research in two terminal devices showed very interesting and useful outcomes, but what about a three terminal device? Once a conducting molecule is placed between two contacts, the source and drain, an additional electrode can be introduced as a gate in the effort of achieving a field effect transistor-like behavior. There are several possibilities on how to incorporate the gate as the third electrode.

**Back Gate:** A molecule attached to the source and drain is placed on a heavily doped silicon substrate acting as the back gate with a layer of SiO$_2$ between them for the gate oxide. This configuration (see Fig. 1.20(a)) is similar to that of the TFTs (Thin Film Transistors). One of the challenges in fabricating such a device is the nanometer separation between the two electrodes in which the molecule(s) should be placed. Several techniques have been proposed such as electromigration [58] and scanning tip based fabrication [59]. Other methods included electron beam lithography and shadow masking [60]. Typical gaps were reported to be between 2 - 4 nanometers. But the fact is due to the sophisticated fabrication processes and the ability of making it reproducible on a large scale, the back gate approach is not widely spread.

**Electrochemical Gate:** In this approach, the molecule is bridged between the source and drain respectively in an electrochemical environment. A large gate field is achieved using an electrochemical gate in which the gate voltage is applied between the source and a gate electrode also inserted in the electrolyte [61], as illustrated in Fig. 1.20(b). The gate voltage leads to a shift of the

---

*Fig. 1.19: I-V characteristics observed by M. A. Reed et al. [57] for either a stored "0" or "1" (left) and their difference with respect to temperature (right)*
energy levels of the molecule towards the electrode Fermi levels. Once resonance is achieved, tunneling occurs contributing to a current. This approach is attractive because it employs “redox” molecules which can switch between oxidized and reduced states, hence leading to transistor-like behavior such as on-off switching, negative differential resistance (NDR), and rectification. Another good feature associated redox molecules is that the highest occupied molecular (HOMO)/lowest occupied molecular orbital (LUMO) are not strongly off resonance with the electrode Fermi levels as compared to non-redox molecules [62].

![Fig. 1.20: (a) Back gate configuration (b) Electrochemical gate configuration](image)

### 1.2.5.4 Self-Assembled Monolayers (SAMs)

Self-assembled monolayers (SAMs) play a huge part in the field of nanoelectronics. Even though a lot of advancement has been done in lithography techniques, self-assembly present numerous advantages. Firstly, as the name suggests, self-assembly involves the formation of molecular monolayers by spontaneous adsorption onto a surface in a highly ordered manner. Another advantage is the tendency to produce relatively defect free structures due to the requirement of thermodynamic stability of the target structure [54]. Furthermore, with SAMs it is possible to tailor the interfacial properties of the surface which the monolayers are attached to [63]. Other advantages include the ease of preparation and the possibility of lateral structuring.

There are two ways in fabricating SAM structures. The first is by solution deposition (Fig. 1.21), which is the most commonly used. The desired molecules are prepared is a specific solution. Afterwards the substrate in question is then immersed into the solution for a period of time at which the molecules will start to assemble on the substrate. The second is by gas phase deposition.
Here the substrate is placed in a UHV chamber so that it will be clean from any impurities. Molecules are then deposited onto the surface of the substrate at a controlled rate from a container attached to the UHV chamber.

![Fig. 1.21: Illustration on how self-assembled monolayers are formed on a substrate by solution based deposition](image)

What makes SAMs even more promising is the variety of applications that in which they can be utilized. Studies showed that SAMs can be incorporated into organic light emitting diodes (OLEDs) to enhance carrier injection [64, 65], or in organic thin film transistors (OTFTs) as a gate dielectric due to its insulating properties [66]. SAMs where also investigated as a candidate for resists in electron beam lithography [67]. To go even further, they were also proposed as biosensors to detect molecules such as cholesterol or glucose, or even to identify infections and diseases [68, 69].

Even though SAMs have been investigated in different areas, there was still the ever growing need to achieve what would be regarded as one of the most significant breakthrough in nanoelectronics; that is a “self-assembled monolayer field effect transistor (SAMFET)”, a device that can mimic the operation of a FET. In fact over the past few years a lot of research has been done on SAMFETs. In these devices the monolayers are used in the channel region for conduction. Different architectures of SAMFETs are illustrated in Fig. 1.22.
There are several factors that must be taken into consideration when dealing with such a device. Firstly, since charge transport is limited to the first layer at the dielectric-semiconductor interface \([70, 71]\), and the presence of structural imperfections such as voids or grain boundaries in such a layer (that is comparable with the accumulation layer in thickness) \([72]\), highly ordered and densely packed SAMs were of critical importance. This depends on the orientation of the layers themselves, where flat \(\pi\)-conjugated units require a 2D stacking orientation \([73]\), while others were less sensitive to the molecular orientation such as fullerenes \((C_{60})\) due to their spherical molecular shape \([74]\). Another factor is the choice of both the dielectric material and the anchor groups. Anchor groups are responsible for the adhesion of the SAMs to the dielectric. Depending on the dielectric material different groups can be used. Carboxylic acids or phosphonic acids were used on aluminum oxide \([75, 76]\). Others used chlorosilane groups on silicon dioxide \([77]\). Processing temperatures also play a role in fabricating SAMFETs. Organics are usually handled at low temperatures, while substrate preparation and electrode depositions are sometimes done at high temperatures. Therefore processing temperatures must be handled with care.

As for electrical characterizations of SAMFETs, some promising results have been reported recently for different types of SAMs. For example, in \([78]\) both p-type and n-type based SAMFETs
were fabricated and characterized for different channel lengths. As for the p-type SAMFET, hole mobility of \(10^{-5}\ \text{cm}^2/\text{Vs}\) and \(\text{ON/OFF}\) current ratio of \(10^3\) were reported. For the n-type SAMFET, electron mobility of \(10^{-4}\ \text{cm}^2/\text{Vs}\) was measured and the device had an \(\text{ON/OFF}\) current ratio of \(10^4\). As the channel lengths were increased, both hole and electron mobilities did not change but on the other hand this increase in length had a negative effect on the \(\text{ON/OFF}\) current ratios with both devices dropping to \(10^2\). Another research study investigated n-type SAMFETs of which flexible polymer based substrates were used [79]. In this study electron mobilities as high as \(10^{-4}\ \text{cm}^2/\text{Vs}\) and \(\text{ON/OFF}\) ratios in the order of \(10^5\) were obtained.

Of course there are other numerous studies that have been carried out obtaining similar results or maybe even better, that all depends on the different materials used for the SAMs, substrates, and dielectrics. The reason the above studies were highlighted, was not only to give an overview about the electrical characteristics, but also to show that both p- and n-type transistors were successfully fabricated paving the way for CMOS-based SAMFETs. Adding to that the success of using flexible substrates which of course has its advantages as compared to silicon substrates.

SAMFETs have just been under investigation recently, so there is still a lot to learn about such devices. Since there are a lot of factors each playing a role of which the device operation depends on, it is a good idea to separate the devices into more simpler pieces that can be investigated separately. That will be the purpose of this thesis, in which only a single molecular chain will be considered in order to know how this chain binds to the substrate. Furthermore transport analysis will be carried out to understand the nature of conduction. Once we have a clearer picture, results can then be put into context when dealing with a more complex configuration consisting of several chains at once.
Chapter 2: Electronic Structure Theory

2.1 Electronic Structure Problem

The first step in studying molecular structures is finding the atomic arrangements with the lowest total energy and hence the most stable state. In doing so we must solve the non-relativistic time independent Schrödinger equation.

\[
\hat{H}\Psi(\{R_A\},\{r_i, \omega_i\}) = E\Psi(\{R_A\},\{r_i, \omega_i\})
\]  \hspace{1cm} (2.1)

For a system consisting of \( M \) nuclei and \( N \) electrons, the many body wave function (\( \Psi \)) is a function of all spatial coordinates of nuclei (\( \{R_A\}, \text{where} \ A = 1, 2, \ldots, M \)) and spatial and spin coordinates of electrons (\( \{r_i, \omega_i\}, \text{where} \ i = 1, 2, \ldots, N \)). The Hamiltonian is the sum of all possible interactions between nuclei and electrons, given by

\[
\hat{H} = -\sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 - \sum_{A=1}^{M} \frac{1}{2M_A} \nabla_A^2 + \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{1}{|r_i - r_j|} + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{|R_A - R_B|} \\
- \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{|r_i - R_A|}
\]  \hspace{1cm} (2.2)

In Eq. (2.2), \( M_A \) is the ratio of the mass of nucleus \( A \) to the mass of an electron. \( Z_A \) is the atomic number of nucleus \( A \), while \( \nabla_i^2 \) and \( \nabla_A^2 \) are the Laplacian operators. The first and second terms are the kinetic energies of all electrons and nuclei respectively. The third and fourth terms represent the Coulomb repulsion between electrons and between nuclei. Finally, the fifth term is the Coulomb attraction between electrons and nuclei.

It is important to note that atomic units are employed here, where the charge \( e \), electron mass \( m_e \), and Planck’s constant \( \hbar \) are all unity. As a consequence all energies are given in Hartree (1H = 27.2114 eV) and all distances are given in Bohr radii (\( a_0 = 0.529\text{Å} \)).

Even though, Eq. (2.2) is quite easy to understand, it is very complex to solve and that is why several approximations have been introduced in order to reduce this complexity. These approximations are discussed in more detail in the following sub-sections.
2.1.1 Born-Oppenheimer Approximation

The Born-Oppenheimer approximation [80] plays an important role in electronic structure calculations. It is based on the fact that the mass of nuclei are much heavier than that of electrons. Even the proton, the lightest form of a nucleus is approximately 2000 times larger than an electron. Therefore nuclei move much more slowly than electrons. This realization allows the dynamics of electrons and nuclei to be separated, in which the nuclei are treated as fixed particles with respect to the moving electrons around them. Under this approximation, the second term (kinetic energy of nuclei) in Eq. (2.2) can be neglected and the fourth term (nuclei repulsion) becomes a constant. This constant value does not affect the wave function ($\Psi$) but shifts the value of the total energy. With the Born-Oppenheimer approximation Eq. (2.2) becomes

$$\hat{H}_{\text{elec}} = -\sum_{i=1}^{N} \frac{1}{2} \nabla_i^2 + \sum_{i=1}^{N} \sum_{j \neq i}^{N} \frac{1}{|r_i - r_j|} - \sum_{i=1}^{N} \sum_{A=1}^{M} \frac{Z_A}{|r_i - R_A|}$$

(Eq. 2.3)

Eq. (2.3) is called the electronic Hamiltonian, and thus the Schrödinger equation becomes

$$\hat{H}_{\text{elec}} \Psi_{\text{elec}}(\{x_i\}) = E_{\text{elec}} \Psi_{\text{elec}}(\{x_i\})$$

(2.4)

where both the spatial ($r_i$) and spin ($\omega_i$) coordinates are put together and described by one variable $x_i$ for simplicity as a function of the electronic wave function ($\Psi_{\text{elec}}$) respectively. Furthermore $R_A$ has been suppressed due to the fixed configuration of the nuclei. The total energy is the sum of electronic energy ($E_{\text{elec}}$) and the constant value representing the nuclei repulsion term, leading to:

$$E_{\text{tot}} = E_{\text{elec}} + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{|R_A - R_B|}$$

(Eq. 2.5)

To put it in a simple way how the Born-Oppenheimer approximation is useful, a benzene molecule consists of 12 nuclei and 42 electrons, hence the time independent Schrödinger equation that has to be solved in order to get the energy and the wave function of this molecule is expressed in terms of 204 variables; that is the spatial and spin coordinates of the electrons and the spatial coordinates of the nuclei. Under the BO approximation, if we are interested in solving the electronic Schrödinger equation, it will only be expressed by 168 variables (electrons only). On the other hand, if the effects of the quantum mechanical nuclear motion are to be studied, the computation must be
done in nuclear coordinates, hence the Schrödinger equation will only be expressed by 36 variables (nuclei only).

However, it is important to note that the Born-Oppenheimer approximation is not a universally valid one, and will breakdown when there are multiple energy surfaces close to each other or crossing each other. This give a rise to electronic and vibrational interaction, sometimes referred to as “vibronic coupling”, that cannot be neglected. Dissociative adsorption of molecules on metal surface is an example where such coupling takes place, therefore more caution must be taken into consideration when dealing with such systems [81, 82, 83, 84].

2.1.2 Hartree-Fock Approximation

From the Born-Oppenheimer approximation, the complexity of the Schrödinger equation was reduced but there still lies a major difficulty in solving Eq. (2.4), which is the interactions between electrons responsible for the quantum effects that needs to be understood. Imagine a system with a large number of N-electrons, we need to know the coordinates of each electron, and that requires a treatment of $4^N$ variables describing the x, y, z, and spin coordinates respectively. This makes the Schrödinger equations increasingly difficult to deal with. That is where the Hartree-Fock approximation comes in. It is the basis of molecular orbital (MO) theory, which posits that each electron’s motion can be described by a single-particle function (orbitals) which does not depend explicitly on the instantaneous motion of other electrons. Therefore, not only does it take the Born-Oppenheimer as a starting point, additionally it maps the N-electron Schrödinger equation into one electron Schrödinger-like equation.

In Hartree-Fock, the wave function is described by a single Slater determinant. The key essence of the single Slater determinant is that it obeys the Pauli Exclusion Principle. Electrons are fermions, and according to the exclusion principle if two sets of electron coordinates are interchanged, their wave functions must be anti-symmetric,

$$\Psi(x_1, \ldots, x_i, \ldots, x_j, \ldots, x_N) = -\Psi(x_1, \ldots, x_j, \ldots, x_i, \ldots, x_N)$$  \hspace{1cm} (2.6)

The Slater determinant simply employs a linear combination of the product of independent spin orbitals ($\chi_l$) with all possible combinations of electrons with their respective coordinates,
\[ \psi^{HF}(x_1, \ldots, x_j, \ldots, x_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_i(x_i) & \chi_j(x_i) & \cdots & \chi_N(x_i) \\ \chi_i(x_j) & \chi_j(x_j) & \cdots & \chi_N(x_j) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_i(x_N) & \chi_j(x_N) & \cdots & \chi_N(x_N) \end{vmatrix} \] (2.7)

In the Slater determinant, the rows denote electrons, while the columns denote the spin orbitals. To realize the anti-symmetry property of the Slater determinant, consider a two-electron example:

\[ \psi^{HF}(x_1, x_2) = \frac{1}{\sqrt{2!}} \begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) \\ \chi_1(x_2) & \chi_2(x_2) \end{vmatrix} \] (2.8)

Expanding Eq. (2.8) will give,

\[ \psi^{HF}(x_1, x_2) = \frac{1}{\sqrt{2!}} [\chi_1(x_1)\chi_2(x_2) - \chi_2(x_1)\chi_1(x_2)] = -\psi^{HF}(x_2, x_1) \] (2.9)

It must be stated as a caution, even though using a single Slater determinant is a good approximation, it is not enough to compute the “exact” solution of the Schrödinger equation as it requires working with several Slater determinants.

Now that we know the functional form for the wave function in Hartree-Fock, the next step is to examine the Hamiltonian to make it look as simple as possible. Firstly, we define a one-electron operator as follows,

\[ h(x_i) = -\frac{1}{2} \nabla_i^2 - \sum_A \frac{Z_A}{r_{iA}} \] (2.10)

Secondly, a two-electron operator defined as,

\[ v(x_i, x_j) = \frac{1}{r_{ij}} \] (2.11)

From Eq. (2.10) and Eq. (2.11), the electronic Hamiltonian is written as

\[ \hat{H}_{elec} = \sum_i h(x_i) + \sum_{i<j} v(x_i, x_j) \] (2.12)
As it was stated before, by using a single Slater determinant we get an approximation to the wave function. Therefore, we need to obtain the “best” approximate wave function and this is done by employing the variation principle. This principle states that we can vary the spin orbitals while at the same time enforcing the constraint that the orbitals are orthonormal (i.e., $\langle \chi_i(x_i) | \chi_j(x_j) \rangle = \delta_{ij}$, normalization of single particle orbitals), until we minimize the energy. Furthermore, the energy obtained will always be an upper bound to that of the true energy. This leads to expressing the energy in a symmetric fashion in terms of the wave function and the electronic Hamiltonian as:

$$E^{HF} = \langle \Psi^{HF} | H_{elec} | \Psi^{HF} \rangle$$

$$= \sum_{i=1}^{N} \int \chi_i^*(x_i) \left[ -\frac{\nabla_i^2}{2} + \frac{Z_A}{r_i} \right] \chi_i(x_i) dx_i$$

$$+ \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \int \chi_i^*(x_i) \chi_j^*(x_j) \frac{1}{|r_i - r_j|} \chi_i(x_i) \chi_j(x_j) dx_i dx_j$$

$$- \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \int \chi_i^*(x_i) \chi_j^*(x_j) \frac{1}{|r_i - r_j|} \chi_j(x_i) \chi_i(x_j) dx_i dx_j$$

(2.13)

The terms of Eq. (2.13) are explained as follows:

The first term,

$$\sum_{i=1}^{N} \int \chi_i^*(x_i) \left[ -\frac{\nabla_i^2}{2} + \frac{Z_A}{r_i} \right] \chi_i(x_i) dx_i = \langle \chi_i(x_i) | h(x_i) | \chi_i(x_i) \rangle$$

(2.14)

represents the kinetic energy and potential energy for the attraction between one electron and the nuclei. The second term called the Coulomb term represents the interactions between two electrons,

$$\frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \int \chi_i^*(x_i) \chi_j^*(x_j) \frac{1}{|r_i - r_j|} \chi_i(x_i) \chi_j(x_j) dx_i dx_j = \langle \chi_i(x_i) | J_j(x_i) | \chi_i(x_i) \rangle$$

(2.15)

where $J_j(x_i)$ is the Coulomb operator defined as,

---

1 Dirac Notation
\[ J_f(x_i) = \int \left| \chi_f(x_j) \right|^2 \frac{1}{|r_i - r_j|} dx_j \] (2.16)

The last term does not have a simple classical analogy. It arises from the antisymmetric requirement of the wave function. It looks much like the Coulomb term, except it exchanges the spin orbitals \( \chi_i \) and \( \chi_j \). Hence, it is called the exchange term.

\[ \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \int \chi_i^*(x_i) \chi_j^*(x_j) \frac{1}{|r_i - r_j|} \chi_j(x_i) \chi_i(x_j) dx_i dx_j = \langle \chi_i(x_i) | \mathcal{K}_j(x_i) | \chi_i(x_i) \rangle \] (2.17)

where \( \mathcal{K}_j(x_i) \) is the exchange operator in terms of its action on an arbitrary orbital \( \chi_i(x_i) \) and is defined as,

\[ \mathcal{K}_j(x_i) \chi_i(x_i) = \left[ \int \chi_j^*(x_j) \frac{1}{|r_i - r_j|} \chi_j(x_i) dx_j \right] \chi_i(x_i) \] (2.18)

In terms of the Coulomb and exchange operators the Hartree-Fock equation can be written in a more compact form.

\[
\left[ h(x_i) + \sum_{j \neq i} J_j(x_i) - \sum_{j \neq i} \mathcal{K}_j(x_i) \right] \chi_i(x_i) = \varepsilon_i \chi_i
\] (2.19)

where \( \varepsilon_i \) is the energy eigenvalue associated with orbital \( \chi_i \), and the Hartree-Fock equation becomes an eigenvalue equation. Moreover, considering that

\[
\left[ \sum_{j \neq i} J_j(x_i) - \sum_{j \neq i} \mathcal{K}_j(x_i) \right] \chi_i(x_i) = 0
\] (2.20)

then we can eliminate the restrictions \( j \neq i \) in the summations, and introduce a new operator, the Fock operator defined as,

\[ \hat{F}_i = h(x_i) + \sum_j J_j(x_i) - \mathcal{K}_j(x_i) \] (2.21)
Working through the above equations leads to a mapping of an N-electron Schrödinger equation into a one-electron Schrödinger-like equation, and the final form of the Hartree-Fock is expressed as

\[
\hat{F}_i \chi_i(x_i) = \varepsilon_i \chi_i(x_i)
\]  

(2.22)

Even though Eq. (2.22) is a linear eigenvalue equation, it needs to be solved iteratively, because the Fock operator is in terms of \( J_j \) and \( K_j \) respectively which in turn are a function of the orbitals. But we cannot get the orbitals without solving Eq. (2.22). Therefore, we have to start with an initial guess for the orbitals, get the Fock operator and then solve to obtain the energy and a new set of orbitals. This process is done iteratively until we reach self-consistency. That is why solving the Hartree-Fock equations is called a “self-consistent field” procedure.

2.2 Basis Sets

As it was discussed in the previous section, getting the best approximation for the wave function is important to solve the Hartree-Fock equation. This is essentially done by describing the molecular orbitals of the system in question. That is why in this section a brief overview will be given about the concept of “Basis Sets”. A basis set is defined as a set of known functions (i.e. typically atomic orbitals) used to represent unknown functions (i.e. molecular orbitals). There are two important criteria when it comes to selecting basis functions:

- The basis functions used to describe the wave function should be computationally tractable.

- They should be physically meaningful. For example, for bound atomic systems the function should go to zero when the distance between the nucleus and the electron becomes large.

2.2.1 Slater-Type Orbitals

Slater-Type orbitals (STOs) were previously used as basis functions in quantum chemistry because they showed similarities in the solutions compared to that of the hydrogen atom; that is the wave function exhibits a cusp at the nucleus and also decays exponentially at large distances from the nucleus.
STOs are defined as,

\[ \chi(\zeta, n, l, m, r, \theta, \varphi) = NY_{lm}(\theta, \varphi) r^{n-1} e^{-\zeta r} \]  

(2.23)

where \( N \) is the normalization factor, \( Y_{lm}(\theta, \varphi) \) is the spherical harmonics, \( r \) is the distance of the electron from the nucleus, \( \zeta \) controls the width of the orbital (large \( \zeta \) gives a tighter function, small \( \zeta \) gives a diffuse function), and finally \( n, l, m \) are the quantum numbers. STOs have an advantage in that they have direct physical interpretation and thus are naturally good basis for molecular orbitals, but from a computational point of view solving the required integrals needed in the course of the self-consistent field procedure (especially for two-electron integrals) are extremely slow and decrease the speed of computation.

### 2.2.2 Gaussian-Type Orbitals

In contrast to STOs, Gaussian-Type orbitals (GTOs) work much faster in solving these integrals, and that is why GTOs are still widely used until now as basis functions for the computation of molecular orbitals. GTOs are defined as,

\[ \chi(\zeta, l_x, l_y, l_z, x, y, z) = Ne^{-\zeta r^2} x^{l_x} y^{l_y} z^{l_z} \]  

(2.24)

Eq. (2.24) is quite similar to Eq. (2.23) except that in the exponential term the distance is squared and the GTOs are expressed in terms of Cartesian coordinates rather than spherical harmonics. The sum of the exponents \( L = l_x + l_y + l_z \) defines the function as an s-type \( (L = 0) \), p-type \( (L = 1) \), d-type \( (L = 2) \) and so on.

Continuing the comparison between Gaussian- and Slater-Type functions, GTOs do not exhibit the cusp and moreover the functions decay far too quickly as they moves further away from the nucleus. To overcome this problem, a linear combination of several “primitive” Gaussian functions each described by Eq. (2.24) can be grouped together to form a single “contracted” Gaussian function. This contraction makes them look more like Slater functions, as it is illustrated in Fig. 2.1. The improvement is easily noticed from using one Gaussian function to using a contracted one that for example is made up of three primitive Gaussians. The cusp starts to take shape and the exponential decay as compared to the Slater function is almost identical.
Chapter 2: Electronic Structure Theory

2.2 Basis Sets

Today there are hundreds of basis sets composed of GTOs. The smallest of these are called minimal basis sets. A minimal basis set is one in which a single basis function is used for each orbital in a Hartree-Fock calculation of the free atom. For example, the minimal basis sets for elements in the first row of the periodic table (i.e. Li, Be, B, C, etc.) consists of two s-functions (1s, 2s) and one set of p-functions (2p$_x$, 2p$_y$, 2p$_z$ respectively). These sets are referred to as “single-zeta (SZ)”. But the minimal bases sets are not flexible enough for accurate representation of orbitals, therefore we need to add more sets of orbitals. The next degree of accuracy is called “double-zeta (DZ)”, and for the same example we will then have four s-functions (1s, 1s’, 2s, 2s’) and two sets of p-functions (2p$_x$, 2p$_y$, 2p$_z$, 2p$_x’$, 2p$_y’$, 2p$_z’$). Of course the accuracy can go to the next level into “triple-zeta (TZ)”, “quadruple-zeta (QZ)” and so on.

An alternative approach is to use split-valence basis sets. This approach separates the core orbitals from the valence orbitals, in which single basis functions are applied to the core orbitals while multiple basis functions are applied to the valence orbitals. Core orbitals are relatively independent of the chemical environment and these states do not require to be very flexible in their description. On the other hand valence orbitals participate in a wide range of bonding environments and therefore require higher flexibility. Fig. 2.2 shows the difference in the number of functions needed as a comparison between normal basis sets and split valence basis sets approaches for a Beryllium atom.

Fig. 2.1: Comparison between 3 different levels of contracted Gaussian functions with a 1s Slater function. Taken from [85]
Fig. 2.2: Comparison between normal basis sets and split valence basis sets
Further improvements can be done to better describe the molecular orbitals. The shape of the orbitals of a free isolated atom is different to that if the atom is present in a chemical environment. Orbitals become distorted from their atomic shapes. This is called polarization and in order to describe this effect one needs to add polarization functions to the basis sets. In general to polarize an orbital with angular momentum $l$, mix it with another orbital with angular momentum $l + 1$. Hence, for example a double-zeta basis set becomes a “double-zeta polarized (DZP)”. One can even choose an accuracy of “double-zeta double-polarized (DZDP)”, and so on thereby giving more flexibility to the orbitals.

What about electrons that exist far from nucleus centers? Normal basis sets are inadequate to describe them. By employing so-called diffuse functions that are made up shallow Gaussians, they can accurately represent the tail portion of the orbitals that are distant from the atomic nuclei. Examples where diffuse functions are used include anions, weak bonds, and excited states.

Over the years, a large amount of basis sets have been proposed which include some or all of what has been mentioned above, with the most famous of them being the Pople-style [86, 87] and Dunning-style [88, 89, 90] basis sets. Other sets use plane waves and pseudopotentials to describe the orbitals.

One can try to use the different types of basis sets and implement the highest level of accuracy but will eventually reach the “complete basis set (CBS)” limit at which the energy will not go lower any further, but also must keep in mind that this has a severe computational cost as the size of the matrices increase accordingly with the size of the basis sets. In the end, it comes down to experience when choosing the right approach with a compromise being done between the desired level of accuracy and the computational power available.

### 2.3 Density Functional Theory

Density functional theory (DFT) presented a huge leap in computing the electronic structure of molecular and condensed matter. Despite the fact that Hartree-Fock (which is a wave function based method) is a good approach, the idea of dealing with $4^N$ variables was becoming increasingly difficult, especially for large systems. DFT on the other hand differs from Hartree-Fock in that it takes into consideration the density as the main quantity, and this simplifies it to a great extent because regardless of how many electrons are in a system the density will always be 3-dimensional. This is one of many reasons why DFT is presently the most successful and well established approach that is being used extensively in studying atoms, molecules, and solids. In this section, the
basic concepts underlying DFT and the features that have led to its widespread adoption will be presented. Before going into the details, let’s start by defining the electron density which is given by

\[ \rho(r) = N \int \cdots \int |\Psi(x_1, x_2, \ldots, x_N)|^2 dx_1 dx_2 \ldots dx_N \quad (2.25) \]

where \( \rho(r) \) determines the probability of finding any of the \( N \) electrons within a volume \( r \) but with arbitrary spin while the other \( N - 1 \) electrons have arbitrary positions and spin in the state represented by \( \Psi \).

### 2.3.1 Thomas-Fermi Model

The first approach taken towards computing the electronic energy in terms of the density was done by Llewellyn Thomas and Enrico Fermi in 1927 [91, 92]. Based on the uniform electron gas they proposed the following functional for the kinetic energy.

\[ T_{TF}[\rho(r)] = C_F \int \rho^3(r) \, dr, \quad C_F = \frac{3}{10} (3\pi^2)^\frac{2}{3} \quad (2.26) \]

From Eq. (2.26) it is obvious that the kinetic energy is only dependent on the electron density. Adding to that the classical treatment of the electron-nuclear attraction and the electron-electron repulsion we obtain the energy functional as

\[ E_{TF}[\rho(r)] = C_F \int \rho^3(r) \, dr - Z \int \frac{\rho(r)}{r} \, dr + \frac{1}{2} \int \int \frac{\rho(r_1) \rho(r_2)}{r_{12}} \, dr_1 dr_2 \quad (2.27) \]

In order to determine the correct density to be included in Eq. (2.27), they employed the variation principle, assuming that the ground state of the system is connected to the electron density for which the energy is minimized under the constraint of

\[ N = \int \rho(r) \, dr \quad (2.28) \]
The Thomas-Fermi model provides good predictions for atoms. However, this method is considered rather crude for more complex systems because it does not incorporate the actual orbital structure of electrons. It is regarded more of an approximation instead of an accurate model, but with the importance that the energy can in fact be derived using only the electron density.

### 2.3.2 Hohenberg-Kohn Theorem

Much of what DFT is based upon relies on the two theorems Hohenberg and Kohn put forward in 1964 [93]. The first Hohenberg-Kohn theorem states that the external potential $v_{\text{ext}}(r)$ is (to within a constant) a unique functional of the electron density $\rho(r)$. Let us assume that there were two external potentials $v_{\text{ext}}(r)$ and $v'_{\text{ext}}(r)$ differing by more than a constant, and each giving the same electron density $\rho(r)$ for the ground state. We would have two Hamiltonians $\hat{H}$ and $\hat{H}'$ whose ground-state densities were the same although the normalized wave functions $\Psi$ and $\Psi'$ would be different. Taking $\Psi'$ as a trial wave function for the Hamiltonian $\hat{H}$ problem we get,

$$E_0 < \langle \Psi'|\hat{H}|\Psi' \rangle = \langle \Psi'|\hat{H}|\Psi' \rangle + \langle \Psi'|\hat{H} - \hat{H}'|\Psi' \rangle = E'_0 + \int \rho(r) [v_{\text{ext}}(r) - v'_{\text{ext}}(r)] \, dr$$  \hspace{1cm} (2.29)

where $E_0$ and $E'_0$ are the ground states for $\hat{H}$ and $\hat{H}'$ respectively. Similarly, taking $\Psi$ as a trial wave function for the Hamiltonian $\hat{H}'$ problem,

$$E'_0 < \langle \Psi|\hat{H}'|\Psi \rangle = \langle \Psi|\hat{H}|\Psi \rangle + \langle \Psi|\hat{H}' - \hat{H}|\Psi \rangle = E_0 + \int \rho(r) [v_{\text{ext}}(r) - v'_{\text{ext}}(r)] \, dr$$  \hspace{1cm} (2.30)

Adding Eq. (2.29) and Eq. (2.30) we would obtain,

$$E_0 + E'_0 < E'_0 + E_0$$  \hspace{1cm} (2.31)

Eq. (2.31) shows an inconsistency, and so there cannot be two different potentials that can give the same electron density for the ground state. This is the proof that $\rho(r)$ uniquely determines the number of electrons $N$ and the potential $v_{\text{ext}}(r)$ and hence all the properties of the ground state.
Now we can write the total energy as,

\[ E[\rho(r)] = T[\rho(r)] + E_{en}[\rho(r)] + E_{ee}[\rho(r)] \]

\[ = \int \rho(r) v_{\text{ext}}(r) \, dr + F_{HK}[\rho(r)] \] \hspace{1cm} (2.32)

where

\[ F_{HK}[\rho(r)] = T[\rho(r)] + E_{ee}[\rho(r)] \] \hspace{1cm} (2.33)

The functional \( F_{HK}[\rho(r)] \) is in terms of the kinetic energy and the electron-electron interactions and it plays a central role in DFT. If it were known we would have solved the Schrödinger equation exactly, but since the complexities of the many-electron problems are associated with it, hence it requires further explanation. Furthermore, Eq. (2.33) is solely dependent on the density which makes it a universal functional, hence it applies equally well to a small atom as to a large molecule.

The second Hohenberg-Kohn theorem states that the functional \( F_{HK}[\rho(r)] \) that delivers the ground state energy of the system, delivers the lowest energy if and only if the input density is the true ground state density. This is nothing but the variation principle.

\[ E_0 \leq E[\rho'(r)] \] \hspace{1cm} (2.35)

In other words, this means that for any trial density which satisfies the necessary conditions (i.e. \( \rho'(r) \geq 0 \) and \( N = \int \rho'(r) \, dr \)) and is associated with some external potential, the energy obtained represents the upper bound of the true ground state energy. Recalling that any trial density defines its own Hamiltonian and hence its own wave function, this wave function can now be taken for the Hamiltonian generated for the true external potential. Thus,

\[ \langle \Psi'| \hat{H} | \Psi' \rangle = \int \rho'(r) v_{\text{ext}}(r) \, dr + F_{HK}[\rho'(r)] \]

\[ = E[\rho'(r)] \geq E_0[\rho(r)] = \langle \Psi_0'| \hat{H} | \Psi_0' \rangle \] \hspace{1cm} (2.36)

To reiterate, from Eq. (2.36) the energy will reach the minimum only when the electron density is the ground-state electron density.
2.3.3 Kohn-Sham Equations

Granted, from the Hohenberg-Kohn theorems, it was proved that the total energy can be obtained from the ground state density. But there was still the problem about how to get the density $\rho(r)$ needed for such calculation. Adding to that, the explicit form of the universal functional $F_{HK}[\rho(r)]$ which is in terms of the kinetic energy and the electron-electron interactions required further elaborations. This was achieved by Kohn and Sham in 1965 [94].

The first step towards finding a solution was that they realized the bad description of the kinetic energy according to the Thomas-Fermi model. They suggested that if the electrons are “non-interacting” in the Hartree-Fock sense, then one can just use the Hartree-Fock expression for the kinetic energy employing orbitals, hence it can be expressed as,

$$ T_s[\rho(r)] = \sum_{i}^{N} \left\langle \chi_{i} | -\frac{1}{2} \nabla^{2} | \chi_{i} \right\rangle $$

where the electron density is expressed in terms of the orbitals as,

$$ \rho(r) = \sum_{i}^{N} \sum_{s} |\chi_{i}(r,s)|^{2} $$

As for the second term in the functional $F_{HK}[\rho(r)]$, the electron-electron interactions $E_{ee}$ consists of the electron-electron repulsion (classical part) and the exchange term (non-classical part) that arises from the anti-symmetry.

$$ E_{ee}[\rho(r)] = \frac{1}{2} \iint \frac{\rho(r)\rho'(r')}{|r-r'|}drdr' + E_{exchange}[\rho(r)] $$

$$ = J[\rho(r)] + E_{exchange}[\rho(r)] $$

Kohn and Sham also stated that the kinetic energy from Eq. (2.37) is not the true energy because there is in fact a correlation (interaction) between the electrons and that there should be a correction accounting for this correlation. Taking into account that condition along with Eq. (2.37) and Eq. (2.39) the functional $F_{HK}[\rho(r)]$ becomes,

$$ F_{HK}[\rho(r)] = T_s[\rho(r)] + J[\rho(r)] + E_{xc}[\rho(r)] $$
where \( E_{xc}[\rho(r)] \) is called the *exchange-correlation* functional,

\[
E_{xc}[\rho(r)] = (T[\rho(r)] - T_s[\rho(r)]) + (E_{ee}[\rho(r)] - J[\rho(r)])
\]  
(2.41)

Eq. (2.41) contains the unknown part about the functional \( F_{HK}[\rho(r)] \) (basically the correlation) and will be discussed in further details in the following subsection. From here the total energy is nothing but the summation of the universal functional \( F_{HK}[\rho(r)] \) and the electron-nuclear interaction.

\[
E[\rho(r)] = T_s[\rho(r)] + E_{en}[\rho(r)] + J[\rho(r)] + E_{xc}[\rho(r)]
\]

\[
= \sum_{l} \langle \chi_i | -\frac{1}{2} \nabla^2 | \chi_i \rangle - \sum_{l} \int \sum_{A} \frac{Z_A}{|r - R_A|} |\chi_l(r)|^2 \, dr 
\]

\[
+ \frac{1}{2} \sum_{l} \sum_{f} \int \int |\chi_i|^2 \frac{1}{|r - r'|} |\chi_j|^2 \, dr \, dr' + E_{xc}[\rho(r)]
\]  
(2.42)

The only term to which no explicit form can be given is the \( E_{xc}[\rho(r)] \). The variation principle is applied to minimize the energy expression under the usual constraint (i.e., \( \langle \chi_i(x_i) | \chi_j(x_j) \rangle = \delta_{ij} \)). The result is the central equation in Kohn-Sham DFT which is the one-electron Schrödinger-like equation expressed as,

\[
\left( -\frac{1}{2} \nabla^2 + v(r) + \int \frac{\rho(r')}{|r - r'|} \, dr' + u_{xc}(r) \right) \chi_i = \varepsilon \chi_i
\]  
(2.43)

where \( \chi_i \) are the Kohn-Sham orbitals and \( v(r) \) representing the potential of the electron-nuclei interactions. The term \( u_{xc}(r) \) is the exchange-correlation potential and we also define a term \( u_{eff} \) for the effective potential, which both are give by,

\[
u_{xc}(r) = \frac{\delta E_{xc}[\rho(r)]}{\delta \rho(r)}
\]  
(2.44)

\[
u_{eff} = v(r) + \int \frac{\rho(r')}{|r - r'|} \, dr' + u_{xc}(r)
\]  
(2.45)

This allows Eq. (2.43) to be written in a more compact way,
\[
\left(-\frac{1}{2} \nabla^2 + v_{\text{eff}}\right) \chi_i = \varepsilon \chi_i \tag{2.46}
\]

Clearly Eq. (2.46) is a Hartree-Fock like single particle equation that needs to be solved iteratively because \(v_{\text{eff}}\) depends on the electron density. We begin with an initial guess of the electron density, then construct \(v_{\text{eff}}\) and get the Kohn-Sham orbitals. Based on these orbitals, a new density is obtained and the process is repeated until convergence is reached at which point the total energy will be calculated with the final value of the electron density.

### 2.3.4 Exchange Correlation

The difference between various DFT methods is the choice of the functional form for the exchange-correlation energy. It can be proven that the exchange correlation potential is a unique functional (as it was shown in the previous sub-section), valid for all systems, but an explicit functional form of this potential has been elusive so far. That is why at present time a lot of different approximations have been proposed with different levels of accuracy and computational complexity in the effort to approximate it. When expressing this functional in a written form the exchange and correlation terms are separated and given by,

\[
E_{\text{xc}}[\rho(r)] = E_x[\rho(r)] + E_c[\rho(r)] = \int \rho(r) \varepsilon_x[\rho(r)] \, dr + \int \rho(r) \varepsilon_c[\rho(r)] \, dr \tag{2.47}
\]

where \(\varepsilon_x\) (exchange) and \(\varepsilon_c\) (correlation) are the energies per particle respectively.

J. P. Perdew and K. Schmidt were credited for grouping the different flavors of the exchange-correlation functionals in what is referred to as “Jacob’s Ladder” [95], were at the bottom is the Hartree-Fock theory and as we go up one would expect an improvement in the accuracy for each higher step. Only the first few steps will be briefly discussed to give a perspective of what is the role of this functional is in DFT.

#### 2.3.4.1 Local Density Approximation

Local Density Approximation (LDA) is the simplest approximation for the exchange-correlation functional. At the center of this model is the idea of uniform electron gas. A system is divided up in equal infinitesimal portions with the same density and by applying the uniform electron gas
results on each portion then summing up over all space, we obtain the exchange-correlation energy. The exchange energy for LDA is given by the Dirac formula as,

\[ E_x^{\text{LDA}}[\rho(r)] = -C_x \int \rho^4(r) \, dr, \quad \epsilon_x^{\text{LDA}} = -C_x \rho^\frac{1}{3} \]

Eq. (2.48) is true for closed shell systems, but for open shell systems the \( \alpha \) and \( \beta \) densities (corresponding to the spin) are not equal thereby the approximation is then referred to as the Local Spin Density Approximation (LSDA), thus the exchange energy becomes

\[ E_x^{\text{LSDA}}[\rho(r)] = -2^{\frac{1}{3}}C_x \int \left( \frac{4}{\rho_\alpha^3(r)} + \frac{4}{\rho_\beta^3(r)} \right) \, dr \]

\[ \epsilon_x^{\text{LSDA}} = -2^{\frac{1}{3}}C_x \left( \rho_\alpha^\frac{1}{3} + \rho_\beta^\frac{1}{3} \right) \]  

On the other hand, no such explicit expression is known for the correlation part \( \epsilon_c[\rho(r)] \). However, thanks to Ceperly and Adler [96] highly accurate numerical quantum Monte-Carlo simulations of the homogenous electron gas are available. The values of the correlation energies from these simulations have been interpolated to provide an analytical form for \( \epsilon_c[\rho(r)] \), the Vosko-Wilk-Nusair [97] LDA functional is one example of that.

The LDA is applicable to systems with slowly varying densities but it does not apply for highly inhomogeneous systems such as molecules. In such complex systems, it tends to underestimate the exchange energy and overestimate the electron correlation. Studies showed binding energies were overestimated by 15%-20% while lattice constants were underestimated by 2%-3% for metals and insulators [98, 99, 100]. Due to these inconsistencies LDA was not sufficient enough and therefore more accurate approximations were required.

### 2.3.4.2 Generalized Gradient Approximation

Improvements over the LDA approach must consider a non-uniform electron gas. A step in this direction is to make the exchange and correlation energies dependent not only on the electron density but also on the derivatives of the density. In Generalized Gradient Approximation (GGA), it gives the value of the density a specific location and the direction of which it is changing with
respect to that location. This led to GGA yielding better results with respect to geometries and ground state energies for molecules and solids. The form of which the exchange energies in GGAs are given by is

$$E_{x}^{GGA}[\rho(r)] = \int \rho(r) \varepsilon_{x}(\rho(r)) F_{GGA}(x) \, dr \quad (2.50)$$

The same thing goes for GGA regarding open shell systems, where in that case one must account for the difference in densities of the respective spin. From Eq. (2.50), $F_{GGA}(x)$ represents a correction in the description of the exchange energy as compared to LDA for a given density, with $x$ being a dimensionless gradient variable given as,

$$x = \frac{\rho}{\rho^3} \quad (2.51)$$

There are various different exchange functionals including the very popular A. D. Becke (B88) [101] and Perdew-Burke-Ernzerhof (PBE) [102]. These functionals differ in the way they describe the correction term. This term is also a function of the gradient variable $x$, if this variable is equal to zero we revert back to LDA. Therefore generally speaking GGA will always offer a better representation of the exchange energy.

Furthermore, there are also several proposed functionals for the correlation energy, most notably the Perdew-Wang (PW86, PW91) [103, 104], and Lee-Yang-Parr (LYP) [105], of which all are based on interpolation. Usually both the B88 and LYP are applied together, but of course the numerous combinations between the exchange and correlation functionals stated can also be applied depending on the system at hand.
Chapter 3: Electronic Transport

3.1 The Quantum of Conductance

It is well known that the conductance of a macroscopic conductor is described by Ohm’s Law as

\[ G = \sigma \frac{A}{L} \]  \hspace{1cm} (3.1)

where \( A \) is the area of the cross section of the conductor, \( L \) is the length, and \( \sigma \) is the conductivity. From Ohm’s law this means that the conductance should vanish when the area becomes smaller and smaller and it becomes infinitely large when the length is very short. But that is not the case for atomic-size conductors because neither does the conductance vanish nor reaches infinity, an indication that Ohm’s law must breakdown. There are three important characteristics that define the size of a conductor when the macroscopic description is not applicable anymore [106]: (i) the de Broglie wavelength related to the kinetic energy of the electrons, (ii) the mean free path which is the distance that the electron travels without changing its momentum, (iii) the phase-relaxation length which is the distance the electron travels without altering its phase, that is in turn related to the inelastic scattering of the electron-electron and electron-phonon interactions. Phase relaxation contributes to so-called “non-coherent transport”, but since the focus of this thesis is on “coherent transport” phase relaxation transport regime will not be discussed further.

In order to have a full understanding about the conductance in atomic-scaled devices, a quantum mechanical treatment is required. Consider a one-level device under a voltage bias consisting of a single channel sandwiched between two contacts (i.e. source and drain) with electro-chemical potentials \( \mu_1 \) and \( \mu_2 \) respectively. One would argue that by coupling the channel more and more strongly to the contacts the conductance would continue increasing leading to a higher output current. But in reality the conductance eventually reaches a maximum value and the actual current observed is then below to that what was expected. This is due to “level broadening” which is related to the density of states (DOS) of both the channel and the contacts [107]. Before the coupling takes place to the source and drain we have a one sharp level in the channel and a continuous distribution of states in the source and drain contacts (see Fig. 3.1(a)). Once the coupling is done, the channel loses part of its state to the contacts, which is significant for the channel but which will not influence anything in the contacts. At the same time it gains states from the contacts over...
a range of energies and this leads to the broadening of the sharp level, as it is illustrated in Fig. 3.1(b). Once the level is broadened, parts of it will lie outside the $\mu_1 - \mu_2$ range, and these parts will not contribute to any current, hence the given proof of a maximum conductance.

Fig. 3.1: (a) A channel placed between source and drain contacts under a voltage bias, (b) Broadening of the energy level after the coupling process. Taken from [107]

### 3.2 Landauer Formalism

Now that we have a firm understanding about the conductance from the quantum mechanical point of view, there still the question about how the electrons actually flow from one contact into the channel and further into the other contact, or to describe it by a different word, how are they “transmitted”? Quantum mechanical formalism of electronic transport through one-dimensional devices was proposed by Rolf Landauer [108, 109]. He suggested a relation between the conductance and the transmission probability of an electron moving from one contact to the other.

Let us first concentrate on the contacts. Both the source and drain are filled up with electrons up to the Fermi level and align perfectly with each other at the point of equilibrium. If a voltage bias is applied the source and drain levels are separated from their equilibrium state by an amount of

$$\mu_1 - \mu_2 = qV_D \quad (3.2)$$

This gives a rise to two different Fermi functions for each contact.
Chapter 3: Electronic Transport  

3.2 Landauer Formalism

\[ f_1(E) = \frac{1}{1 + \exp[(E - \mu_1)/k_BT]} \]  \hspace{1cm} (3.3a)

\[ f_1(E) = \frac{1}{1 + \exp[(E - \mu_2)/k_BT]} \]  \hspace{1cm} (3.3b)

From this point each contact will try to bring the channel into equilibrium with itself. The source keeps pumping electrons into the channel while the drain keeps pulling them out. And since this equilibrium is never achieved, this constant adding and extracting of electrons contributes to a current flow. All of this was already a given to Landauer, but what he incorporated is that there lies a finite probability for an electron to be transmitted. The conductance according to Landauer is expressed by,

\[ G = \frac{2e^2}{h} \sum_{n,m} |t_{n,m}|^2 \]  \hspace{1cm} (3.4)

where \( t_{n,m} \) are the normalized transmission probability amplitudes between all right-going modes \( n \) and \( m \) in the left and right leads as depicted in Fig. 3.2. The factor 2 in Eq. (3.4) account for the spin degeneracy of the electrons.

Fig. 3.2: A two probe system at which the electrons flow from the left to the right according to the transmission probability with respect to the modes \( n \) and \( m \)

The idea is when an energy level comes in resonance with respect to an applied voltage, this gives a rise in the transmission probability and hence a rise in conductance. Further voltage application will then push this level out and the conductance will remain constant until another level comes...
into play at which the conductance will rise again. This resembles a step-like function for the conductance. From here the current can be expressed as

\[ I = \frac{2e}{h} \int_{-\infty}^{+\infty} T(E) [f_1(E) - f_2(E)] dE \]  \hspace{1cm} (3.5)

The maximum value of \( T(E) \) is 1 and it depends on the energy levels and the corresponding density of states present between both electro-chemical potentials. Finally, one must also take the temperature into consideration because not only does it define the Fermi functions for both the left and right contacts but also it has an effect on the shape of the broadening of the energy levels in between.

Calculating the transmission function and understanding how DFT (from what has been discussed in the previous chapter) relates to it is achieved by applying the Green’s function method, and in this method one must take into consideration the following quantities:

- \( \epsilon \rightarrow [H] \) \hspace{1cm} Hamiltonian matrix \hspace{1cm} (3.6a)
- \( \gamma_{1,2} \rightarrow [\Gamma_{1,2}(E)] \) \hspace{1cm} Broadening matrix \hspace{1cm} (3.6b)
- \( 2\pi \delta(E - \epsilon_k) \rightarrow [A(E)] \) \hspace{1cm} Spectral function \hspace{1cm} (3.6c)
- \( N \rightarrow [\rho] = \int \frac{dE}{2\pi} [G(E)] \) \hspace{1cm} Density matrix \hspace{1cm} (3.6d)

All of the above functions are in a matrix form because we are realistically dealing with a multi-level conductor. I will go through the equations leading to the formulation of the transmission function in a brief manner, therefore it would be advised to refer to [107] if a detailed explanation of all the derivations is required.

The Hamiltonian and wave function of a system can be divided into three parts, \( (H_1, |\Psi_1\rangle) \) for one contact, \( (H_2, |\Psi_2\rangle) \) for the other contact, and \( (H_d, |\Psi_d\rangle) \) for the device.

\[
\begin{pmatrix}
H_1 & \tau_1 & 0 \\
\tau_1^\dagger & H_d & \tau_2^\dagger \\
0 & \tau_2 & H_2
\end{pmatrix}
\begin{pmatrix}
|\Psi_1\rangle \\
|\Psi_d\rangle \\
|\Psi_2\rangle
\end{pmatrix} = E
\begin{pmatrix}
|\Psi_1\rangle \\
|\Psi_d\rangle \\
|\Psi_2\rangle
\end{pmatrix}
\]  \hspace{1cm} (3.7)

Where \( \tau_1 \) and \( \tau_2 \) describe the interaction between the device and contacts. The Green’s function is used in transport calculations because it gives a response of a system to a constant perturbation (i.e. \( H|\Psi\rangle = E|\Psi\rangle + |\nu\rangle \)). Furthermore using the Green’s function makes it easier than solving the
Chapter 3: Electronic Transport

3.2 Landauer Formalism

whole eigenvalue problem. This is due to the fact that we can calculate the device Green’s function \((G_d)\) separately without going into the complex procedure of solving for the whole system.

\[
\begin{pmatrix}
-E - H_1 & -\tau_1 & 0 \\
-\tau_1^\dagger & E - H_d & -\tau_2^\dagger \\
0 & -\tau_2 & E - H_2
\end{pmatrix}
\begin{pmatrix}
G_1 \\
G_{1d} \\
G_{2d}
\end{pmatrix}
= \begin{pmatrix}
I \\
0 \\
0
\end{pmatrix} \quad (3.8)
\]

Solving for only the device can be done by taking the three equations in the second column of Eq. (2.59). Hence we have

\[
(E - H_1)G_{1d} - \tau_1G_d = 0 \quad (3.9a)
\]

\[-\tau_1^\dagger G_{1d} + (E - H_d)G_d - \tau_2^\dagger G_{2d} = I \quad (3.9b)\]

\[-\tau_2 G_d + (E - H_2)G_{2d} = 0 \quad (3.9c)\]

Solving Eqs. (3.9a) and (3.9c) we get

\[
G_{1d} = g_1\tau_1G_d \quad (3.10a)
\]

\[
G_{2d} = g_2\tau_2G_d \quad (3.10b)
\]

where \(g_1\) and \(g_2\) are the Green’s function of the contacts in isolation. Substituting \(G_{1d}\) and \(G_{2d}\) into Eq. (3.9b) we obtain

\[
G_d = (E - H_d - \Sigma_1 - \Sigma_2)^{-1}, \quad \begin{cases}
\Sigma_1 = \tau_1^\dagger g_1\tau_1 \\
\Sigma_2 = \tau_2^\dagger g_2\tau_2
\end{cases} \quad (3.11)
\]

\(\Sigma_1\) and \(\Sigma_2\) are called \textit{self-energies}\(^2\) that arise from the coupling of the device to the contacts and can be viewed as a modification of the Hamiltonian of the device. These self-energies consist of both real and imaginary parts. The real part represents the energy shift of the channel due to the perturbation induced by the contacts while the imaginary part represents an energy broadening due to the device-contact coupling. This broadening is linked to the time an electron spends in the device region before being absorbed by one of the contacts.

\[
\Gamma_{1,2} = i(\Sigma_{1,2} - \Sigma_{1,2}^\dagger) = i\tau(g_{1,2} - g_{1,2}^\dagger)\tau^\dagger = \tau a_{1,2}\tau^\dagger \quad (3.12)
\]

\(^2\) Both the Hamiltonian and the self-energies are computed using DFT
After coupling occurs a current starts to flow due to the difference in the chemical potentials of both contacts. This current corresponding to incoming waves from the contact. In steady state, these waves are reflected back but also there is a retarded response of the whole system with respect to the incoming waves, therefore knowing the wave functions of the incoming waves gives us an idea about the change to the wave function of the device and the electron reservoirs filling the contacts. Solutions corresponding to an incoming wave are denoted as $|\Psi_{1,n}\rangle$ with $1$ referring to the contact number and $n$ representing the different modes. This wave will cause a perturbation of $-\tau_1^*|\Psi_{1,n}\rangle$ leading to a change in the wave function of the device expressed as

$$|\Psi_d\rangle = G_d\tau_1^*|\Psi_{1,n}\rangle$$  \hspace{1cm} (3.13)

It is also important to note that this perturbation will also lead to a change of the next incoming wave such that

$$|\Psi_1\rangle = (1 + g_1\tau_1 G_d \tau_1^*)|\Psi_{1,n}\rangle$$  \hspace{1cm} (3.14)

And finally the wave function in the second contact will be

$$|\Psi_2\rangle = g_2\tau_2 G_d\tau_1^*|\Psi_{1,n}\rangle$$  \hspace{1cm} (3.14)

Taking into consideration that in the steady state, the probability to find an electron on the device is conserved (and going through some mathematical steps which can be found in the previously stated reference), the current can then be expressed as

$$\frac{\partial}{\partial t} \sum_i |\Psi_i|^2 = \sum_i \frac{\partial \langle i|\Psi\rangle}{\partial t} \langle i|\Psi\rangle = \sum_i \left( \frac{\partial \langle i|\Psi\rangle}{\partial t} + \langle i|\Psi\rangle \frac{\partial \langle i|\Psi\rangle}{\partial t} \right) = 0$$  \hspace{1cm} (3.15a)

$$i_j = \frac{e}{\hbar} \left( \langle \Psi_j|\tau_j|\Psi_d\rangle - \langle \Psi_d|\tau_j^*|\Psi_j\rangle \right)$$  \hspace{1cm} (3.15b)

where $j$ is an arbitrary contact and $i_j$ is the current from the contact into the device. Consider an example where the current passes through the device from an incoming wave of energy $E$ in contact 1 ($|\Psi_{1,n}\rangle$) through the coupling defined by $\tau_2$, hence the current is:
Chapter 3: Electronic Transport

3.2 Landauer Formalism

\[ i_{2 \text{ from } 1} = \frac{e}{h} \left( \langle \Psi_2 | \tau_2 | \Psi_d \rangle - \langle \Psi_d | \tau_2^+ | \Psi_2 \rangle \right) \]
\[ = \frac{e}{h} \left( \langle \Psi_{1,n} | \tau_1 G_d^+ \tau_2^+ g_2 G_d \tau_1^+ | \Psi_{1,n} \rangle - \langle \Psi_{1,n} | \tau_1 G_d^+ \tau_2^+ g_2 \tau_2 G_d \tau_1^+ | \Psi_{1,n} \rangle \right) \]
\[ = \frac{e}{h} \left( \langle \Psi_{1,n} | \tau_1 G_d^+ \tau_2^+ (g_2^+ - g_2) \tau_2 G_d \tau_1^+ | \Psi_{1,n} \rangle \right) \]
\[ = \frac{e}{h} \left( \langle \Psi_{1,n} | \tau_1 G_d^+ \Gamma_2 G_d \tau_1^+ | \Psi_{1,n} \rangle \right) \]
(3.16)

Adding over the modes \( n \) and integrating over the energy range gives (note the factor 2 for spin):

\[ i_{2 \text{ from } 1} = \frac{2e}{h} \int_{-\infty}^{+\infty} f(E) \sum_n \delta(E - E_n) \langle \Psi_{1,n} | \tau_1 G_d^+ \Gamma_2 G_d \tau_1^+ | \Psi_{1,n} \rangle \, dE \]
\[ = \frac{2e}{h} \int_{-\infty}^{+\infty} f(E) \sum_{m,n} \delta(E - E_n) \langle \Psi_{1,n} | m \rangle \langle m | G_d^+ \Gamma_2 G_d \tau_1^+ | \Psi_{1,n} \rangle \, dE \]
\[ = \frac{2e}{h} \int_{-\infty}^{+\infty} f(E) \sum_m \langle m | G_d^+ \Gamma_2 G_d \tau_1^+ (\sum_n \delta(E - E_n) | \Psi_{1,n} \rangle \langle \Psi_{1,n} |) | m \rangle \, dE \]
\[ = \frac{2e}{h} \int_{-\infty}^{+\infty} f(E) \sum_m \langle m | G_d^+ \Gamma_2 G_d \tau_1^+ \frac{a_1}{2\pi} \tau_1 | m \rangle \, dE \]
\[ = \frac{2e}{h} \int_{-\infty}^{+\infty} f(E) \text{Tr}(G_d^+ \Gamma_2 G_d \Gamma_1) \, dE \]
(3.17)

To get the total current through the device, the current from contact 2 has to be subtracted away, therefore

\[ l = \frac{2e}{h} \int_{-\infty}^{+\infty} \text{Tr}(G_d^+ \Gamma_2 G_d \Gamma_1) [f_1(E) - f_2(E)] \, dE \]
(3.18)

The current equation expressed above is exactly the Landauer formula from Eq. (3.5), with \( \text{Tr}(G_d^+ \Gamma_2 G_d \Gamma_1) \) representing the trace for the transmission function.
Chapter 4: Results and Discussion

This chapter will cover the results obtained from simulations on organophosphonate molecules, more specifically a decyl-(C\textsubscript{10}) phosphonic acid, a 2,6-diphosphonoanthracene and comparing them to that of what has been reported in [110] and [111]. Simulations were split into two parts. The first part covers geometry optimization of the molecules in question, while the second part discusses electronic transport through these molecules. Both parts of the simulations were carried out with the aid of an atomistix toolkit called Quantumwise.

4.1 Building and optimizing the system

The system under investigation consists of a highly p-doped silicon substrate followed by a thin layer of silicon dioxide on which the molecules are then attached to. The molecules are then relaxed to figure out their favorable bonding conditions on the substrate/oxide surface at the lowest energy. The way the system was built up will be presented in this section. The silicon substrate was constructed from a single unit cell, and repeated in the x, y, and z directions as shown in Fig. 4.1. It is also important to note that the orientation of the silicon substrate is in the (100) direction.

![Fig. 4.1: Construction of Si substrate from a unit cell](image)
Chapter 4: Results and Discussion  

4.1 Building and optimizing the system

For the oxide, it was a bit more complicated. As there is no unique unit cell or supercell representation that corresponds to an amorphous oxide, such a sample was not present in the Quantumwise database as it was for silicon. Instead a crystalline oxide had to be constructed and some molecular dynamics simulations carried out. More specifically a cristobalite structure of SiO$_2$ was chosen since its unit cell can be transformed into an orthogonal cell which makes the simulation much more convenient, in contrast to quartz. After the crystalline oxide was constructed, the next step was to transform it into an amorphous structure. This was achieved by melting the crystal and cooling it down again. Simulation of the melting process was done at a very high temperature of 5000K in order to randomize the arrangements of the atoms. After the randomization was completed, the system had to be then cooled down to 300K on several steps by lowering the temperature gradually, in order to maintain as much as possible the local atomic arrangements (i.e. 4 oxygen atoms to 1 silicon). If this process would have been done in a single step (lowering from 5000K to 300K in one go), there will be a lot of uncoordinated atoms. Both the crystalline structure and the final amorphous oxide are illustrated in Fig. 4.2.

![Fig. 4.2: Construction of amorphous oxide](image)

In order to insure that the resulting oxide is correct, a radial distribution function was plotted between the silicon and oxygen atoms that describes how the density varies as a function of distance from a reference atom. As it is shown in Fig. 4.3, instead of sharp lines (which is the case for crystal structures), there is a distribution of broadened peaks, which is a characteristic for an amorphous structure.
Both molecules were constructed fairly straightforward. The alkane chain was constructed from Decane (C\textsubscript{10}H\textsubscript{22}) since they are similar in structure. Minor changes were applied to the Decane in which one hydrogen atom at both ends of the molecule was substituted with a hydroxyl group (OH) and a phosphorus acid group (PO\textsubscript{3}) respectively. The anthracene was constructed in a similar fashion, with a benzene ring used as starting point. It was then repeated two times and all 3 rings were attached together to form a chain. Again one hydrogen atom at both ends of the chain was substituted with a phosphorus acid group.

Both molecules were relaxed before being attached to the Si/SiO\textsubscript{2} substrate. Relaxation of the molecules in gas phase was done to see if there would be any change to the orientation of the molecules and furthermore to make the same process after attachment to the substrate a simpler one (regarding simulation time) by introducing an already relaxed structure. Fermi level and HOMO/LUMO energies were also calculated alongside the Mulliken charges for each atom. Fig. 4.4 shows the molecules before and after the relaxation process, Fig. 4.5 and 4.6 shows the
HOMO/LUMO representations for both molecules, and finally Table 4.1 summarizes the parameters chosen for the simulation and the results obtained.

Fig. 4.4: Alkane chain [left] and anthracene [right] before and after relaxation

Fig. 4.5: HOMO/LUMO for the alkane chain
## 4.1 Building and optimizing the system

### Fig. 4.6: HOMO/LUMO for the anthracene

- **decyl-(C$_{10}$) phosphonic acid (1)**
- **2,6-diphosphonoanthracene (2)**

<table>
<thead>
<tr>
<th>Exchange Correlation</th>
<th>GGA (PBE)</th>
<th>GGA (PBE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basis Sets</strong></td>
<td>Double-Zeta Polarized</td>
<td>Double-Zeta Polarized</td>
</tr>
<tr>
<td><strong>Fermi Level</strong></td>
<td>$-2.372$ eV</td>
<td>$-4.036$ eV</td>
</tr>
<tr>
<td><strong>HOMO</strong></td>
<td>$-5.582$ eV</td>
<td>$-5.155$ eV</td>
</tr>
<tr>
<td><strong>LUMO</strong></td>
<td>$0.837$ eV</td>
<td>$-2.916$ eV</td>
</tr>
</tbody>
</table>

**Mulliken Charges**

- Carbon: $3.845 - 3.860$  
  Phosphorus: $5.028$  
  Oxygen: $6.023 - 6.139$  
  Hydrogen: $0.932 - 1.073$

**Carbon:** $3.899 - 3.970$  
**Phosphorus:** $5.064 & 5.066$  
**Oxygen:** $6.004 - 6.136$  
**Hydrogen:** $0.927 - 1.073$

*Table 4.1: Simulation parameters and results*
Chapter 4: Results and Discussion

4.1 Building and optimizing the system

From the above table, GGA was preferred since LDA, as discussed in the previous chapter, is not a good approximation that can be applied to molecules. Double-Zeta polarized were the basis sets of choice to give more freedom to the orbitals as compared to the single-zeta. Major change in the orientation or shape of both molecules was not observed except for a slight difference in the anchor groups. For the alkane chain the band gap was calculated to be 6.419 eV and by looking at the HOMO/LUMO (Fig. 4.5), they are localized making the probability of finding an electron along the length of the chain low hence providing the proof that this is an insulating chain since it is made up of $\sigma$-bonds. As for the anthracene the band gap is 2.239 eV, and in contrast to the alkane chain the HOMO/LUMO are delocalized (Fig. 4.6) hence being a more conductive chain due to the alternating $\pi$-bonds that would contribute to charge flow.

4.1.1 Si/SiO$_2$ substrate and the decyl-(C$_{10}$) phosphonic acid

Now that the substrate, oxide, and the molecule are constructed, all that is left is to attach them together. Firstly the oxide was attached to the substrate. This is done automatically in Quantumwise by providing the program with the separate structures and it gives back the interfaces bound together. Substrate thickness is $\sim 2$ nm and the oxide thickness $\sim 1$ nm.

![Fig. 4.7: Relaxation of the alkane chain on Si/SiO$_2$](image)
The molecule (length ~ 1.6 nm) was brought close enough to bond to the oxide at an arbitrary position. All phosphorus-oxygen bonds were fixed at 1.63 Å prior to the relaxation process. As it is depicted in Fig. 4.7, after relaxation the molecule preferred a state where all oxygen atoms bonded to the oxide. Furthermore, in the PO₃ group there is a double bond (which is something that Quantumwise does not show, but can be found by identifying the shortest bond) between the phosphorus and one of the oxygen atoms, and this bond is no longer present since all 3 oxygen atoms have bonded to the oxide surface. There was no observable change to the shape of the molecule backbone even though it was left to relax in contrast to the substrate and oxide which were intentionally fixed and not subject to any relaxation. The HOMO/LUMO (Fig. 4.8) levels of the molecule shifted leading to a smaller bandgap along with an observable change in the Fermi level.

### 4.1.2 Si/SiO₂ substrate and the 2,6-diphosphonoanthracene

Same approach was taken in implementing the anthracene. Both the substrate and the oxide were of the same thickness while the molecule’s length was ~ 1.3 nm. Once the molecule was attached to the oxide surface and the P-O bonds fixed at 1.63 Å, the molecule was then relaxed.
Chapter 4: Results and Discussion

4.1 Building and optimizing the system

Before and after relaxations are shown in Fig. 4.9.

Fig. 4.9: Relaxation of the anthracene on Si/SiO₂

Fig. 4.10: HOMO/LUMO for anthracene after bonding
Examining the structure after the simulation, only two out of the three legs of the anchor group bonded to the oxide, with the P-O double bond corresponding to the one with a bond length of 1.47 Å. As for the shape of the molecule backbone, it did exhibit a slight bending but it is also important to stress that that did not lead to any breakage within the molecule. Shift in the HOMO/LUMO and Fermi levels (see Fig. 4.10) were also observed and a reduced bandgap was a result of that change.

To conclude, a summary of the geometry optimization (relaxation) process is given in the following table.

<table>
<thead>
<tr>
<th></th>
<th>decyl-(C\textsubscript{10}) phosphonic acid (1)</th>
<th>2,6-diphosphanoanthracene (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Before Relaxation</strong></td>
<td><strong>After Relaxation</strong></td>
<td><strong>Before Relaxation</strong></td>
</tr>
<tr>
<td>Substrate Thickness</td>
<td>2 \textit{nm}</td>
<td>2 \textit{nm}</td>
</tr>
<tr>
<td>Oxide Thickness</td>
<td>1 \textit{nm}</td>
<td>1 \textit{nm}</td>
</tr>
<tr>
<td>Molecule Length</td>
<td>1.6 \textit{nm}</td>
<td>1.3 \textit{nm}</td>
</tr>
<tr>
<td>Bonding Sites\textsuperscript{3}</td>
<td>2/3</td>
<td>3/3</td>
</tr>
<tr>
<td>Bond Lengths</td>
<td>1.63 Å</td>
<td>1.54 Å</td>
</tr>
<tr>
<td></td>
<td>1.63 Å</td>
<td>1.62 Å</td>
</tr>
<tr>
<td></td>
<td>1.63 Å</td>
<td>1.65 Å</td>
</tr>
</tbody>
</table>

\textit{Table 4.2: Summary of the geometry optimization (relaxation) process}

\textsuperscript{3} Number of bonds between anchor group and oxide
4.2 Transport Analysis

In the previous section both structures were constructed and relaxed in order to prepare them for transport simulations, but before going into these simulations there is still the matter of attaching the second contact, which in this case is gold. The gold contact was constructed in the same way the substrate was as explained in previous section. The starting point was a unit cell of gold that was repeated in the x, y, and z directions keeping in mind that the orientation of the gold contact was chosen to be in the (100) direction (See Fig. 4.11).

Fig. 4.11: Construction of a gold contact from a unit cell

Fig. 4.12: Final arrangement of both molecular configurations divided into specified parts
After the contact has been constructed, it is brought close enough to the respective molecule until they are bonded together. The next step is again done automatically by Quantumwise, at which it takes the bulk configuration and converts it into a device configuration as shown above in Fig. 4.12. Also shown are the different regions defined by the program for the respective device. Firstly, it detects the two periodic structures at both ends hence defining the left and right electrodes (i.e. dark red). The remaining part in the middle is defined as the central region (i.e. white), which in turn is sub-divided into the left and right electrode extensions (i.e. dark blue) and finally the scattering region (i.e. dark green). The reason for including the electrode extensions inside the central region, which undergoes a self-consistent procedure for computing the electron density and potential, is to provide a smooth transition from the molecule to the electrodes where the potential is bulk-like.

The final step that needs to be done before going into transport, is doping the silicon substrate in order to make it a highly p-doped substrate as stated in the beginning of this chapter. Doping in Quantumwise is not performed by adding dopant atoms to the substrate, instead it is done by setting an external potential which results from the adding of an extra charge. Basically by calculating the charge per atom an external potential is set that corresponds to the desired doping density. This is illustrated in Fig. 4.13 where a snapshot of the code describing the doping procedure was taken. First the doping density is defined (note the negative sign since p-doping is desired). The volume of the electrode is calculated and converted from Å$^3$ to cm$^3$. From the doping density and the volume the charge per atom is then calculated, which is then translated into an external potential that is finally set to the electrode and the silicon atoms present inside the central region (as part of the electrode extension).

![Fig. 4.13: Doping of the Si substrate](image-url)
Chapter 4: Results and Discussion  4.2 Transport Analysis

After the gold electrodes have been attached and the silicon substrates doped, all that is left is to carry out transport simulations on both devices and analyze the resulting I-V characteristics. Simulations were done with a positive voltage sweep applied to the substrate ranging from 0V to 0.9V with a step of 0.3V respectively. A self-consistent procedure at each voltage point was carried out at the start and then both the transmission spectra and the density of states (DOS) were plotted along with the I-V curves. All these results will be explained in detail in the following sub-sections.

4.2.1 Transport through Si/SiO$_2$/decyl-(C$_{10}$) phosphonic acid/Au

For the alkane chain, the I-V characteristics and the transmission spectrum are shown in Fig. 4.14. Fig. 4.15 and Fig. 4.16 show the density of states as a sum of all the states in the device compared to each part individually at each voltage point. These figures are ordered in the following fashion:

- Device Density Of States (Blue) vs. Silicon Density Of States (Black)
- Device Density Of States (Blue) vs. Oxide Density Of States (Red)
- Device Density Of States (Blue) vs. Molecule Density Of States (Green)
- Device Density Of States (Blue) vs. Gold Density Of States (Purple)

![Fig. 4.14: Alkane chain I-V characteristics (left) and transmission spectra (right)](image-url)
Chapter 4: Results and Discussion

4.2 Transport Analysis

Fig. 4.15: DOS at 0V and 0.3V
Chapter 4: Results and Discussion

4.2 Transport Analysis

Fig. 4.16: DOS at 0.6V and 0.9V
Chapter 4: Results and Discussion

4.2 Transport Analysis

From the above figures, the first observation would be the shape of the I-V characteristics. Electronic transport through the alkane chain exhibits a negative differential resistance (NDR) behavior, at which the current starts to rise with the applied voltage and then decrease with further increase in the voltage until it reaches a point where it rises again. A factor that may contribute to this NDR is the use of a semiconducting material as one of the contacts. If metal contacts were employed at both ends of the device, this NDR behavior would not be observed since the density of states in metals are continuous. That is not the case with semiconductors since for a large enough voltage the levels of the molecule can drop at the same energy with the band gap stopping transmission of electrons. Furthermore, this behavior can be understood by examining Fig. 4.15 and Fig. 4.16. As the voltage is applied the current rises while the difference between the electrode chemical potentials also increase. Between 0.3V and 0.6V the current decreases since all states that are present within this window lie inside the bandgap of silicon, hence there cannot be charge transport within this region of applied bias. It is important to note that the bandgap of silicon was calculated to be approximately 0.6 eV. This is mainly due to the nature of the DFT calculations as it usually underestimates bandgaps. From a bias of 0.6V upwards, the current starts to increase again since the voltage applied causes a shift in the levels of the silicon and therefore introducing a number of states outside of the bandgap that would contribute to charge transport.

Another observation would be the lack of peaks in the transmission spectra (Fig. 4.14) within the region between the left and right electrode chemical potentials (i.e. $-4.38 \text{ eV} \rightarrow -5.28 \text{ eV}$). Even by observing the density of states of the molecule, it is almost flat. Then from where is the current generated? There are in fact states provided by the molecule for transport, but the density of these states is extremely small giving the flat representation of the transmission spectra. To locate these states they must be plotted on a logarithmic scale, as it is show in Fig. 4.17.

Fig. 4.17: Log scale of the molecule DOS
Nevertheless the molecule still mainly acts as barrier and the carriers supplied by the contact reach the oxide/molecule interface at which they then probably tunnel through towards the opposite contact.

Finally, the highest current obtained was approximately $2.4 \times 10^{-15}$ A. Converting this into current density with respect to the size of the structure at hand (i.e. $7.6 \AA \times 7.6 \AA = 57.7 \AA^2$), would result in a current density equal to $0.415 \text{ A/cm}^2$. Comparing this value to what has been reported in [110] (i.e. $10^{-3} \text{ A/cm}^2$), it shows a 3 orders of magnitude higher current.

### 4.2.1 Transport through Si/SiO$_2$/2,6-diphosphonoanthracene/Au

The results for the anthracene are presented here in the same fashion as it was done in the previous sub-section. Fig. 4.18 through to Fig. 4.20 show the I-V characteristics, transmission spectra, and the density of states (in the same order as before).
Chapter 4: Results and Discussion

4.2 Transport Analysis

Fig. 4.19: DOS at 0V and 0.3V
Chapter 4: Results and Discussion

4.2 Transport Analysis

Fig. 4.20: DOS at 0.6V and 0.9V
Generally speaking, one would expect a much improved transport through the anthracene as compared to the alkane chain, since this molecule consists of alternating π-bonds hence making it a more conductive structure. This is true with much higher current values being observed. Before going into more details regarding the current values, by first examining the I-V characteristics leads to the conclusion that this molecule behaves like a diode, at which the current is either zero or negligible until the applied bias reaches a certain point where the current suddenly increases. This behavior is supported with what is presented by the density of states from Fig. 4.19 and Fig. 4.20. More specifically by observing the silicon density of states from 0V to 0.6V, it can be clearly seen that all the states that would be available for transport again lie in the same region as that of the silicon bandgap, therefore no current is generated. On the other hand between 0.6V and 0.9V, the window provided by the difference in electrode Fermi energies starts to include states that do not coincide with the silicon bandgap, hence this gives a rise in the current generated.

Further analysis of Fig. 4.19 and Fig. 4.20 lead to the conclusion that both the oxide and the molecule play an important role by providing the required states for transmission. The states of the molecule can be observed on a logarithmic scale to support this conclusion (see Fig. 4.21). Clearly, the anthracene provides more available states compared to the alkane chain with 2 orders of magnitude higher in the density of states.
Finally, from the I-V characteristics in Fig. 4.18 the highest calculated current is approximately 1.1 nA at a bias of 0.9V. It is important to note that the negative current in Fig. 4.18 is merely due to the miss-interpolation of the curve by Quantumwise according to the identified points (i.e. the current is actually zero in this segment). Comparing that to the results for the alkane chain, there is a rather large difference up to 6 orders of magnitude in the current. This is expected due to the fact that the anthracene is a more conductive molecule.
Chapter 5: Conclusion

Two molecules were under investigation in the worked carried out in this thesis. One is an alkane chain (decyl-(C\textsubscript{10}) phosphonic acid) and the other is a anthracene (2,6-diphosphonoanthracene). Both molecules were relaxed on top of silicon/silicon dioxide substrate and then put through transport simulations.

For the geometry optimization, both molecules favored bonding at the oxide surface with at least 2 out of the 3 legs of the anchor group. Nevertheless, it must be noted that the bonding might be different if the molecules were placed at different sites along the oxide surface. Since the oxide surface is not uniform, it is not possible to accurately predict how the molecules will bond to the surface. There was not a significant observed change in the backbone of the molecules after the relaxation process. HOMO and LUMO calculations showed that the molecules had a reduced bandgap as a result of their relaxation onto the oxide surface. Furthermore, they provided the sufficient proof that one would act an insulating chain while the other as a more conductive chain.

After the molecules have been relaxed they were put through a transport simulation to observe their I-V characteristics. For the alkane chain, it showed characteristics of a negative differential resistance, and according to the density of states most of what is contributing to the current in the device employing this molecule is coming from the substrate and oxide regions and appear to tunnel through the molecule (which acts as a barrier) into the opposite contact. Highest current was calculated to be \(2.4 \times 10^{-6} \text{ nA}\) and a current density of \(0.415 \text{ A/cm}^2\). The difference between the calculated current density and the previously reported one could be accounted for according to conditions under which the results were obtained. Such conditions include the under estimation of the silicon bandgap (which is common with DFT) and the spreading of the molecular orbitals, hence over estimating the conductivity of the molecular junction. Another would be the size of the contacts, oxide and molecules, since they might not be exactly identical and this difference (even the slightest) will lead to a large variation in the results obtained between the experimental procedure and theoretical simulations. As for the anthracene a much higher current was observed with values reaching \(1.1 \text{ nA}\). In this device both the states inside the molecule and the oxide play an important role in the conduction, but only after a certain value of applied bias to the substrate. This is due to the fact the states lie within the region of the silicon bandgap, and as the voltage applied increases the window between the left and right electrode Fermi energies will start to include more and more available states for conduction. It is also shown that even with further increase in the
applied voltage the current will increase more as there is a large number of states provided by the molecule on either side of its bandgap.

Even though the results obtained in the current work paints a small part of the bigger picture, that is the investigation of SAMFETs, these results provide a good qualitative assessment which is in agreement with previous work and can be further expanded to get a better understanding of such devices, especially concerning charge injection in the SAM layers as well as leakage current effects.
Bibliography


Bibliography


Acknowledgments

I would like to express my sincere gratitude towards Prof. Alessio Gagliardi. First and foremost, for giving me the opportunity to finish my master’s thesis. I would also like to thank him for his guidance, discussions, and for giving me the freedom to express myself. During this period I have learned a lot from him and with his support I was able to finish this thesis.

I would also like to thank my colleague and friend Tim Albes. From the first day he has been very friendly and made it easier for me to feel welcomed and settle down. Furthermore whenever I had technical problems he was always there to assist me in every way possible.

Finally, a massive appreciation goes out to my family. Since the start of my master’s program they have always stood by me and gave me immense support which will always be remembered.