

A Deformation Induced Quantum Dot

Investigating the Electronic Properties of a Carbon
Nanotube Cross

by

Daniel James Woodsworth

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

Bachelor of Science (Honours)

in

The Faculty of Science

(Physics)

The University Of British Columbia

(Vancouver)

April, 2008

© Daniel James Woodsworth 2008

Abstract

Due to their extraordinary electronic properties, Quantum Dots (QDs) are potentially very useful nanoscale devices and research tools. As their electrons are confined in all three dimensions, the energy spectra of QDs is discrete, similar to atoms and molecules. Because the gaps between these energy levels is inversely related to the size of the QD, very small QDs are desirable.

Carbon nanotubes have long been touted as fundamental units of nanotechnology, due to their structural, optical and electronic properties, many of which are a result of the confinement of electrons in the trans-axial plane of the nanotube. It is known that their band gap structure is altered under deformation of their cross section.

It is proposed that one way to fabricate a very small quantum dot is by confining electrons in the nanotube so that they may not freely move along its length. A structure to produce this confinement has been described elsewhere, namely the carbon nanotube cross, consisting of two carbon nanotubes, with the one draped over the other at ninety degrees. It is thought that this structure will induce local physical deformations in the nanotube, resulting in local changes in electronic structure of the top nanotube at the junction of the cross. These band gap shifts may cause metal-semiconductor transitions, resulting in tunnel barriers that axially confine electrons in the nanotube. This thesis investigates the possibility that the carbon nanotube cross may exhibit QD behavior at the junction of the cross, due to these local band gap shifts.

A device for carbon nanotube growth, using Chemical Vapor Deposition, has been designed, and may be built using microfabrication techniques. This device consists of electrodes (for electrical measurements of the nanotubes) and catalyst regions (to initiate nanotube growth), lithographically patterned in a configuration that promotes carbon nanotube formation. Unfortunately, due to fabrication issues, this effort is a work in progress, and these devices have not yet been constructed. However, an experimental methodology has been developed, which provides a framework for eventually building a carbon nanotube cross, and investigating the possibility of QD behavior at the junction of the cross.

This structure has also been investigated computationally. Molecular dynamics simulations were used to obtain equilibrium geometries of the carbon nanotube cross, and it was found that there are many different meta stable states, corresponding to different types of nanotube, and different physical arrangements of these nanotubes. The electronic structure of the carbon nanotube cross was calculated using the density functional theory. Band gap energies similar to experimental values were obtained. A one-to-one spatial correlation between deformation and band gap and conduction band shifts were observed in the top carbon nanotube of the nanotube cross. Small tunnel barriers, inferred from both the calculated band gap and LUMO energies, are observed, and could well be sufficient to confine electrons along the axis of the nanotube.

The results described in this thesis, while not definitive, certainly indicate that a QD probably would form at the junction of a carbon nanotube cross, and that further investigation, both experimental and computational, is warranted.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Figures	v
Acknowledgements	vi
1 Introduction	1
1.1 Quantum Dots	1
1.2 Carbon Nanotubes	2
1.3 The Carbon Nanotube Cross	2
1.4 Investigation of the Carbon Nanotube Cross	5
2 Theory	6
2.1 Energy States of Quantum Dots	6
2.2 Band Structure	6
2.3 Physical Structure of Carbon Nanotubes	8
2.4 Band Structure of Carbon Nanotubes	9
2.5 Effects of Deformation on the Band Structure of Carbon Nanotubes	10
3 Simulation Methods	12
3.1 Molecular Dynamics	12
3.2 Quantum Chemistry Theories	13
3.2.1 Hartree-Fock Method	13
3.2.2 The Density Functional Theory	14
4 Fabricating the Carbon Nanotube Cross	16
4.1 Design and Fabrication of the Chip	16
4.1.1 Design	16
4.1.2 Fabrication	17
4.2 Experimental Design	19
4.2.1 Microfabrication	19
4.3 Carbon Nanotube Growth	21
4.4 Measuring the Current-Voltage Characteristics of the CNT Cross	22

Table of Contents

5	Modeling the Carbon Nanotube Cross	24
5.1	Geometry	24
5.2	Electronic Structure	29
6	Conclusions and Future Work	36
	Bibliography	38

List of Figures

1.1	The Carbon Nanotube Cross.	3
2.1	Physical and Electronic Structure of Graphene and CNTs.	9
4.1	Device and Chip Design Schematics.	17
4.2	Microfabrication Process.	20
4.3	Current-Voltage (I-V) Measurements of Quantum Dots.	21
5.1	Comparison of Gaussian Profiles of CNT Cross	26
5.2	MD Relaxation of Carbon Nanotube Crosses	28
5.3	Investigating the Inter-Molecular Interactions in NanoHive	29
5.4	Comparison of GAUSSIAN methods.	30
5.5	Electronic Structure of CNT Crosses: Band Gap Energy.	32
5.6	Electronic Structure of CNT Crosses: LUMO Energy.	34

Acknowledgements

I would like to thank my supervisor, Dr. Alireza Nojeh, for his extensive and continued support and patience throughout this project. Dr. Nojeh has been extremely generous with his time and knowledge, for which I am very grateful. Under his guidance I have gained a far deeper understanding of this field than I otherwise would have. He has my sincere appreciation for his involvement in all aspects of this work.

Chapter 1

Introduction

As the semiconductor industry continually strives to construct ever smaller structures for microelectronics applications, it has provided the scientific community with a wealth of nanofabrication techniques, as well as the ability to shape matter on the atomic scale. This capacity has been one of the main factors in the explosion of research and development at the nanoscale, which has already had a deep impact on both fundamental science, as well commercial technology. One of the objects which is perhaps most representative of this union of technological promise and scientific insight is the so called artificial atom, or Quantum Dot.

A novel structure has been proposed[28], consisting of a cross of Carbon Nanotubes, with one nanotube draped over the other at ninety degree angles, as in Fig. 1.1(a). This thesis describes the investigation of this structure, specifically examining its potential to exhibit Quantum Dot behavior.

1.1 Quantum Dots

Starting around twenty years ago semiconductor fabrication technology had progressed sufficiently such that it was possible to construct objects which could confine countable numbers of electrons to three dimensional regions small enough that their discrete energy spectrum could be resolved[2, 17]. These so-called Quantum Dots (QDs) are often described as zero-dimensional structures, because, as opposed to a bulk three dimensional solid, the electrons are confined in all three dimensions[8]. Because of this confinement, instead of the virtually continuous energy levels of electrons in a bulk solid, the electrons in a QD, having no translational elements in their Hamiltonian, are confined to a finite set of quantized states[8]. In this way these QDs are very similar to natural atoms, except that the energy spectra of QDs, which defines most of their important physical properties, are a function of size, composition and morphology[5]. As these properties may be controllably adjusted, it follows that it is possible to engineer nanoscale, ‘artificial atoms’, which would have a myriad of possible applications.

To the researcher, QDs present a controlled environment in which to study all manner of quantum and atomic phenomena. Examples of the former include time reversal symmetry breaking under applied magnetic field[1], spin and coherence effects[18] and electron transport and the Kondo effect[32]. The artificial atom analogy is indeed apt, as QDs provide the opportunity to systematically study atomic phenomena and electron-electron interactions[22], as well as mimic experiments in atomic physics with macroscopically tunable QDs[23]. Furthermore, it has been argued that it may even be possible to see novel physical effects in QDs that are not found in nature[2, 16].

The fields in which QDs have probably made the furthest progress towards widespread technological application are biological analysis[15] and medicine[31]. In both cases, QDs are used

as a means of *in vivo* fluoroscopic imaging: examining cellular mechanisms in the former and diagnostic imaging in the latter. Here the tunable energy spectrum of QDs is exploited to create particles that will absorb and emit arbitrary wavelengths of light, enabling extremely accurate imaging.

QDs provide one of the most promising routes to a solid state quantum computing platform, in which multiple entangled QDs are combined to form a single qubit[9]. A wide range of optoelectronic applications of QDs are also envisioned, including lasers and optical detectors[6], as well as nonlinear optics, telecommunications and photovoltaics[8]. Finally, there are ongoing attempts to assemble groups of QDs, ‘artificial molecules’, and arrays of QDs, ‘artificial solids’, raising the possibility of fully engineered materials and macro-molecules[2, 5].

1.2 Carbon Nanotubes

Although they had been observed, and subsequently misclassified, as early as the 1970s[35], Carbon Nanotubes (CNTs) were first reported by Iijima in 1991[13], and can be divided into two main types: Multi Walled Carbon Nanotubes and Single Walled Carbon Nanotubes. A Single Walled Carbon Nanotube, an example of which is shown in Fig. 1.1(d), may be thought of as a hollow cylinder made of a hexagonal array of carbon atoms, one atomic layer thick. A Multi Walled Carbon Nanotube is simply many Single Walled Nanotubes, arranged in concentric shells. I will be dealing exclusively with Single Walled Carbon Nanotubes in this thesis, and so will simply refer to them as Carbon Nanotubes, or nanotubes for brevity.

Carbon Nanotubes have garnered a great deal of attention and excitement, both from the public, as well as industry and academia. This is in large part due to their extraordinary structural, electronic, thermal and optical properties[3]. Typical diameters for CNTs are on the order of nanometers, while their length is on the order of hundreds, even thousands, of microns, giving them very large aspect ratios, which is in many cases responsible for their unique properties[3]. Particularly important for this thesis is that CNTs can be either semi-conducting or metallic[25].

Due to these characteristics, a wide array of applications incorporating CNTs has been envisioned, including: “electronic devices and interconnects, field emission devices, electrochemical devices, such as supercapacitors and batteries, nanoscale sensors, electromechanical actuators, separation membranes, filled polymer composites, and drug-delivery systems”[10].

1.3 The Carbon Nanotube Cross

Since they were first studied, all manner of QDs have been reported, possessing a wide variety of compositions, structures and properties, to the point that there is currently some ambiguity in the literature as to what exactly a QD is[27]. One unifying theme is the crucial notion of three dimensional confinement of electrons, and how this is achieved is always the central challenge when trying to design or manufacture a QD. Original means of confinement included lithographically patterning electrodes on a silicon wafer, Fig. 1.1(c), and confining a two dimensional electron gas to a small point in the plane using electric fields. More recently there has been much work done on self assembled QDs, which are usually small, semi-hemispherical objects made of some semiconducting material (*GaAs*, *InGaAs* for example)[8]. Here confinement is generally due to strain induced band gap changes in the semiconductor.

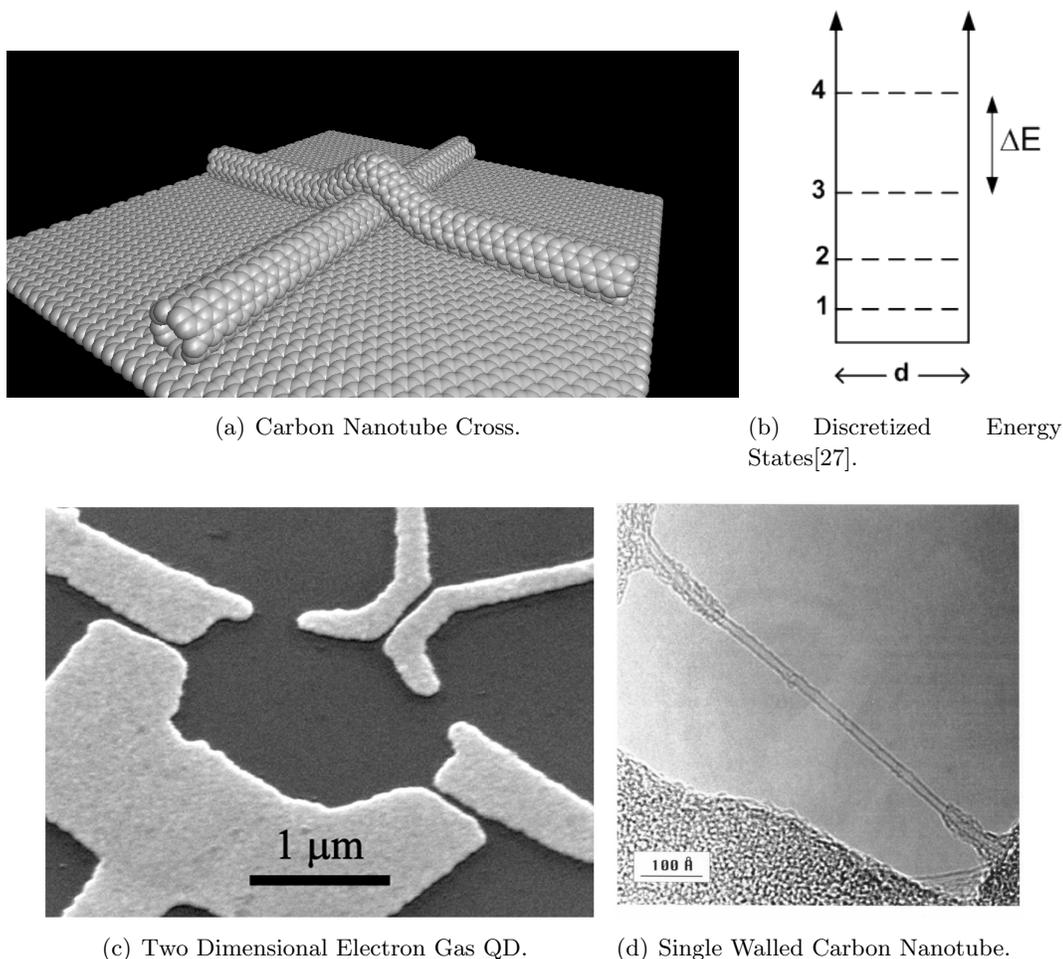


Figure 1.1: Motivation for the Carbon Nanotube Cross. The method proposed in this work aims to exploit the inherently small nature of carbon nanotubes (d), by investigating a cross structure of two CNTs (a). As a QD at the junction of a CNT cross would potentially have a much smaller characteristic size than traditional QDs (c), it would have much larger energy level spacing (b). Images (c) and (d) are taken from www.unibas.ch and www.almaden.ibm.com respectively.

Many of these methods rely heavily on microfabrication techniques, and despite their undeniable success, as science and technology strives to work at ever smaller scales, and researchers try to achieve confinement in ever smaller volumes, they try to thread a needle through a hole of ever decreasing size. Inevitably this effort is subject to a law of diminishing returns: researchers must spend more time and resources in an effort to achieve ever smaller feature sizes on silicon wafers. While they continue to succeed, it is likely that eventually this progress will slow to a painful, perhaps prohibitive rate.

Alternatively, rather than forcing macroscale technology into the nanoscale, it may be possible to exploit the unique features of this regime, and adapt existing mechanisms, structures and

objects to the purposes of research and technology at the nanoscale. This is the philosophy adopted throughout this thesis, and is manifested in the basic structure of this project: the carbon nanotube cross (Fig. 1.1(a)).

Consider attempting to confine electrons starting with a macroscopic piece of a three dimensional material, perhaps a chunk of silicon. If one could manage to machine a cube with length of perhaps, a nanometer, then confinement would have been achieved. Unfortunately this would be a very difficult prospect, and would involve very refined versions of those techniques described above. A somewhat easier route, is if one was given a material in which electrons were already confined in one direction, and all that remained was to achieve confinement in the remaining two dimensions. This is indeed how many QDs are fabricated, but again still relies on microfabrication techniques, and results in relatively bulky QDs, as seen in Fig. 1.1(c). Finally, imagine being given an object whose electrons were confined to a single axis. All that would remain would be to somehow stop a certain number of these electrons from moving along this one axis.

Perhaps somewhat surprisingly, these last objects do indeed exist. Both carbon nanotubes and silicon nanowires, at least to a very good approximation, are one dimensional objects, with their electrons effectively confined to a line. However, the problem remains to confine these electrons along this line. There is of course the possibility of manufacturing some kind of mold into which one could press a nanotube, and so bend it or shear it, which, as will be described, would probably confine the electrons. Alternatively one could attempt to directly interact with it, via Scanning Tunneling Microscopy (STM) or Atomic Force Microscopy (AFM) probes. However all these methods have the same limitations and looming dead end as described above.

One potential way to achieve confinement is to grow two nanotubes in a crossed formation, the one draped across the other, as is shown in Fig. 1.1(a). The resulting deformation causes the necessary change in electronic structure to confine the electrons from moving freely along the axis of the nanotube. In this way a QD can be constructed in a manner that harnesses the features of the nanoscale world, rather than forces an outside pattern onto it. Carbon nanotubes may simply be grown, and provided appropriate stimulus, it should be possible to grow them in this crossed formation[28].

At this point, it is reasonable to wonder why it is necessary to avoid using conventional microfabrication techniques. Certainly some esthetic desire to exploit the uniqueness of the nano world is not enough. However, there is one fundamental reason for attempting to do this, and that is size. As will be shown later, the energy gap (ΔE in Fig. 1.1(b)) between two energy states in a QD's spectrum is inversely related to its characteristic size. Thus, given a relatively large QD, with a size in the hundreds of nanometers, the energy states will have a very small, or fine separation. In many cases this separation will be much less than the thermal energy of electrons¹, and so these QDs must be operated at very cold temperatures, or else the discrete nature of their energy spectra is completely washed out. On the other hand, small QDs, with sizes on the order of a nanometer, should have energy spectra that would be measurable at room temperatures.

Aside from the pure convenience of room temperature operation, for many of the eventual applications envisioned for QDs, this would be absolutely essential. This small characteristic size requirement is one of the fundamental reasons for studying the carbon nanotube cross structure. If one were to try and construct a QD on the order of 1 nanometer by confining a two dimensional

¹At room temperature $k_B T \approx 26 \text{ meV}$.

electron gas using electrodes, one would have to pattern these electrodes to around one nanometer, which is very difficult: notice the size of the electrodes in Fig. 1.1(c). However, the diameter of (single walled) nanotubes are usually $\sim 0.8 - 1.5 \text{ nm}$. A QD formed at the junction of a carbon nanotube cross would *naturally* have a characteristic size of around one nanometer, and so it presents a unique way of obtaining an extremely small QD, which could possibly be operable at room temperatures.

This structure displays several other desirable features, for the most part in terms of potential application. First, there is the synergistic combination of QDs and CNTs: two objects that are touted as some of the fundamental building blocks of nano-electronics and nano-electromechanical systems (NEMS). For example it has already been proposed that CNTs would make excellent field emitters, while QDs are potentially very good single electron transistors. Combined in a single package, and therefore avoiding the enormous difficulty of interfacing at the nanoscale, one might have a means of a nanoscale field emitter, whose emission is regulated to an extraordinarily fine degree by the QD.

As most QDs are fabricated from semiconductor materials, it is often necessary to coat them with organic materials for biological compatibility, and even with these additional measures, toxicity is a concern[14]. Furthermore, for cellular imaging applications, the necessary size of the biologically suitable QDs that have been produced is such that the QDs cannot be introduced to the cells via the natural transmembrane diffusion processes. Rather, QDs must be introduced using far more invasive approaches such as microinjection, which has obvious detrimental consequences for *in vivo* imaging[14]. Due to its carbon composition and extremely small size, it is possible that QDs in a nanotube cross might remedy these problems.

1.4 Investigation of the Carbon Nanotube Cross

Recognizing its potential, this thesis describes the ongoing study of the carbon nanotube cross. The ultimate goal of this work is to achieve three dimensional electron confinement in a carbon nanotube, specifically at the junction of the cross, and investigate the properties of this QD.

Following previous experiments by my supervisor, Dr. Alireza Nojeh, [28], two main avenues of study were undertaken. First, a methodology and design for growing carbon nanotubes in a crossed formation was pursued, consisting of growing CNT crosses, using Chemical Vapor Deposition (CVD), on lithographically patterned chips and then performing Current-Voltage measurements on the these crosses. Second, in order to gain more insight into the physical processes involved in the CNT cross, it was modeled by generating relaxed structures using Molecular Dynamics (MD), whose electronic structure were then calculated using Computational Quantum Chemistry (QC) algorithms.

Through these two complimentary approaches, it is hoped that this work will lay the foundations for a better understanding of a structure so rich in scientific interest and technological utility: the Quantum Dot in a Carbon Nanotube Cross.

Chapter 2

Theory

2.1 Energy States of Quantum Dots

Much of the interest in QDs is due to their controllable, discretized, electronic structure, and there has been an enormous amount of theoretical and experimental study of the energy states in QDs. Although the details of these works are quite complicated, I would like to at least heuristically show the dependence of a QD's energy level spacing on its size, in order to give some physical grounding to the claim put forward in the introduction, namely that a very small QD should have easily resolvable discrete energy levels at higher temperatures.

The Schrödinger equation for the wave function, $|\Psi(\mathbf{r})\rangle$, of a particle in a time independent potential, $V(\mathbf{r})$, is:

$$H|\Psi(\mathbf{r})\rangle = E|\Psi(\mathbf{r})\rangle \quad (2.1)$$

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}) \quad (2.2)$$

For our purposes here, the important feature of a QD, namely the three dimensional confinement of the electrons, can be modeled using the potential, V . Two suitable potentials are: i) a three dimensional box of dimensions L_x, L_y, L_z , where $V = 0$ inside the box, and is infinite elsewhere and ii) a sphere of radius R_o where $V = 0$ inside the sphere and is infinite elsewhere. Substituting these potentials into the Schrödinger equation and performing the straightforward separation of variables, and invoking appropriate boundary conditions gives:

$$E_{n_x, n_y, n_z} = \frac{\pi^2 \hbar^2}{2m} \frac{n_x^2 + n_y^2 + n_z^2}{L_x^2 + L_y^2 + L_z^2}, \quad n_i = 1, 2, 3, \dots \quad (2.3)$$

$$E_{nl} = \frac{\hbar^2}{2m} \frac{k_{nl}^2}{R_o^2} \quad (2.4)$$

where E_{n_x, n_y, n_z} are the energy states of the box potential and E_{nl} are those of the spherical potential, $n_i = n_x, n_y, n_z$ and k_{nl} is the n^{th} zero of the Bessel function of order l . The important result, confirmed by experiment, is that the energy states of a QD are inversely proportional to (in this case) the square of its length. Therefore to achieve a large separation of energy states, and create a viable room temperature QD, its characteristic size should be very small.

2.2 Band Structure

As most of the theory described here lies within the realm of condensed matter physics, and relies heavily on the concepts of band structure and band gap energy (E_g), I will briefly review

these concepts here. However, this is a very large field, and for more complete treatments see, for example, Kittel[19], whom I have followed closely here.

As a first approximation, we may model a solid crystal as a collection of electrons confined to a cube of side length L . The valence electrons of each atom are completely delocalized, and are free to move throughout the solid. As we are considering a bulk crystal, we assume periodic boundary conditions, so that the solid ‘appears’ the same from all points throughout the solid (neglecting edges). We further assume that the individual electrons are non-interacting. Recall that the state of an electron is given by the electron’s wave function $\Psi(\mathbf{r})$, where $\Psi(\mathbf{r})$ solves the Schrödinger equation (Eq. 2.1). As the electrons are free to move, the potential, $V(\mathbf{r})$, is defined to be zero inside the cube. Through separation of variables we arrive at solutions that are travelling plane waves: $\Psi = e^{i\mathbf{k}\cdot\mathbf{r}}$ with $E = \frac{\hbar^2\mathbf{k}^2}{2m}$. Here \mathbf{k} is the wave vector and has components $k_i = \frac{2n\pi}{L}$ for $i = x, y, z$ and n is any integer.

While this so-called free electron model provides a good understanding of many properties of metals, such as thermal and electrical conductivity, it fails to take into account the periodicity of the crystal lattice, and the associated potentials of each atom, and so fails to give a good description of most optical and electronic properties[19]. To account for this periodicity, we assume that electrons in the above model are only weakly perturbed by the periodic potential of the atoms. Thus the travelling wave solutions now ‘see’ periodic crystal planes, as opposed to the uniform zero potential of the free electron model. The travelling waves undergo Bragg reflection off these planes, and this process is described by Bragg’s law, $2d\sin\theta = n\lambda$, where d is the lattice spacing of the crystal and $\lambda = \frac{2\pi}{k}$ is the wavelength of the travelling wave. It follows that waves that have $k = \frac{n\pi}{d}$, regardless of the direction they are travelling in, are completely reflected, according to the Bragg Law, resulting in a standing wave for this energy eigen state. The two possibilities for the form of this standing wave are[19]:

$$\psi(+)=e^{i\pi x/d}+e^{-i\pi x/d}=2\cos(\pi x/d) \tag{2.5}$$

$$\psi(-)=e^{i\pi x/d}-e^{-i\pi x/d}=2i\cos(\pi x/d) \tag{2.6}$$

These standing waves produce localized regions of elevated electron density (as can be seen by taking the modulus squared of either wave function), and the important thing to note is that each wave produces a different distribution. In particular $|\psi(+)|^2$ has high charge density at the location of the ion cores, and due to the attraction between the negative electrons and positive ion cores, electrons with this wave function have lower potential energy. Conversely, $|\psi(-)|^2$ has high charge density away from the ion cores: a state of higher potential energy. This difference in potential energies is essentially the origin of the band gap in crystal solids, because, as we have just seen, electrons with wave vectors that satisfy $k = \frac{n\pi}{d}$ are confined to have wave functions of either $\psi(+)$ or $\psi(-)$. If we take the expectation values of the energies for $\psi(+)$ or $\psi(-)$, we find that they are, in general, separated by some energy E_g , which is to be expected, as the two wave functions correspond to electron distributions with different potential energies. Therefore, there is a so-called band gap in the set of allowed energy states for electrons. If we continually add energy to an electron, it will be excited to higher and higher states, until it meets the Bragg condition, and then it must have sufficient energy, namely E_g , in order to reach the next band of allowed states.

The set of states with lower potential energy, namely those of $\psi(+)$, is called the valence band, while the set with higher potential energy, $\psi(-)$, is called the conduction band, and the energy difference between the highest energy state of the valence band and the lowest energy state of the conduction band is the band gap energy. In reality, the band structure of any material is extremely complicated, with many different allowed bands and gaps, and is usually measured using some form of diffraction technique (neutron, electron or x-ray). However, the important point is that the relative positioning of the valence and conduction bands, and the size of E_g generally defines the nature of a crystalline solid. Metals have zero band gap energy and overlapping valence and conduction bands, so there are many electrons in the conduction band available to move about freely, as is characteristic of metals, which are excellent conductors. Insulators are just the opposite, with a very large band gap, meaning that a large amount of energy must be supplied to excite an electron into the conduction band, which results in relatively few free charge carriers, and low conductivity. Semi-conductors are in between these two cases, with a moderate band gap. Thus by supplying reasonable amounts of energy, electrons can be promoted or prevented from reaching the conduction band, giving semiconductors their extremely useful controllable or diode-like characteristics.

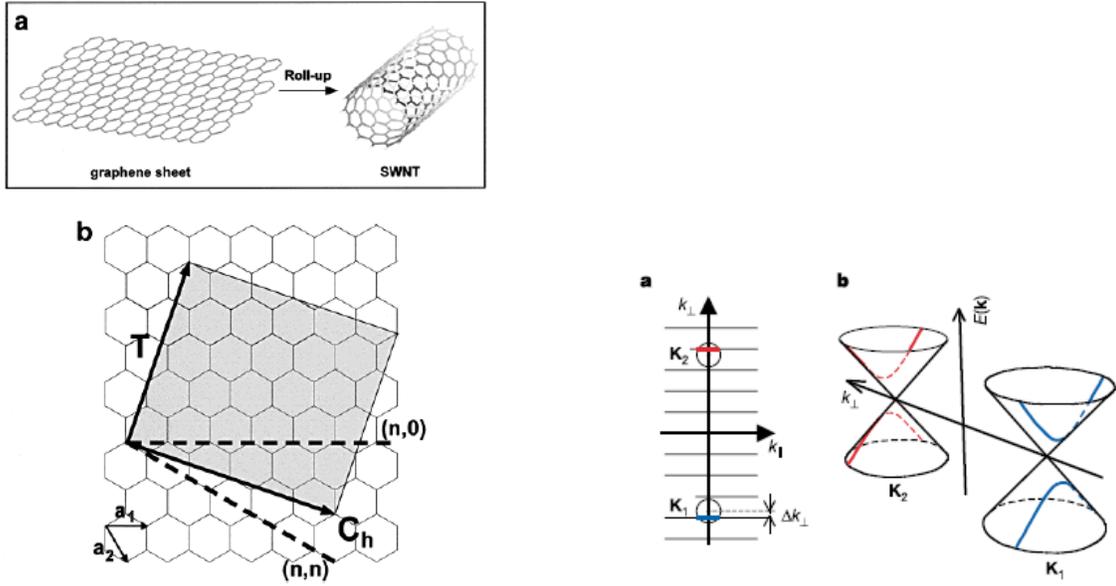
This work essentially exploits the fact that deformations and strains in the *physical* structure of carbon nanotubes cause spatially correlated local changes in the *electronic* band structure causing local transitions from semi-conducting to metallic nanotubes, or vice versa. I will first give a brief overview of the relevant physical and electronic properties of CNTs, before describing how the carbon nanotube cross produces these shifts in band gap energy, thereby confining the electrons.

2.3 Physical Structure of Carbon Nanotubes

Carbon has several allotropes, one of which is graphite, a common structure found in lead pencils and lubricants. Graphite consists of planar sheets of carbon atoms stacked upon each other, where each sheet is only one atomic layer thick and is known as graphene. These graphene sheets consist exclusively of carbon atoms, sp^2 bonded in a hexagonal network. A CNT is essentially a sheet of graphene rolled into a tube, as depicted in Fig. 2.1(a). As such many of the properties of CNTs can be understood in the context of the lattice structure of graphene, which can be described² by making a few definitions with reference to Fig. 2.1(a). The chirality vector, $\mathbf{C} = n\mathbf{a}_1 + m\mathbf{a}_2$, connects two crystallographically equivalent sites in the two dimensional graphene sheet. Here n and m are integers and \mathbf{a}_1 and \mathbf{a}_2 are the graphene lattice vectors, which define a unit cell in the hexagonal graphene structure. In other words, for any location \mathbf{r} in the graphene sheet, any other location $\mathbf{r}' = \mathbf{r} + \mathbf{C}$, will be completely equivalent for all m, n . This set of points defines the graphene lattice[19]. The translation vector, \mathbf{T} , is defined to be orthogonal to \mathbf{C} , and is extended until the first crystallographically equivalent site is encountered.

The area spanned by \mathbf{C} and \mathbf{T} (shaded gray in Fig. 2.1(a)) defines the unit cell of the CNT when it is constructed by rolling up the graphene sheet. When the sheet of graphene is rolled up the lattice points defined by the two ends of the chirality vector coincide, resulting in the continuous hexagonal structure of CNTs, seen in Fig 2.1(a). Note that \mathbf{T} points along the axis

²The discussion here closely follows that of Leiber[29].



(a) Hexagonal structure of graphene. Taken from [29].

(b) Band structure of graphene and CNTs. Taken from [26].

Figure 2.1: Physical and Electronic Structure of Graphene and CNTs. (a) The grey region spanned by \mathbf{C} and \mathbf{T} defines the unit cell of a carbon nanotube (CNT). The free parameters in this unit cell are (n, m) , which define the chirality and diameter of the CNT. (b) The valence and conduction bands meet at \mathbf{K}_1 and \mathbf{K}_2 . The quantization of k_c gives a discrete set of allowed states (horizontal lines at left and parabolas at right). The 1-D band structure of CNTs are slices of the 2-D band structure of graphene at points defined by the quantization of k_c .

of the CNT and \mathbf{C} is its circumference.

As a CNT is simply the unit cell repeated along \mathbf{T} , and the unit cell is purely a function of \mathbf{C} , it follows that a CNT is completely defined by the integer pair (n, m) . On the other hand, upon considering the honeycombed structure in Fig. 2.1(a), it is clear that there are only two physical degrees of freedom of the rolled up CNT: the diameter of the CNT (which is just $\frac{C}{\pi}$) and the chirality of the CNT. To understand the latter, consider the two dashed lines in Fig 2.1(a), each of which corresponds to a different chirality vector. All that is meant by chirality is that for the (n, n) case, \mathbf{C} is parallel to one side of the hexagons (which would appear ‘twisted’ about the CNT axis), while for the $(n, 0)$ case, \mathbf{C} runs straight across the hexagons (which would appear ‘straight’). It is indeed confirmed, by both theoretical and experimental studies[29], that the two physical degrees of freedom of a CNT, chirality and diameter, are completely defined by the integer pair.

2.4 Band Structure of Carbon Nanotubes

As with the physical properties of CNTs, their electronic structure can also be understood by starting with the electronic structure of graphene, and then accounting for the radial confinement of electrons in the CNT caused by the rolling up of the graphene sheet. The explicit calculation

of the band structure of graphene is far beyond the scope of this paper, and here I will only qualitatively discuss some of the structure's relevant characteristics. Using a tight binding model (see, for example, [33]) it is found that graphene is a zero-band gap semiconductor in which the valence and conduction bands overlap at two points in k-space³, \mathbf{K}_1 and \mathbf{K}_2 in Fig. 2.1(b)[26]. The dispersion function around these points is linear, which implies a conical band gap energy surface in k-space as in Fig 2.1(b).

When a finite piece of graphene is rolled up to form a CNT, there must be a perfect overlap of the two lattice points defined by the ends of \mathbf{C} , which gives the periodic boundary condition:

$$\mathbf{C} \cdot \mathbf{k} = 2\pi q \quad q = 1, 2, 3 \dots \quad (2.7)$$

where \mathbf{k} is the electron wave vector. This gives

$$\pi k_c D = 2\pi q \quad q = 1, 2, 3 \dots \quad (2.8)$$

where k_c is the circumferential component of the electron wave vector and D is the diameter of the CNT. As the radial confinement of the electrons has quantized k_c , \mathbf{k} is also quantized, which means that, in general, there will not necessarily be an allowed state with $\mathbf{k}_j = \mathbf{K}_i$, $i = 1, 2$ and j denotes the discrete nature of \mathbf{k}_j due to this new quantization. Rather the closest states to \mathbf{K}_i will be $\mathbf{k}_j = \mathbf{K}_i \pm \Delta k_c$. In other words the quantization of \mathbf{k} only allows certain planes bisecting the cones at $k_c = k_{c,j}$, as is shown in Fig. 2.1(b). Note that the cones are the band structure of graphene, and so these bisected curves define the band structure of the associated CNT, with the upper and lower curves defining the conduction and valence bands respectively. Thus unless there is some j such that $\mathbf{k}_j = \mathbf{K}_i$ there will be a band gap of $E_g = \hbar v_F \Delta k_c$ [26], between the conduction and valence bands of the CNT. The size of Δk_c is a function of chirality and any physical perturbations, such as curvature, axial strain and twist[26].

In summary CNTs can exhibit metallic or semiconducting behavior depending on the quantization of \mathbf{k} , which is dependent on \mathbf{C} , which is in turn a function (n, m) , and it is indeed found that a CNT is metallic when $n = m$, has a small gap when $n - m = 3i$ and is truly semiconducting when $n - m \neq 3i$, where i is an integer[3].

2.5 Effects of Deformation on the Band Structure of Carbon Nanotubes

CNTs are often termed 'quantum wires' or one-dimensional structures. These terms refer to the fact that any electrons in the conduction band of a CNT are confined in two dimensions, namely the trans-axial plane, and are only free to move in one single dimension: along the axis of the CNT. In this section I briefly show how deforming the CNTs from their relaxed state can cause significant shifts in their electronic structure, which can in turn cause axial confinement of electrons.

Deformation, or the transition from a circular to elliptical cross section of a CNT, can be achieved via AFM/STM manipulation, or other more creative techniques[28]. Using the density functional theory to simulate various types of semiconducting nanotubes, Mazzonni et. al.[25]

³k-space is related to momentum space by $\mathbf{k} = \mathbf{p}/\hbar$.

report notable band gap closure due to AFM deformation. Notably, the larger the degree of deformation, greater the decrease in band gap energy. They conclude that this band gap closure is due to the hybridization of bonds due to the increase in the curvature of the CNT as a result of the applied deformation. Cho[34] reports DFT calculations that support these results, confirming that the band gaps in semiconducting CNTs close under deformation of their cross sections. In contrast, deforming a metallic CNT causes its band gap to open, as shown by Chang[30]. Cho[34] also reports band gap opening in metallic CNTs as a result of deformation, noting that the magnitude of these shifts is less than those in semiconducting CNTs. There are many other results in the literature, all of which show that under deformation, semiconducting nanotubes show band gap closure, while the band gap of metallic nanotubes opens.

If these deformations are introduced locally, over a short axial distance, then the change in band gap energy can be used to confine electrons to a very small region of a CNT. For example, suppose that a deformation is introduced in a CNT by laying it across a ridge. Due to the finite width of the ridge, the deformation will not be discontinuous. Rather there will be a gradual transition from its normal relaxed, nearly circular cross section, to a more severe elliptical cross section. Due to the complex, nonlinear behavior of $\frac{dE_g}{d\eta}$ (where E_g is the band gap energy and η characterizes the deformation of the CNT), the band gap energy will oscillate creating numerous, alternating regions of semiconducting and metallic behavior. This results in various regions of confinement, isolated from the rest of the CNT by tunnel barriers consisting of local semiconducting regions: electrons in a one dimensional quantum wire have been further confined to create a zero dimensional QD.

The rest of this work will be dedicated to investigating the veracity of these theoretical predictions as applied to the carbon nanotube cross, both by attempting to fabricate and grow CNTs in this configuration, as well as by modelling the structure using various classical and quantum simulation methods, to which I now turn.

Chapter 3

Simulation Methods

I would now like to give brief descriptions of each of the theories used to model the carbon nanotube cross. As these methods are all standard within the field, I will try to avoid replicating pages of equations, leaving this to the many excellent references⁴, instead trying to provide some qualitative insight into the physical basis of these methods.

3.1 Molecular Dynamics

Consider a collection of interacting atoms. The basic idea of molecular dynamics (MD), is to consider the motion of each atom in a potential landscape, or mean field, that results from the forces of all the other atoms. These calculations are made using purely classical physics, and so, in a nutshell, MD consists of integrating Newton's Second Law, $F = ma$. The general idea of MD is to choose a particular atom, note its current position, sum the forces of all the other atoms acting on this chosen atom, and then move the atom in a direction and distance calculated from Newton's Second Law. This motion is calculated for a brief period of time, δt , short enough that we assume that the external force field due to the other atoms does not change, and therefore the acceleration remains constant. We then do this for all other atoms, therefore evolving the entire system from a time t to a time $t + \delta t$, and then repeat this for the duration of the simulation.

From mechanics we know that $\mathbf{F}(\mathbf{r}) = -\nabla U(\mathbf{r})$ and $\mathbf{a} = d^2\mathbf{r}/dt^2$, so for a given atom, we have

$$\mathbf{F}(\mathbf{r}) = m\mathbf{a} \tag{3.1}$$

$$-\nabla U(\mathbf{r}) = \frac{d^2\mathbf{r}}{dt^2} \tag{3.2}$$

Depending on the form of the potential U , it may be possible to solve this analytically. However, usually U involves the positions of every other atom, and so all of the differential equations are coupled, and it is usually necessary to solve the system numerically. Mathematically it is straightforward to discretize both space and time by choosing an appropriate time step and spatial grid, and then apply this discretization to the above set of differential equations, to obtain a set of finite difference equations which can be used to step the system forward from its initial co-ordinates. The system energy may easily be obtained simply by adding the kinetic energy, obtained from the time derivatives of the position coordinates, and the sum of the potential energies of each atom.

Given the initial co-ordinates, the only thing that remains is to choose a way of calculating the potential, U . Probably the simplest technique, for small scale objects where gravity is negligible,

⁴For example, see Balbuena[4] for Molecular Dynamics, Lundqvist[24] for Density Functional Theory and Ostlund[37] for Hartree-Fock, as well as Nojeh[27] for excellent introductions to all three theories.

would be to apply a Coulomb interaction to each pair of atoms. However, this ignores all manner of electromagnetic interactions such as van der Waals and dipole-dipole interactions, as well as the bonds between individual atoms. There are many different potentials in the literature, and the one used in this study is the AIREBO potential[36], an extension of the REBO[7] potential, which incorporates the Tersoff-Brenner potentials, the standard for modelling hydrocarbon atoms. The AIREBO potential has the form $U = U^{REBO} + U^{LJ} + U^{tors}$, where U^{REBO} is a Tersoff potential which models the bonds between atoms, and is essentially a sum of short range Coulomb type attractive and repulsive terms, with some modification using empirical constants. To model the inter-molecular interactions between non-bonded atoms two potentials are used. U^{tors} is a single minimum potential to model torsional interactions between atoms, while U^{LJ} is a classic 6-12 Lennard Jones potential used to model the van der Waals interactions between atoms. The reader is referred to the references for the mathematical details of these terms, but the important thing to realize is that having chosen this potential, one may discretize it, and insert it into the discretized form of Eq. 3.1. Then at each time step, the position of all atoms are fixed, U is calculated for all atoms, and then used to step forward the positions and velocities of all atoms. This process is then repeated for the duration of the simulation.

3.2 Quantum Chemistry Theories

While the MD model described above is adequate for calculating the geometrical configuration of atoms, as well as an estimate of the total system energy, it is purely classical, and so cannot give any information about many crucial properties of a nanoscale system, for example its band structure, and in order to examine these properties we must turn to models that incorporate quantum mechanics. Due to the complexity of nanoscale systems, it is necessary to develop approximation schemes to make the problem tractable, while still accounting for the small scale quantum effects that are crucial to the system's behavior. I will describe two so called *ab-initio* methods that start by seeking the exact solution to the Schrödinger equation, and make a series of approximations to render the problem workable.

3.2.1 Hartree-Fock Method

The Hartree-Fock (HF) method starts by seeking the exact solution, Ψ , to the many body problem. We first assume that the nuclei of each atom is fixed, the so called Born Oppenheimer approximation. Dynamic interactions between electrons are neglected, and are replaced by a mean field, which is an average over space and time of the effects of all the electrons in the system. In this so called Hartree approximation, we assume that the many body wave function is separable, and can therefore be written as a product of single electron wave functions, each of which is a solution of the Schrödinger equation with a Hamiltonian consisting of only the electron's kinetic energy and the mean field potential. Recall that, due to the indistinguishability of electrons and the Pauli exclusion principle, any wave function for fermions must be anti-symmetric with respect to electron exchange. Therefore the many body wave function may be written as a Slater determinant of single particle wave functions:

$$\Psi = \det \begin{pmatrix} \psi_1(\mathbf{r}_1) & \psi_1(\mathbf{r}_2) & \psi_1(\mathbf{r}_3) & \dots & \psi_1(\mathbf{r}_N) \\ \psi_2(\mathbf{r}_1) & \psi_2(\mathbf{r}_2) & \psi_2(\mathbf{r}_3) & \dots & \psi_2(\mathbf{r}_N) \\ \psi_3(\mathbf{r}_1) & \psi_3(\mathbf{r}_2) & \psi_3(\mathbf{r}_3) & \dots & \psi_3(\mathbf{r}_N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_N(\mathbf{r}_1) & \psi_N(\mathbf{r}_2) & \psi_N(\mathbf{r}_3) & \dots & \psi_N(\mathbf{r}_N) \end{pmatrix}$$

Here $\psi_i(\mathbf{r}_j)$ is the i th single electron wave function, applied to the j th electron. This expansion of Ψ is really the key of the HF method, as it separates the many body wave function into a product of single body wave functions.

The variational principle is invoked, which states that for any wave function Ψ , its energy expectation value is greater than or equal to the true ground state energy of the system, using the above Slater determinant as a trial wave function. Thus the (HF) method gives an upper bound on the system's energy. This variational principle is then applied iteratively, 'tweaking' the wave function at each iteration, in order to minimize the system energy. This results in a set of N single electron equations, the so called Fock equations:

$$f\psi_n = \epsilon_n\psi_n, f = h_{core} + \sum_{i=1}^{n/2} [J_i + K_i] \quad (3.3)$$

Here $h_{core} = -\frac{1}{2}\nabla_n^2 - \sum_A \frac{Z_A}{r_{An}}$ is the single electron Hamiltonian for an electron moving in the mean field electronic potential due to all the atoms, and the sum is over all electrons. J_i is the Coulomb operator, which accounts for electron-electron repulsion, by replacing the true two electron instantaneous Coulomb operator (proportional to r_{ij}^{-1}) with a one electron operator consisting of the average of this interaction over all space and spin. K_i is the so called exchange operator, which arises from the anti-symmetry requirement (ie. the indistinguishability of electrons), and gives the exchange energy of the electrons. Notice that the Fock operator, f , is essentially a single electron hamiltonian, with some extra energy terms to account for electron-electron interactions. However, as opposed to an exact treatment, which would involve double sums over all electrons, the Fock operator is only dependant on a single electron. This separability into N coupled equations is the crucial feature of the HF approximation that makes it possible to solve. In practice these N equations are recast as a matrix eigenvalue problem, and then the basis sets are iteratively adjusted using the variational principle subject to Eq. 3.3 until an energy minimum is reached. The set of ψ_i and ϵ_i are then recombined to find the Hartree-Fock estimate of the wave function and ground state energy, respectively.

3.2.2 The Density Functional Theory

Rather than treating the many body problem and seeking the many body wave function, Ψ , which is a function of all the coordinates of all particles, the Density Functional Theory (DFT) seeks the particle density, ρ over all space. This transformation from wave function to particle density is rigorously justified by two theorems published by Kohn and Hoenberg[12]:

1. There is a one-to-one correspondence between the ground state particle density and the ground state many body wave function. This is justified by proving that the external

potential acting on the exact many body system is uniquely determined by the particle density.

2. There is an analog to the wave function variational principle, described above, that states that the true particle density minimizes ground state energy. In other words the ground state energy of the correct particle density is less than the ground state energy of any other particle density.

Notice that this transformation results in a huge simplification. While Ψ is dependant on $3N$ coordinates, ρ is dependant on only 3. Essentially Kohn and Hoenberg proved that the wave function's dependence on the particle density may be inverted. That is given $\rho(\mathbf{r}) = \int |\Psi(\mathbf{r})|^2 d\mathbf{r}$ we can calculate a unique $\Psi = \Psi[\rho(\mathbf{r})]$. It can be shown that all other observables are also functionals of ρ , specifically the total energy: $E = E[\rho]$.

The actual method for finding the solution was layed out by Kohn and Shan[20], and involves the introduction of an effective potential. With this modification, the problem of finding the many body solution can be transformed to finding a set of non-interacting single body solutions, the Kohn-Sham equations. Each Kohn-Sham equation is an eigenvalue equation for one of the system's orbitals and its associated energy. Using the above variational principle in Theorem 2 the system energy is minimized using Lagrange's method of multipliers. This set of solutions to the Kohn-Sham equations can be combined to give the global particle density.

Practically this is implemented as follows. An initial density is chosen, and is used to calculate the effective potential. Note that this effective potential incorporates all the many body effects of the other electrons and atoms (such as electrostatics and exchange energy) into a single, lumped term. Next the Kohn-Sham equations are solved, from which a new particle density is calculated. This process is continued until the particle density converges.

While the details of DFT can be rather complex, the important point is that by seeking the particle density rather than the many body wave function, one can greatly simplify the problem. Furthermore, due to the above two theorems, all other relevant observables can be obtained from this particle density.

The methods described in this chapter, namely MD, HF and DFT, will be employed later to investigate the geometric and electronic structure of the carbon nanotube cross. However, as noted in the introduction, the ultimate goal of this work is to observe QD behavior at the junction of a real, physical CNT cross, and so I would like to describe the progress made to that end.

Chapter 4

Fabricating the Carbon Nanotube Cross

While the merits of the CNT cross have been discussed, the question remains how to actually physically realize a CNT cross. Furthermore, having built one, how do we actually see it, interact with it, and measure it. Clearly the very property which is its strength, its size, will also be an issue in light of the above challenges. In brief, the requirements for any experimental method are as follows. CNTs must be fabricated, and preferably in a manner that promotes the formation of a CNT cross. After growth it is necessary to make electric contact with the CNT cross, so that measurements can be made.

CNTs will be grown using Chemical Vapor Deposition (CVD) on a chip fabricated using microfabrication techniques. CNT crosses will be located using Scanning Electron Microscopy (SEM) imaging. Electrical contact can be made using both electrodes on the chip and an Atomic Force Microscope (AFM) tip. In order to investigate the possibility of QD behavior at the CNT cross junction, its Current-Voltage (I-V) characteristics can be measured. In this chapter I will describe this process in more detail, as well as the progress made and remaining challenges to observing a Quantum Dot in the CNT cross.

4.1 Design and Fabrication of the Chip

4.1.1 Design

Recalling that the diameter of CNTs is on the order of a nanometer, it is clear that making electrical contact with them will not be as easy as simply attaching wires at either end of the CNT using alligator clips. What is needed is a sort of ‘Nano Lab Bench’: a step down transformer, something that I can interact with, which in turn can interact with the CNTs. A rectangular ‘chip’ shown in Fig. 4.1(b) was designed to meet these requirements and is at the macro end of the step down transformer: having dimensions of $\sim 1\text{ cm}$ square, it is large enough to handle and put in a CVD furnace for CNT growth. Each chip consists of an array of 27×30 smaller ‘devices’, with equal numbers of four different variations of the same device design, shown in Fig. 4.1(c).

The design of this chip was intentionally kept simple, to reduce potential fabrication complications. The electrodes (red in Figs. 4.1(b) and 4.1(c)) are $100 \times 200\ \mu\text{m}$, and are large enough to make ohmic contact with using a probe station. Along the inside edge of the electrodes are strips of catalyst (black in Fig. 4.1(c)). The catalyst will be described later; for now consider it simply a place that initiates CNT growth. These strips are between five and ten microns wide, and are between 2 and 4 microns away from the edge of the electrode. This placement is crucial, as it ensures that any CNT that grows from these catalyst strips towards the gap between the electrodes will first travel over the electrode, thus ensuring good electrical contact between the

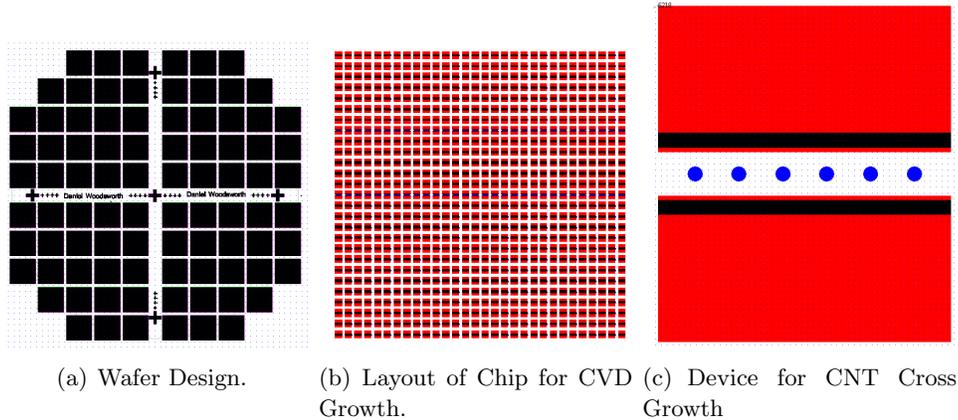


Figure 4.1: Fabrication Designs. An array of chips (b) is patterned over the entire four inch wafer (a), along with alignment marks for aligning the two layers (crosses in (a)). Within each chip (b) is an array of devices (c), consisting of two opposing electrodes (red), with strips of catalyst (black) along the edges of the electrode. In between the electrodes are catalyst islands (blue), and it is hoped that CNT cross growth will occur in the region defined by the catalyst strips and two adjacent islands.

CNT and the electrode. The blue circles are catalyst islands, of chemical composition identical to the catalyst strips.

The rationale for the design is as follows. Each rectangular area defined by two adjacent catalyst islands and two opposing electrodes is a potential region for CNT cross growth. CNTs will grow randomly from all catalyst regions, but it is hoped that the diamond type array of the four catalyst regions (two strips and two islands) will result in the occasional formation of CNT crosses at the center of this region. Unfortunately, it is probable that this will not occur with any regularity. To increase the yield of CNT crosses per device, there are between five and fifteen of these regions on each, depending on which the variant of the device. These variations are simply different combinations of: i) catalyst island size ($5 - 10 \mu m$); ii) catalyst strip width ($5 - 10 \mu m$); iii) distance between catalyst islands ($10 - 20 \mu m$) and iv) distance between catalyst island and electrode edge ($5 - 10 \mu m$). The reason for incorporating all of these variations is *a priori* there is no way of knowing what are the optimal growth conditions for CNTs. As a further means of offsetting the probable low yield of CNT crosses, there are 810 devices per chip, giving on the order of thousands of potential regions per chip for CNT cross growth. Finally, in order to obtain many chips from one fabrication process, the chip was replicated over the entire four inch wafer (see below), as is shown in Fig. 4.1(a). The spacing between these chips ($\sim 1 mm$) was chosen to ensure the chips were not damaged when the whole wafer was sliced into individual chips (see below).

4.1.2 Fabrication

The chips and devices can be manufactured using standard microfabrication techniques, which are the very well established methods that make all of micro electronics possible. While there are many excellent resources on the subject (see for example Franssila[11]), I will briefly outline the

process I have designed, and describe the challenges encountered.

The fundamental technique employed in this process is photolithography. In a nutshell, this is a resistive technique, similar to the plate lithography of the middle ages, and many wax resist techniques used in the arts. Starting with your substrate ('blank canvas'), in this case a silicon wafer, a layer of protective material is applied to the wafer. Then a template or stencil (a photomask) is used to selectively expose various parts of the protective material (photoresist) to light, which changes the chemical structure of the exposed photoresist, making it susceptible to certain chemical etching agents (developers). The entire wafer, with photoresist on top, is then washed in the developer, removing the parts of the photoresist that were exposed. Then, whatever material is desired is deposited over the entire wafer, in this case either metal, for the electrodes, or catalyst. Finally, in a procedure known as liftoff, the entire assembly is washed in acetone. This removes the remaining unexposed photoresist, and anything that sits on top of it, while not affecting the last deposited layer (metal or catalyst). Thus one is left with a layer of 'lithographically' patterned objects. This procedure may be repeated as many times as is required, each time resulting in a new layer of material.

For each different set of patterns, or layer, a different photomask is required and in this work, two were necessary: one for the electrodes and one for the catalyst strips and islands. The drawings of the chips and devices in Fig. 4.1 are directly taken from my photomask designs, which were fabricated at the University of Alberta Nanofabrication Facility. These photomasks are 5 inch square glass plates, and the exposure pattern is controlled by a thin layer of chrome, which has been etched away in certain places (wherever one wishes to have material in any given layer). In order to facilitate good alignment of the two layers, crucial to ensure that, say the catalyst strip, ends up being deposited where it is supposed to be, several alignment marks were included on each mask. These alignment marks consist of nested crosses, with the size of the cross on the catalyst mask (second layer) slightly smaller than on the electrode mask (first layer). By adjusting the mask until the smaller cross is centered in the larger cross, one can ensure alignment.

Fabrication was carried out in the AMPEL cleanroom, in order to avoid contaminating the relatively sensitive process. While simple in concept, microfabrication is in general tedious and time consuming. One significant challenge encountered was that, during the liftoff process, the unexposed photoresist⁵ was not completely removed in the acetone bath. This was a problem, as it is probably an organic polymer residue, which could have various, uncontrollable, negative effects, including contaminating the chip during CVD growth and invalidating the assumption that the substrate is an insulating equipotential. Various methods were tried to remove this residue, including washing the wafer in hot acetone, as well as sonicating the wafer in hot acetone, neither of which were successful. The one possibility that would work, an oxygen plasma etch to remove the residue, was not possible, as this would oxidize the metal electrode, hampering ohmic contact between the CNT and electrode, as well as quite possibly damaging the catalyst.

As a result of this residue issue, a different process was designed, involving a double resist method for each layer. Specifically, a layer of Poly(methyl methacrylate) (PMMA) resist was first deposited, followed by conventional photoresist (PR). The PR was then patterned photolitho-

⁵That it was the unexposed resist was confirmed by performing a blanket exposure on a wafer of patterned photoresist, and then washing it in the developer, which resulted in a clean wafer. Unfortunately this was not a viable solution to the problem, as the developer could have unforeseen effects on both the electrode and catalyst.

graphically, leaving exposed PMMA, which was then etched using an oxygen plasma. However, as the etch rates for the two resists are different, and there is no available data for these values for the machine in AMPEL, it has been very difficult to reliably etch all of the exposed PMMA, without completely etching away the PR/PMMA patterns (in other words completely etching all the resist, resulting in a clean, blank wafer). Furthermore, it is important to obtain accurate etch rates for both resists, in order to avoid oxidizing the electrode while patterning the second catalyst layer.⁶

Due to these challenges, the experimental side of this thesis is a work in progress and is still in the fabrication stages. However, before describing the work done modelling this structure, which is sufficiently encouraging to provide motivation to continue fabrication, I would like to outline a procedure that can serve as a basis for ultimately fabricating and measuring the characteristics of the CNT cross.

4.2 Experimental Design

This section is intended to be a brief illustration of the experimental design that has been devised. It is included because I believe that it is a relatively promising route to growing and measuring a CNT cross. However, as much of it has not been realized, I will be fairly terse. For example I will not explain the necessity of the PMMA layer, as that has already been discussed in the previous section.

4.2.1 Microfabrication

As described above, this is essentially a two layer photolithographic process, which is schematically depicted in Fig. 4.2.

Molybdenum Electrode Layer

Starting with a four inch silicon wafer, coated with a thin layer of thermal oxide, a layer of AZ P4110 photoresist ($\sim 1.4 \mu m$ thick) is spun on the wafer, followed by a layer of MicroChem 950 PMMA in Anisole (4% solids) ($\sim 350 nm$ thick). Following each spin step, the resist is baked on a hot plate and allowed to cool. The whole wafer is then exposed in a mask aligner, using the first layer mask (with electrode patterns). The wafer is then immersed in a bath of 1:4, (AZ 400K Developer: De-Ionized Water), removing the exposed PR. The wafer is etched with an oxygen plasma, removing the PMMA⁷. Following this 50 nm of Molybdenum is evaporated onto the wafer in a vacuum chamber. Finally the remaining resists (as well as any *Mo* on top of the resists) is removed using an acetone bath.

Catalyst Layer

This layer is almost identical to the *Mo* layer, with a few notable exceptions. The PR and PMMA are applied in exactly the same manner, and the photolithography is identical, except that the

⁶In this case, the electrode layer is already deposited, with the PR/PMMA layer on top. We wish to etch down to the electrode, so that we may deposit catalyst. However, if the oxygen plasma etch is continued for too long, then the metal electrode will be oxidized, which is undesirable as described above.

⁷Much of the remaining PR will also be removed

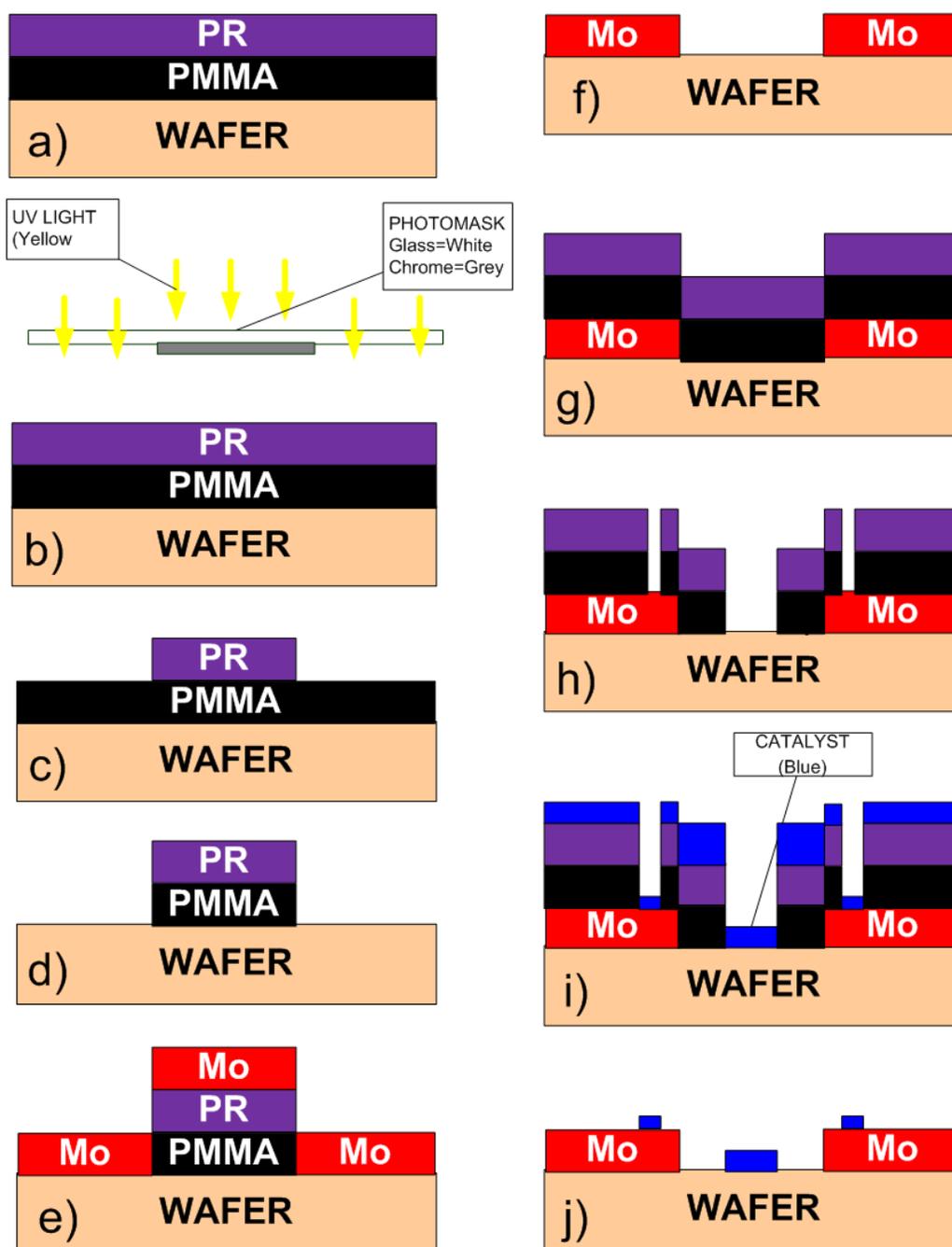


Figure 4.2: Schematic of Microfabrication Process of Chip for CNT Growth in CVD. a) Spin PR and PMMA resist layers; b) Expose to UV light, using mask with electrode patterns; c) Develop PR; d) Etch PMMA using O_2 plasma; e) Evaporate Mo ; f) Liftoff of remaining PR/PMMA (and Mo on top of resists) using acetone; g) Spin PR and PMMA resist layers again (over both wafer and Mo); h) Repeat photolithography (steps b-d)) using mask with catalyst patterns; i) Spin catalyst over entire assembly and j) Liftoff of remaining PR/PMMA (and catalyst on top of resists) using acetone, leaving completed device. Note that the profile in j) corresponds to the device in Fig. 4.1(c).

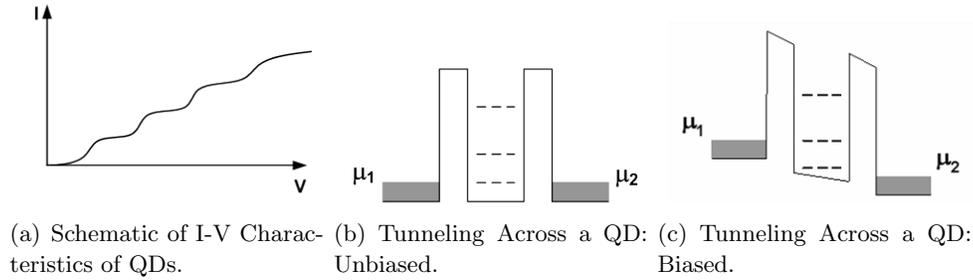


Figure 4.3: Current-Voltage (I-V) Measurements of Quantum Dots. The quantum well formed by two potential barriers, (b), is a model for a QD. Electrons cannot tunnel through the entire assembly but may use the states inside the QD (dashed lines) as stepping stones. As a voltage is applied across the QD, the chemical potential shifts, (c), and more and more states become available as stepping stones, subject to $\mu_1 > E_i > \mu_2$. Each new state that becomes available with increasing voltage results in a current spike, giving rise to the characteristic staircase-like I-V curve, (a).

second layer mask (with catalyst strip and island patterns) is used. The developing of the PR and etching of the PMMA is also identical. Then the catalyst, which initiates CNT growth, is spun on to the entire wafer, after which it is baked, to remove the methanol solvent (see below). Finally liftoff is performed with acetone, leaving the finished wafer. Note that the profile in Fig. 4.2j) corresponds to the device shown in Fig. 4.1(c).

Finally the four inch wafer is cut into the individual chips using a dicing saw.

4.3 Carbon Nanotube Growth

The general methodology of growing CNTs on micro-fabricated devices using chemical vapor deposition (CVD) was initially developed by Dai and colleagues[21], and is essentially followed here. The broad strokes of this scheme are as follows. A chip is placed in a CVD furnace, which is heated to around 700°C , at which point methane (CH_4) is injected into the furnace. It is here that the catalyst⁸, which consists of iron and molybdenum nanoparticles in an alumina support matrix, is crucial. Due to the high temperature, the gas molecules break down at the surface of the catalyst nanoparticles, and the now free carbon atoms spontaneously form CNTs. While the physical mechanism of this CNT growth at the catalyst particle surface is still under debate, the salient feature is that it is spontaneous. Intervention is not required: as long as the chemicals are inside the CVD in the right concentrations, at the right temperature, CNTs will grow.

4.4 Measuring the Current-Voltage Characteristics of the CNT Cross

While it is known that CNTs will grow in the above conditions, and that, given enough chances, there will be a few CNT crosses, these last still need to be located on the chip. While the lithographically patterned features can be resolved with an optical microscope, CNTs are too small for this, and so the chip must be imaged using a Scanning Electron Microscope (SEM)⁹. Due to the regular array the devices form on the chip, it should be fairly easy to quickly cycle through the chip, visually searching for CNT crosses row by row.

Having located a CNT cross, in order to make measurements it is necessary to make electrical contact with this structure, and there are two cases to consider. Ideally one of the CNTs in the cross will be in electrical contact with both of the electrodes. In other words, one end of the CNT, on one side of the junction, will be touching one electrode, and the other end, on the other side of the junction, will be touching the opposing electrode. In this case electrical contact may be made with each electrode separately. More probably, only one of the ends of one of the CNTs will be in contact with a single electrode. In this case, in order to complete the circuit for measurement, contact with the other, free, end of the CNT must be made either using direct contact via an Atomic Force Microscope (AFM) tip or via a tunneling current, using a Scanning Tunneling Microscope (STM). As both of these microscopes are also imaging modalities, it should be possible to see the CNT cross, and then ‘zoom in’ and make contact with the appropriate region of the CNT. The circuit is completed by making contact with a probe station to the electrode that is in contact with the other end of the CNT.

Once contact has been made, Current-Voltage (I-V) measurements can be taken. A DC voltage bias is applied across the CNT cross (the possible QD), and the voltage is gradually increased, all the while measuring the current flowing through the CNT cross. The current is then plotted versus the voltage¹⁰. What shows the presence of a QD is the staircase type behavior shown in Fig. 4.3(a): each ‘step’ is characteristic of one of the quantized energy levels in the QD.¹¹

How this staircase-like current confirms the presence of a QD can be heuristically understood as follows. Consider, as a toy model of a QD, a uniform potential, with two very high, closely spaced potential barriers, as depicted in Fig. 4.3(b). The region in between the two barriers is the model QD, with its set of discrete states (dashed lines). Initially the chemical potential of the electrons is the same (Fig. 4.3(b)) on either side of the barriers (in other words on either side of the CNT junction). However, when the voltage bias is applied, the chemical potentials will

⁸The catalyst consists of 0.05 mmol of $Fe(NO_3)_3 \cdot 9H_2O$, 0.015 mmol of $MoO_2(acac)_2$ and 15 mg alumina nanoparticles (aluminum oxide C) in 15 ml of methanol. The entire mixture was sonicated for twenty four hours following initial mixing, and then for one hour prior to each use. See [21] for details.

⁹Alternatively a Scanning Tunneling Microscope (STM) or Atomic Force Microscope (AFM) could be used

¹⁰This is not crucial, V vs. I is also possible- the important feature is the staircase behavior

¹¹In reality this behavior is a combination of both the quantum effect of the discrete energy states in the QD as well as the purely classical electrostatic interaction between electrons. As the QD has a very small area, it therefore has a very small capacitance, and so each additional electron that is added to the QD has an energy greater than the previous by a value of $\frac{e^2}{C}$, where e is the fundamental charge, and C is the capacitance of the QD. Since C is small, this is an appreciable energy. This effect is known as the Coulomb blockade. Thus the simple capacitive charging of the QD will also show a discrete behavior, and it is necessary to decouple the two effects, which while sometimes difficult, is possible.

shift on both sides, as shown in Fig. 4.3(c). From statistical mechanics, particles tend to flow from high to low chemical potential. Now if, instead of two separate barriers, there was simply a large, single barrier of equal height and as wide as both barriers and the well, then this tunnel gap would be too large for any kind of appreciable tunnel current. However, the discrete states of the QD can act as ‘stepping stones’ through the QD: an electron may first tunnel through one barrier, into one of the allowed states of the QD, and then through the second barrier, and around the circuit, which will register as a current. Recall that, without extra energy, a particle can never move from a state of lower chemical potential to higher chemical potential. Thus the only states that are able to act as stepping stones are those with $\mu_1 > E_i > \mu_2$, where μ_1 and μ_2 are the chemical potentials on either side of the QD and E_i is the energy of the i^{th} allowed state in the QD. Initially as $\mu_1 = \mu_2$ no current flows. As the voltage increases, still no current flows, until the voltage is large enough so that the above inequality is satisfied for at least one state, at which point some electrons will tunnel through the QD, as described above, and so there will be a spike in the current. However this state will quickly saturate, and so the current will plateau. As the voltage is further increased, μ_1 keeps rising, and eventually another state will satisfy the inequality, leading to another spike in current. This leads to the staircase current shown in Fig. 4.3(a), indicating the presence of the QD.

This experiment is still in the fabrication stages, and so the above discussion is somewhat speculative. However, it should provide at least a framework for eventually performing I-V measurements on a CNT cross. There is good reason for believing that one may indeed see QD behavior in these measurements, in large part due to the simulation studies described in the following sections.

Chapter 5

Modeling the Carbon Nanotube Cross

To gain another perspective of, and more insight into, the physical processes that govern its geometrical and electronic structure, the CNT cross structure was modeled using a variety of methods. Broadly these simulations fall into two categories: Molecular Dynamics studies of the system's geometry and Quantum Chemistry calculations (using Hartree-Fock and Density Functional Theory) of its electronic structure.

Modelling nanoscale devices presents unique challenges, as their length and energy scales tend to be right on the cusp, or overlap, of two different regimes. They are small enough to exhibit all manner of inherently quantum phenomena, for example the very discretization of energy levels that is integral to this work, and therefore any efforts to model systems on this scale must at least consider quantum effects. However, they are large enough, and have enough protons, electrons and neutrons so as to be virtually impossible to model using the full, unaltered machinery of quantum mechanics. In other words one cannot simply write down the Hamiltonian for all the particles of the system and solve for the wave function. That being said, one cannot go too far in the other direction, and treat the system in a continuum fashion, using say the classical theories of electrodynamics, mechanics, and electronic transport equations.

Thus any attempt at modeling devices at this scale usually must incorporate a variety of different methods for different parts of the simulation. In this section I will describe how I have modelled the CNT cross structure, for the most part avoiding technical discussion, as this has already been done in Sec. 3, and rather focusing on the methodology adopted, and the various results that have been obtained and seem to support the possibility of QD behavior in the CNT cross.

5.1 Geometry

To determine the geometry of the system, that is the physical configuration and orientation of all the atoms, the Molecular Dynamics methods described in Sec. 3.1 were employed. This is certainly somewhat of an approximation, as in theory the atoms will align themselves and create and break bonds according to the laws of quantum mechanics. As there were on the order of seven thousand atoms in the system it would be impossible to model using quantum methods, and the AIREBO potential is an excellent approximation.

As a starting point for any simulation an initial configuration should be chosen that mimics experiment as closely as possible. This was chosen to be a flat sheet of graphene (used as an approximation to the silicon substrate), with one CNT on top of it, and the other draped over the substrate and the first CNT, as shown in Fig. 1.1(a). A Matlab script was written to generate

a hexagonal array of carbon atoms in a plane, which is exactly the structure of graphene. A program called `Wrapping`¹², developed by Dr. Shigeo Maruyama from The University of Tokyo, was used to generate co-ordinates of flat CNTs of arbitrary length and chiral indices (m, n) . The CNTs were 10 nm in length and the graphene sheet was 11×11 nm.

Recall from statistical mechanics that, within its constraints, a system will evolve to the state with the lowest possible energy, corresponding to its equilibrium state. This is a very convenient way of obtaining the physical, or ‘true’, geometrical state of a system: let it evolve in a MD simulation until its energy saturates (approaches some value asymptotically). Ideally it would be nice to simply use two flat CNTs, the one over top of the other, both over top of the graphene substrate, as seen in Fig. 5.3(a), as the starting configuration for the MD simulation, and then let the top CNT slowly drift downward to the substrate, eventually conforming itself to the substrate and the bottom CNT. Unfortunately this does not happen, for two reasons. First, the inter-molecular interactions in the AIREBO potential are relatively short range. Therefore, as will be discussed below, the top CNT, being around 1 nm above the substrate, simply does not ‘see’ the substrate, and behaves as though it is not there, and therefore does not ‘feel’ the attractive van der Waals forces that physically should drag the CNT downward. Secondly, even if this interaction were realized mathematically, it would be very weak, requiring the simulation to be run weeks or even months.

Because of this, an initial configuration needed to be developed that was as close as possible to the suspected physical configuration, such that the MD simulation is ‘closer to its final goal’, and need only minorly re-arrange atoms to reach an equilibrium state. Noting that the SEM image of the CNT cross shown in Fig. 5.3(c) has a bell curve shape, I decided to map the top CNT co-ordinates to a Gaussian profile according to the transformation $(x, y) \rightarrow (x, y')$ where

$$y' = h \left(1 - e^{-\frac{x^2}{4 \ln(2) w^2}} \right) \quad (5.1)$$

Here h is the height above the substrate of the centre (peak) of the top CNT, and w is the Full Width at Half Maximum (FWHM) of the Gaussian function. Having completed this mapping, the results of which are shown in Fig. 1.1(a), another question immediately arises, namely what values to use for the free parameters in the simulation. These are i) the type of CNT, which, from Sec. 2.3, is a function of the integers (m, n) and ii) the width of the Gaussian profile (w), that is whether the CNT has a very gradual rise from the substrate to the junction, or rather remains flat on the substrate, and then closely contours the bottom CNT at the junction, and then falls quickly back to the substrate and is flat on the substrate for the rest its length.

There is no way to know *a priori* what combination of these parameters one would obtain from CNT growth, nor is there any way to ascertain this information from imaging or measurements. Thus, it is appropriate to investigate a wide array of these parameters, in various pairings, examples of which are shown in Fig 5.1, and see the effect this has on the geometrical and electronic structure of the CNT cross. The two types of CNT investigated were the semi-conducting (10,0) and the metallic (5,5), and for each of these a range of ‘widths’ were investigated ($w = 15, 20, \dots, 45, 50$) as well as a ‘flat’ CNT cross (Fig. 5.3(a), identical to other CNT crosses, but without the Gaussian mapping, for comparison. Because in an actual physical experiment the CNTs are growing on macroscopic silicon wafers, which are structurally rigid, the

¹²Freely available at <http://www.photon.t.u-tokyo.ac.jp/%7Emaruyama/wrapping3/wrapping.html>.

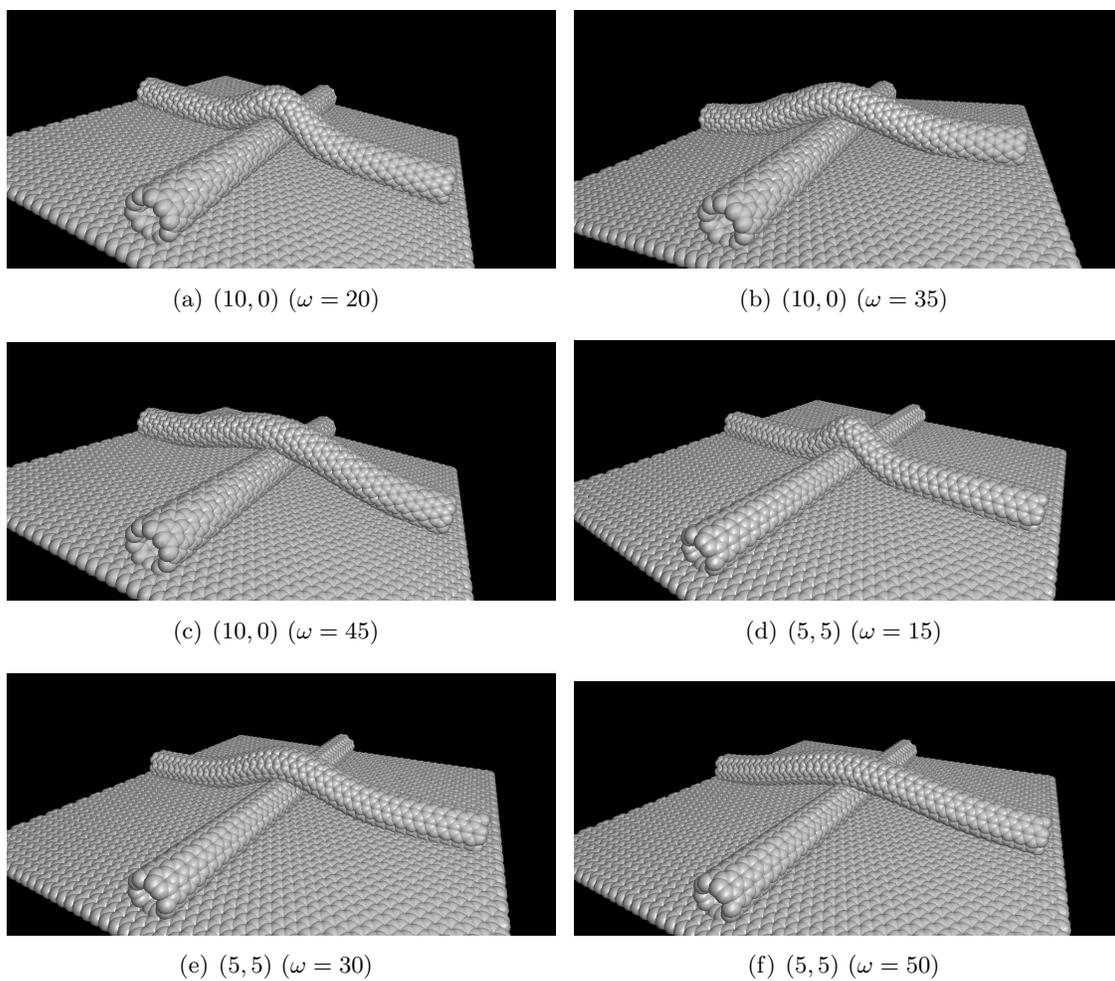


Figure 5.1: Comparison of Gaussian Profiles of CNT Cross. Here (m, n) are the integer indices of the given CNT, and ω is the Full Width at Half Maximum of the Gaussian function used to generate the co-ordinates. These are examples of the profiles which were obtained using MD relaxation.

graphene sheet was fixed throughout the simulation, as would be the wafer.

The MD simulations were done in NanoHive¹³ and the results are shown in Fig. 5.2. At first glance, these results are extremely worrying. As the width of the Gaussian profile increases, the potential energy of the system decreases, both initially, and throughout the simulation. Furthermore, the ‘flat’ CNT cross (Fig. 5.3(a)) has the lowest energy of all. As it is known that CNT crosses do indeed grow (see Fig. 1.1(d) for example), and are therefore physically realistic states, it would seem that the MD implementation is flawed, casting doubt on these results.

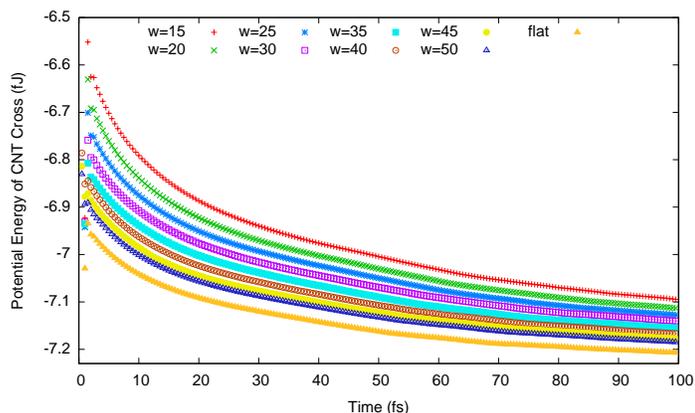
It turns out that the MD method is flawed, but in a controllable and non-catastrophic manner. To investigate this issue, a series of simulations were performed, consisting of a straight, flat CNT, placed at various heights h above the substrate, depicted in Fig. 5.3(b). These systems were then relaxed via MD in exactly the same manner as the CNT crosses, and the results of these simulations are shown in Fig. 5.2(c). Notice that only the nanotube that is 1\AA above the substrate, exhibits any kind of potential energy minimization. Furthermore the potential energy curves for the two higher nanotubes (5.5\AA and 10\AA above the substrate) are identical. These results are explained by the fact that the effective range of the inter-molecular interactions, modelled using the Lennard-Jones potential (see Sec. 3.1), is relatively short, and can be interpreted as follows.

First, the only nanotube which changed its configuration in any way, and evolved to a lower energy state, is that which was initially closest to the substrate. Thus, the effective range of the inter-molecular interactions is sufficient such that this close CNT interacts with the molecules of the substrate, but small enough that the two more distant CNTs do not. Since they are not interacting with the substrate, they behave as if they were isolated. As the initial co-ordinates of a simple CNT are already essentially in a relaxed state, there is no energy minimization possible. This observation is reinforced by the identical energy curves of both of the more distant nanotubes. While the geometric configuration of these two systems *is* different- one is twice as far as the other from the substrate- they show identical behavior. This is only possible if neither of the CNTs interact with the substrate, implying that the effective range of the inter-molecular interactions in the AIREBO potential is smaller than 5.5\AA .

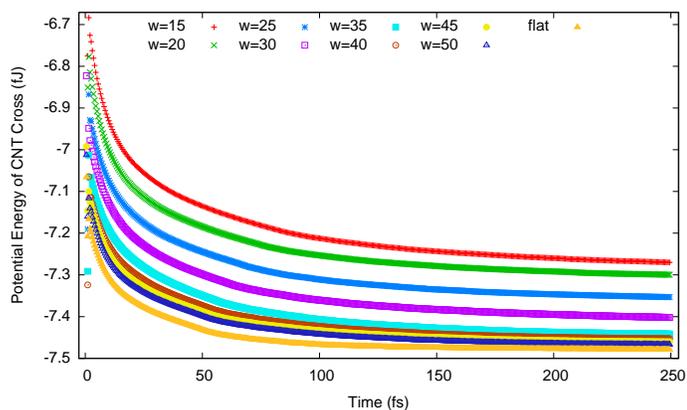
Returning to the relaxation of the various Gaussian profiles, shown in Figs. 5.2(a) and 5.2(b), this also explains the decreasing energy of the CNT crosses with increasing spread. For a very narrow spread, for example $w = 15\text{\AA}$ in Fig 5.1(d), the CNT is interacting with the substrate for most of its length, forming covalent bonds. Thus the CNT is slightly deformed, having an oval cross-section, and is not free to relax to its circular, minimum energy, profile. Conversely, for very large spreads, say $w = 50\text{\AA}$, Fig 5.1(f), much of the CNT is well above the substrate, and so a greater portion of the CNT is free to evolve to its lowest possible energy state, without any ‘interference’ from the graphene, as it is out of the effective range of interaction. Carrying along in this vein, the flat CNT cross is out of range of the substrate, and so is virtually free to evolve to its global minimum energy.¹⁴ Therefore this decrease in energy with increasing spread is an artifact of the MD simulation, specifically the AIREBO potential.

¹³Freely available from <http://nanoengineer-1.com/nh1/>

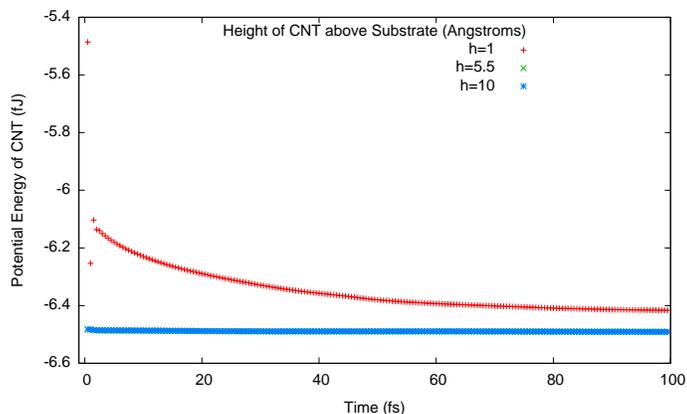
¹⁴That the potential energy curve does show a minimization is also due to the interactions of the two CNTs as well as the interactions between the bottom CNT and the substrate. Also recall that the substrate is fixed in space, and so has no effect on these arguments, as it cannot rearrange to any other configuration to reach a lower energy state.



(a) MD Relaxation of various (5, 5) CNT Crosses

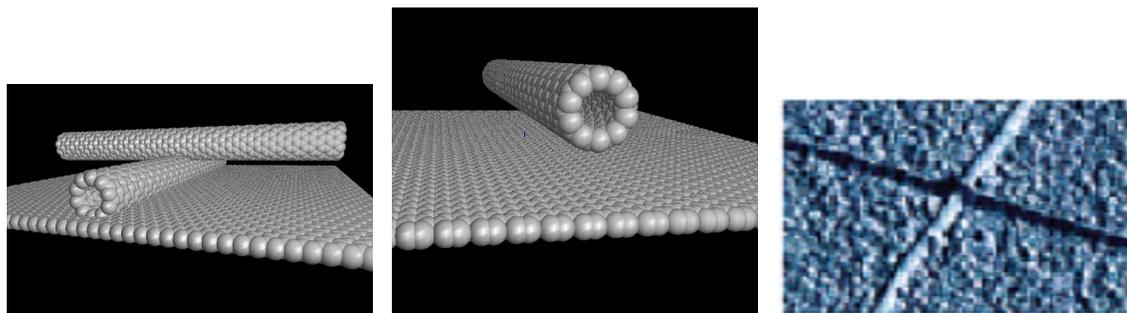


(b) MD Relaxation of various (10, 0) CNT Crosses



(c) MD Relaxation of Flat CNT above Substrate

Figure 5.2: MD Relaxation of Carbon Nanotube Crosses. For both types of nanotube, the potential energy of the structure, for all values of w , saturates, implying a set of equilibrium, meta-stable states, which are physically realistic. Note that, although the flat CNT has the lowest energy, this is an artifact of the relatively short range of the inter-molecular term in the AIREBO potential. This is confirmed by (c): notice that only the CNT that is 1\AA above the substrate evolves to a lower energy state, by interacting with the substrate. The other two CNTs do not ‘see’ the substrate, and so behave as if they were isolated.



(a) Flat Carbon Nanotube Cross. (b) Single Flat CNT Above Substrate. (c) SEM Image of a CNT Cross.

Figure 5.3: Investigating the Inter-Molecular Interactions in NanoHive. While the MD relaxation results (Figs. 5.2(a) and 5.2(b)) indicate that a CNT cross would not form in nature, instead staying in a flat configuration (a), they have been observed (c). To resolve this issue MD simulations were run with a single CNT above the substrate (b), at varying heights. These results are shown in Fig. 5.2(c).

Having eliminated this one concern, the other obvious feature of these results is that the potential energy of all the various CNT crosses saturates with time. This implies that each of these CNT crosses, both (10,0) and (5,5) CNTs, and for all values of w , are meta-stable states. That is, once the carbon atoms are in this initial configuration, barring some fairly severe perturbation, these CNT crosses will not spontaneously switch to some other physical arrangement. Therefore all of these structures are potentially viable, and could possibly occur experimentally. As such, a selection of these structures was studied, to see if their electronic structures varied. At this point, Molecular Dynamics has fulfilled its role: it has provided equilibrium CNT cross configurations and demonstrated that viable CNT crosses may occur in a variety of combinations of both nanotube type, and width of Gaussian profile.

5.2 Electronic Structure

One might wonder why it was ever advisable to use a classical theory such as Molecular Dynamics to model a QD dot device, whose very name suggests a quantum mechanical description of the CNT cross structure is required. While the MD package used in this work does an excellent job of empirically accounting for quantum effects such as chemical bonds, it is certainly an approximation, and MD was chosen out of necessity due to the large size of the CNT structure, and because for geometry calculations, quantum effects are less crucial

When calculating the electronic structure of the system, however, it is imperative to turn to the quantum chemistry methods described in Sec. 3.2. Unfortunately, due to the vastly increased complexity of these methods, the total system cannot be simulated at the same time. Indeed these methods are so computationally expensive, that even the single top CNT, namely that which is draped over the bottom CNT and substrate, cannot be simulated in its entirety. This is because for each electron that is added to the system, the overlap and exchange integrals of this new electron must be calculated with every other electron already present. It follows that the

		GAUSSIAN Basis Set				
		STO-3G	6-31G	6-311G(d)	6-311++G(d,p)	
Numerical Simulation Method	Hartree Fock	HOMO (E_n)	0.07380	-0.03566	FAILED	FAILED
		LUMO (E_n)	0.20173	0.08826		
		Band Gap (eV)	3.479696	3.370624		
		CPU Time	00:21:02	4:54:50		
	DFT (BLYP)	HOMO (E_n)	0.08067	-0.02828	FAILED	FAILED
		LUMO (E_n)	0.08793	-0.02612		
		Band Gap (eV)	0.197472	0.058752		
		CPU Time	1:26:11	14:32:24		
	DFT (B3LYP)	HOMO (E_n)	0.06808	FAILED	-0.05846	FAILED
		LUMO (E_n)	0.09792		-0.04136	
		Band Gap (eV)	0.811648		0.46512	
		CPU Time	00:53:42		27:09:10	
			Tight Binding Method			
			AM1	PM3		
	Semi Empirical	HOMO (E_n)	-0.25440	0.07380		
		LUMO (E_n)	-0.01842	0.20173		
Band Gap (eV)		6.418656	3.479696			
CPU Time		00:01:05	00:00:49			

Figure 5.4: Comparison of different electronic structure calculation methods in the GAUSSIAN package. While the STO-3G basis sets are faster, they greatly overestimate the band gap energy. The method used to perform electronic structure calculations on the CNT cross was the DFT method B3LYP using a 6-311G(d) basis set, as electronic structure calculations using this combination gave a bandgap of 0.57 eV, similar to the empirical values of $E_g \approx 0.5 - 1.5$ eV.

number of integrals that must be computed, and therefore the computation time of the simulation, exhibits a power law type dependence on the number of electrons in the system.

The only option is to split the CNT into more manageable sections, and calculate the electronic structure of each of these sections, and then re-assemble these results into a picture of the electronic structure of the CNT as a whole. The top CNT was split into 5\AA slices, centred about the CNT cross junction. In order to approximate these slices as though they were part of a continuous CNT, the atoms at the ends of each slice were terminated with hydrogen atoms, thus avoiding any dangling bonds, which are notorious for causing problems in electronic structure calculations.

Each of these slices contained forty carbon atoms, a far more manageable number than the seven thousand odd atoms that make up the complete system that was relaxed using MD. Electronic structure calculations were performed in GAUSSIAN¹⁵, using the hydrogen terminated CNT

¹⁵Not freely available anywhere. See www.gaussian.com. Calculations were performed on Abacus, a cluster

slices as input. GAUSSIAN has a vast array of different methods available, most based on the Hartree-Fock (HF) and Density Functional Theory (DFT) methods described in Sec. 3.2. While there are certain trends in these methods, for example using more basis functions in the expansion of a wave function often produces more accurate results, there is no definitive best choice for any particular system.

A series of calibration simulations were performed, to determine which method to use for the electronic structure calculations. A straight CNT, 7.5Å long, was generated in Wrapping, and its electronic structure calculated using HF, DFT and so-called Semi Empirical Methods¹⁶, the last of which was used for calculations not described here. Within the two former methods, one must make a choice of basis set. These basis functions are used in the expansion of the wave function in both algorithms. The mathematical details of these bases are fairly complicated, and are described elsewhere¹⁷. Suffice to say that from the ST0-3G basis set towards 6-311++(d,p) basis set, there is an increasing number and variety of basis functions assigned to each atom, which enables more accurate modeling of higher level molecular orbitals. The tradeoff for this increased accuracy is, not surprisingly, drastic penalties in CPU time. Within DFT there are two common variations, namely BLYP and B3LYP, with the latter an extension of the former, as the latter also solves the HF problem.¹⁸

Essentially, what GAUSSIAN returns as output from these electronic structure calculations, in this case using DFT, is a hierarchical arrangement of molecular orbitals¹⁹, and their associated energies, along with other values calculated from these two fundamental quantities. Among *many* other things, GAUSSIAN outputs the energy (E_H) of the Highest Occupied Molecular Orbital (HOMO) and the energy (E_L) of the Lowest Unoccupied Molecular Orbital (LUMO). These two orbitals have a straightforward physical interpretation. Consider the discrete set of allowed molecular orbitals of a molecule, ordered from lowest to highest energy. Then consider taking all of the electrons from all of the atoms, and filling these discrete *electronic* states with electrons, from lowest to increasing energy, thus constructing the molecule in its *global* ground state. The molecular orbital state in which the final electron is placed is the HOMO state, while the next highest energy state is the LUMO state. In this work the band gap energy was taken to be $E_g = E_L - E_H$, as this is the energy required to excite an electron across the ‘band gap’ from the ‘valence band’, or HOMO, which is full and so electrons cannot move to conduct charge, to the ‘conduction band’, or LUMO, which is, in general, empty and so electrons are free to move and

maintained by UBC Chemistry.

¹⁶These are essentially hybrid methods which use many first principles results, but also use empirically obtained parameters to simplify the calculations.

¹⁷The Gaussian online manual (http://www.gaussian.com/g_ur/g03mantop.htm) has basic descriptions of these bases, and references to all of the original papers, in which each basis (and for that matter, exchange functional, simulation method etc.) was first described.

¹⁸In general the two main choices to be made when running a DFT simulation are the exchange functional and the basis sets. The former is used when making the transformation to particle density from the wave function, while the latter is used to expand the single particle wave function (in an effective potential which accounts for the other particles), which is solved as a final step in DFT.

¹⁹These molecular orbitals have a somewhat convoluted relationship to other, more mathematical and less physical so-called spin orbitals, which are the actual orbitals that are used in the wave function expansions. This relationship is somewhat complicated, and far beyond the scope of this work, especially as they are simply a means to an end, namely the band structure of the CNT cross, and furthermore they are employed in a very large, well validated commercial package.

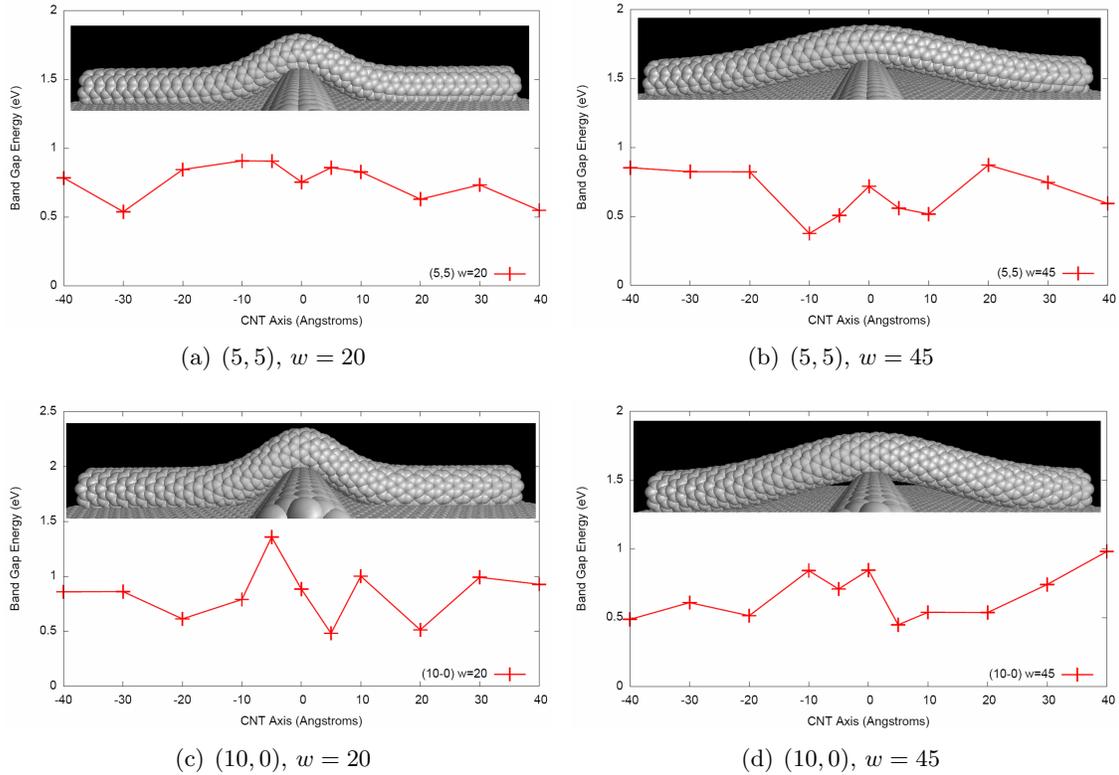


Figure 5.5: Band Gap Energy of CNT Crosses. The actual CNT that was used as the physical structure for the electronic structure calculations is shown above each plot. Notice that there is a one-to-one correlation between local deformation of the CNT and local band gap change, which supports the possibility of a QD at the CNT cross junction.

conduct charge. The validity of this choice of definition of band gap energy will be discussed in Sec. 6; for now the important point is that it represents the energy necessary to move an electron into a region in which it is free to move throughout the structure.

Returning to the results of the calibration simulations, shown in Fig. 5.4, it is immediately evident how the more complex basis functions result in more “computational complexity”: for even an array of forty atoms, all methods fail after around five days, having failed to converge. The most significant result of these simulations is that, as is generally accepted, the HF methods greatly overestimate E_L , compared to DFT methods, which also results in greatly inflated band gap energies. A similar effect is observed with the relatively inaccurate ST0-3G basis set, making all HF methods and any DFT methods using this basis set undesirable. As all of the simulations using the highly accurate 6-311++(d,p) basis set crashed, along with the DFT method(B3LYP, basis set 6-31G) there were really only two choices: 1) B3LYP with a 6-311G(d) basis set or 2) BLYP with a 6-31G basis set. In this case, the added accuracy provided by the extra basis functions was warranted, as the simulation using the B3LYP/6-311G(d) combination gave $E_g = 0.47 eV$, much closer to the empirical values of $E_g \approx 0.5 - 1.5$ [38] than those obtained using the BLYP/6-31G combination ($E_g = 0.59 eV$). Therefore the B3LYP method using a 6-311G(d) basis set was used for all subsequent electronic structure calculations.

Having chosen a DFT method, and adopted a definition for the band gap energy, it is possible to calculate the electronic structure for various CNT cross structures, specifically those with $w = 20, 45\text{\AA}$ for both (10, 0) and (5, 5) CNTs. The band gap energy for each slice was calculated in GAUSSIAN and then plotted as a function of the location of the center of the slice on the ‘parent’ CNT. These plots are shown in Fig. 5.5, and on the whole are promising. Probably the most important result is the global observation that there is a one-to-one correlation between local deformation of the top CNT and local band gap change of the CNT, across all different CNT cross structures. This is an encouraging first step towards showing QD behavior at the junction of the CNT cross, as the original stated goal was to create a QD in a CNT by introducing local physical deformations, in the hopes of inducing local band gap change, to create tunnel barriers which would axially confine electrons in the CNT. This observation theoretically confirms the first half of this goal.

There are several other important features of these plots. Consider the case of the metallic (5, 5) CNT, with a wide Gaussian profile ($w = 45$), shown in Fig. 5.5(b). As discussed in Sec. 2.5, deforming a metallic CNT causes its band gap to open, and this is observed here. Moving from left to right initially the CNT is flat on the substrate, and as such the CNT is slightly deformed, ‘squished’ from a circular to oval cross section, similar to an inflatable ball which has lost much of its air resting on the ground. This causes an opening of the band gap. As the CNT rises off the substrate, it is free to be in a more relaxed shape, circular in cross section, and so has a much smaller band gap. At the junction, the CNT is once again deformed, this time by the bottom CNT, and so the band gap rises again. Then, symmetrically, the band gap lowers once again (as the CNT is in a free, non-deformed state), before rising again, as it hits the substrate and is deformed.

Now examine the semi-conducting (10, 0) tight profile ($w = 20$) cross, shown in Fig. 5.5(c). For semi-conducting CNTs, deformation causes a band gap closure (Sec. 2.5), and so the band gap profile should be just the opposite of that discussed in the previous paragraph, and it is. Starting in a mildly deformed state resting on the substrate, it has a decreased band gap. As the Gaussian profile is quite narrow, the rise off the substrate is quite sharp, leading to a significant deformation, and a further decrease in the band gap at around -15\AA , as the CNT comes off the surface. Once it is free of the surface, it is in a non-deformed state, and so the band gap increases, until the CNT hits the other, bottom CNT, causing deformation and a decreased band gap once again. As above, this process is mirrored as the CNT falls back to the substrate.

While the other two plots do not have so clean an interpretation, they do exhibit similar features. The question then becomes what is the effect of the type of CNT (m, n) and the configuration of the CNT cross (w). About the only valid observation that can be made is to note that the two different types of CNTs behave differently in response to deformation, as they should as one is metallic and one is semiconducting. It is not possible from Fig. 5.5 to make any conclusions as to which Gaussian profile, or which type of CNT, is optimal for QD behavior, nor is it possible to state that these parameters have no effect.

As a final note, observe that for all of the CNT crosses, the calculated band gaps ($\sim 0.5 - 1.5\text{ eV}$) are in a normal range, as compared to the experimentally observed range of $\sim 0.5 - 2\text{ eV}$ [38]. This indicates that these computational investigations of CNT crosses, from the MD calculation of an equilibrium geometry to the choice of GAUSSIAN method to the definition of band gap energy, while not necessarily quantitatively accurate, are certainly producing very physically

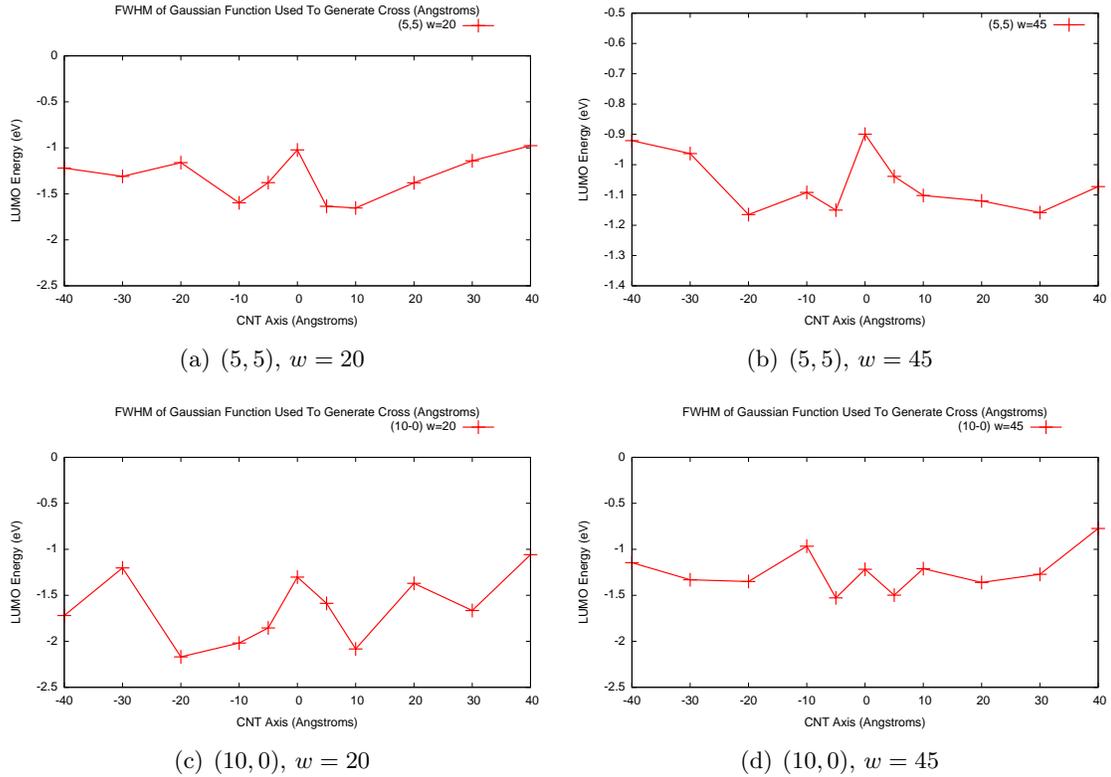


Figure 5.6: LUMO Energy of CNT Crosses. Notice, as in Fig. 5.5 that there is a one-to-one correlation between local deformation of the CNT and local band gap change, which supports the possibility of a QD at the CNT cross junction. The LUMO energy is interpreted as the ‘conduction band’ energy. Notice the local minimums, probably a result of the spatial deformation, in the LUMO energy, which could well be sufficient tunnel barriers to produce electron confinement.

reasonable results. Therefore the methods adopted here may be considered viable for future study.

In terms of visualization and physical interpretation,²⁰ it may be easier to see the regions of electron confinement from plots of the LUMO energy (E_L), shown in Fig. 5.6. Recall that the LUMO (Lowest Unoccupied Molecular Orbital) is the first electronic state that is not occupied in the system’s ground state. This leads to the natural interpretation that the LUMO, and all states above it, are the ‘conduction band’ and the HOMO, and all states below it, are the valence band²¹. Thus, heuristically, E_L gives a rough image of the ‘resistance to movement’ that an imaginary electron near the HOMO and LUMO states would ‘see’ as it is traveled down the axis of the CNT. This can be seen as follows. Suppose that an electron is in the HOMO of a slice somewhere along the CNT and is going to move to the left. Further suppose that, as in this slice, the adjacent slice, where the electron is going, is in its ground state, and so all electronic states

²⁰It should be noted that both Figs. 5.5 and 5.6 are derived from the same calculations. This section and these plots are included as they add a different perspective and interpretation of how the deformation of the CNT confines electrons

²¹Recall this motivated the definition of the band gap energy as $E_g = E_L - E_H$

up to and including the HOMO are occupied. In order for this electron to move to the adjacent slice, it must move to the LUMO of that slice, as this is the lowest available state. The lower the energy of this LUMO, the easier it will be for the electron to reach this LUMO, and so move to the adjacent slice. Extending this argument to large numbers of electrons, it follows that, statistically, for sections with larger values of E_L fewer electrons will be able to pass through these sections.

With this interpretation of the LUMO energy, it is almost self evident that the plots in Fig. 5.6 show intriguing behavior that appears to be local regions of confinement at the junction of the CNT cross, due to the local physical deformation of the CNT. For example, consider Fig. 5.6(d), and imagine an electron travelling from left to right in the conduction band (skipping from LUMO state to LUMO state). When it encounters the first rise in E_L at $x = -20\text{\AA}$, statistically it is less probable that it will continue in this direction, as it needs more energy to remain to the LUMO state ('conduction band'). Now consider an electron in one of the two minima at $x = \pm 5\text{\AA}$. If the height of the three triangular peaks (at $x = -10, 0, 10\text{\AA}$) are high enough, they may be considered as tunnel barriers, trapping this electron in a very small region of around 1 nm . This would be a Quantum Dot.

These minima in E_L are present in all plots in Fig. 5.6, and are all potential regions of confinement. However, it is impossible to know if the height of these tunnel barriers is sufficient to confine any appreciable number of electrons. To resolve this question, the transport properties of the CNT must be calculated.

Returning to whether a QD can be induced in a CNT cross formation, it must be said that while promising, these results do not definitively provide confirmation. However, they do provide a sketch of how a QD might arise in a CNT cross. Examining Fig. 5.5(b) once again, the region of the potential QD would essentially be the region over which local band gap changes occur: around 20\AA on either side, for a total width of 4 nm . The E_L minima in Fig. 5.6, which represent upward conduction band shifts, which can, if large enough, produce confinement. In particular the 'W' shape at the centre of Fig. 5.6(d), could well represent a QD with a width of around 1 nm . Thus, although I cannot conclusively infer the presence of a QD in these CNT crosses (which would require electronic transport calculations), the electronic structure change over extremely short ranges, on the order of nanometers, which is clearly caused by the deformation induced by CNT crosses, is certainly a positive indication that QDs may well occur in these structures.

Chapter 6

Conclusions and Future Work

Quantum Dots have very intriguing and promising properties resulting from the confinement of electrons in all three dimensions. One important aspect of QDs is they have discretized energy levels, the spacing of which is inversely proportional to the size of the QD. Therefore, in order to observe and exploit this discretization, very small QDs, on the order of nanometers are desired. Rather than attempt to use traditional microfabrication techniques at ever smaller, more challenging length scales, this work proposed to exploit the inherent small size of carbon nanotubes. As CNTs are already ‘one dimensional structures’, it is only needed to axially confine the electrons in the CNT, to achieve three dimensional confinement. A structure has been proposed, namely the carbon nanotube cross, which is expected to display excellent QD behavior at the junction of the CNT cross, and will have an extremely small characteristic size.

As may be evident from the phrasing of the above paragraph, I am not able to report a functioning QD in a CNT cross structure. However, much progress has been made, and the groundwork for further work has also been well established. An experimental methodology has been laid out, and this aspect of the project is currently a work in progress and is in the fabrication stages.

The CNT cross was also investigated computationally, using classical mechanics (Molecular Dynamics) to calculate the geometrical configuration of the CNT cross and quantum mechanics (Density Functional Theory) to calculate its electronic structure. It was found that there are many meta stable configurations of the CNT cross, corresponding to different types of CNT, as well as different manners in which the CNT cross was arranged. As each of these states showed stable equilibrium energy behavior, it was concluded that all of these states are physically realistic, and therefore worth investigating.

Using electronic structure calculations, I was able to demonstrate a one-to-one spatial correlation between local physical deformation, and local electronic structure change. The calculated band gaps are very similar to experimentally measured values, and the band gap changed significantly at the junction of the CNT cross, over a region as small as $1 - 2\text{ nm}$. Furthermore, significant conduction band shifts (based on the LUMO energy) were observed, over regions of similar sizes. Potentially these local electronic structure shifts could be adequate potential barriers to locally confine electrons, resulting in a QD.

The rich potential of the CNT cross that has been demonstrated in this work certainly justifies further research. I was not able to report any definitive results as to whether the type of CNT, or the width of the CNT cross, has any bearing on its electronic structure. To further investigate this I would propose to collect more data points along the CNT axis, gaining finer resolution, as well as run simulations of a wider variety of Gaussian profiles, and other types of CNTs. Furthermore, it would be worthwhile to investigate using some less computationally expensive

methods to calculate the electronic structure of the entire CNT²², examine the resultant local density of states, and extract the band structure of the whole system from this information. Finally, the definition of the band gap energy adopted in this work is fairly primitive, and is very sensitive to local errors or irregularities in the band structure, either physical or computational. It would be beneficial to develop a more physical and robust definition of the band gap energy. All of these suggestions are fairly straightforward, but would quite probably yield some very interesting results, and some of this work has already been started.

In order to truly validate the CNT cross as a QD device, two things would be required: simulating the electronic transport properties of the CNT cross, and growing and measuring CNT crosses. In both cases, staircase-like current-voltage curves would have to be observed to confirm that the CNT cross displays QD behavior. And while this work has not provided these definitive curves, it hopefully has provided enough evidence to suggest that it is probable that CNT crosses do exhibit QD behavior at their junction, and that further investigation of these structures, both experimental and computational, is merited.

²²This was tried using semi-empirical methods, but convergence could not be achieved. I have been working on modifying the simulation to remedy this.

Bibliography

- [1] Y. Alhassid. The statistical theory of quantum dots. *Rev. Mod. Phys.*, 72:895, 2000.
- [2] R.C. Ashoori. Electrons in artificial atoms. *Nature*, 379:413, 1996.
- [3] P. Avouris, J. Appenzeller, R. Martel, and S.J. Wind. Carbon nanotube electronics. *Proc. IEEE*, 104:1772, 2003.
- [4] P. B. Balbuena and J. M. Seminario. *Molecular Dynamics: From Classical to Quantum Methods*. Elsevier, Amsterdam, 1999.
- [5] J.V. Barth, G. Costantini, and K. Kern. Engineering atomic and molecular nanostructures at surfaces. *Nature*, 437:671, 2005.
- [6] P. Bhattacharya, S. Ghosh, and A.D. Stiff-Roberts. Quantum dot opto-electronic devices. *Rev. Mater. Res.*, 34:1, 2004.
- [7] D.W. Brenner. Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films. *Phys. Rev. B*, 42:9458, 1990.
- [8] T.J. Bukowski and J.H. Simmons. Quantum dot research: Current state and future prospects. *Crit. Rev. Sol. State Mat. Sci.*, 27, 2002.
- [9] V. Cerletti, W.A. Coish, O. Gywat, and D. Loss. Recipes for spin-based quantum computing. *Nanotechnology*, 16:R27, 2005.
- [10] M. Endo et. al. In A. Jurio, editor, *Carbon Nanotubes*, volume 111 of *Topics in Applied Physics*, pages 13–61. Springer-Verlag, 2008.
- [11] S. Franssila. *Introduction to Microfabrication*. Wiley, Sussex, 2004.
- [12] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136, 1964.
- [13] S. Iijima. *Nature*, 1991.
- [14] J.K. Jaiswal and S.M. Simon. Potentials and pitfalls of fluorescent quantum dots for biological imaging. *Trends in Cell Biology*, 14:497, 2004.
- [15] T. Jamieson, R. Bakhshi, D. Petrova, R. Pocock, M. Imani, and A.M. Seifalian. Biological applications of quantum dots. *Biomaterials*, 28:4717, 2007.
- [16] J.H. Jefferson and W. Hausler. *Mol. Phys. Rep.*, 17:81, 1997.
- [17] M.A. Kastner. Artificial atoms. *Physics Today*, 379:24, January 1993.

-
- [18] S. Katsumoto. Coherence and spin effects in quantum dots. *J. Phys.: Condens. Matter*, 19:233201, 2007.
- [19] C. Kittel. *Introduction to Solid State Physics*. Wiley, New York, 1996.
- [20] W. Kohn and L.J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140, 1965.
- [21] J. Kong, H. T. Soh, A. M. Cassell, C. F. Quate, and H. Dai. Synthesis of individual single-walled carbon nanotubes on patterned silicon wafers. *Nature*, 395:878, 1998.
- [22] L. Kouwenhoven and C. Marcus. Quantum dots. *Phys. World*, page 35, June 1998.
- [23] N.N. Ledentsov, V.M. Ustinov, V.A. Shchukin, P.S. Kop'ev, Zh.I. Alferov, and D. Bimberg. Quantum dot heterostructures: Fabrication, properties, lasers. *Semiconductors*, 32:343, 1998.
- [24] S. Lundqvist and N. H. March. *Theory of the Inhomogeneous Electron Gas*. Plenum Press, New York, 1983.
- [25] M.S.C. Mazzonni and H. Chacham. Bandgap closure of a flattened semiconductor carbon nanotube: A first principles study. *Appl. Phys. Lett.*, 87:173109, 2000.
- [26] E.D. Minot, Y. Yaish, V. Sazonova, and P.L. McEuen. Determination of electron orbital magnetic moments in carbon nanotubes. *Nature*, 428:526, 2004.
- [27] A. Nojeh. In M.M. Eshaghian-Wilner, editor, *Bio-inspired and Nano-scale Integrated Computing*. Wiley, New York, 2007.
- [28] A. Nojeh, G.W. Lakatos, S. Peng, K. Cho, and R.F.W. Pease. A carbon nanotube cross structure as a nanoscale quantum device. *Nano Lett.*, 3:1187, 2003.
- [29] T.W. Odom, J-L Huang, P. Kim, and C.M. Lieber. Structure and electronic properties of carbon nanotubes. *J. Phys. Chem. B*, 104:2794, 2000.
- [30] C-J. Park, Y-H. Kim, and K.J. Chang. Band-gap modification by radial deformation in carbon nanotubes. *Phys. Rev. B*, 60:10656, 1999.
- [31] N.G. Portney and M. Ozkan. Nano-oncology: Drug delivery, imaging, and sensing. *Anal. Bioanal. Chem.*, 384:620, 2006.
- [32] M. Pustilnik and L. Glazman. *J. Phys.: Condens. Matter*, 16:R513, 2004.
- [33] S. Reich, C. Tomsen, and I. Maultzsch. *Carbon Nanotubes*. Wiley, Weinheim, 2004.
- [34] B. Shan, G.W. Lakatos, S. Peng, and K. Cho. First-principles study of band gap change in deformed nanotubes. *Appl. Phys. Lett.*, 87:173109, 2005.
- [35] Susan Sinnott and Rodney Andrews. *Crit. Rev. Sol. State Mat. Sci.*, 26(7):145–249, 2000.
- [36] S.J. Stuart, A.B. Tutein, and J.A. Harrison. A reactive potential for hydrocarbons with intermolecular interactions. *J. Chem. Phys.*, 112:6472, 2000.

Bibliography

- [37] A. Szab and N. S. Ostlund. *Ostlund, Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover, New York, 1996.
- [38] J.W.G. Wildöer, A.G. Rinzier L.C. Venema, R.E. Smalley, and C. Dekker. *Nature*, 391:59, 1998.