Carbon Nanotubes for the Generation and Imaging of Interacting 1D States of Matter

Abstract

Low-dimensional systems in condensed matter physics exhibit a rich array of correlated electronic phases. One-dimensional systems stand out in this regard. Electrons cannot avoid each other in 1D, enhancing the effects of interactions. The resulting correlations leave distinct spatial imprints on the electronic density that can be imaged with scanning probes. Disorder, however, can destroy these delicate interacting states by breaking up the electron liquid into localized pieces. Thus, to generate fragile interacting quantum states, one requires an extremely clean system in which disorder does not overcome interactions, as well as a high degree of tunability to design potential landscapes. Furthermore, to directly measure the resulting spatial correlations, one requires an exceptionally sensitive scanning probe, but the most sensitive probes presently available are also invasive, perturbing the system and screening electron-electron interactions.

In this thesis, we show how carbon nanotubes allow us to create pristine 1D electronic systems with unparalleled tunability. By realizing a new approach to device fabrication based on deterministic nano-assembly, we create devices of high complexity and low disorder by selectively attaching nanotubes of chosen bandgaps and cleanliness onto devices with large numbers of local gates. Using precision nano-assembly, we also demonstrate devices with multiple nanotubes placed at predetermined locations. We then demonstrate the use of these devices as scanning charge detectors with the ability to image electrostatic potentials and to spatially resolve charging in a second nanotube device. By placing two such nanotube devices perpendicular to each other and bringing the two nanotubes into close proximity to each other, while distancing metal electrodes using our precise control over device geometry, we can use these devices as highly sensitive, high-resolution charge detectors that are also non-invasive. The capability to make tunable, pristine 1D electron systems and sensitive, non-invasive charge detectors could enable novel experiments with engineered interacting quantum states and direct access to their spatial correlations.
# Contents

1. **Introduction: Towards Condensed Matter by Design**
   1.1 A Way Forward in One Dimension .............................................. 5
   1.2 Carbon Nanotubes: A Promising 1D System .................................. 9
   1.3 Existing Local Probes and the Advantages of a Carbon Nanotube Charge Detector 14

2. **Existing Techniques to Create Quantum Devices with Carbon Nanotubes**
   2.1 Observing the Effects of Disorder ............................................. 21
   2.2 Fabrication Techniques .......................................................... 24
   2.3 The Need for a Deterministic Method ....................................... 29

3. **A New Generation of Clean and Complex Carbon Nanotube Devices**
   3.1 Overview of the Nano-Assembly Technique ................................ 31
   3.2 The Nanotube Chip ............................................................... 33
   3.3 The Circuit Chip ................................................................. 51
   3.4 Nano-Assembly with a Scanning Probe Microscope ....................... 59
   3.5 Resulting Devices ............................................................... 69

4. **Demonstration of Clean and Tunable 1D Electron Systems in Carbon Nanotubes**
   4.1 Localizing and Moving Electrons ............................................. 74
   4.2 Pristine Nanotubes At Dilution Temperatures ............................ 81
   4.3 Extracting and Explaining Electrostatics .................................... 84
   4.4 Quantifying Disorder ........................................................... 91
   4.5 Potential Design ............................................................... 96
   4.6 Double Dots ........................................................................... 100
   4.7 Novel Applications of Two-Nanotube Devices ............................... 103
   4.8 Conclusions ............................................................................ 106
5  Carbon Nanotubes as Highly-Sensitive and Non-Invasive Charge Detectors

  5.1  Operation of a Nanotube SET .................................................. 108
  5.2  Detection of Charging in a Nanotube Quantum Dot ....................... 112
  5.3  Spatially Resolving Charge in a Multi-Gated Double Dot ................. 114
  5.4  Mapping Gates and Enhancing Resolution with a Perpendicular Nanotube . 116
  5.5  Future Directions ................................................................. 118

Appendix A  Overview of the Electronic Properties of Carbon Nanotubes 120

  A.1  Tight-Binding on Graphene .................................................... 121
  A.2  Nanotubes: Quantized Graphene ............................................. 128
  A.3  Curvature .............................................................................. 136
  A.4  Curvature with Spin: Spin-Orbit Coupling ................................. 139
  A.5  External Perturbations: Mechanical Strain ................................. 143
  A.6  External Perturbations: Magnetic Fields .................................... 145
  A.7  Interactions ............................................................................ 147

Appendix B  Overview of Electrons in Quantum Dots 155

  B.1  Coulomb Blockade and Quantum Dots ....................................... 156
  B.2  Electrostatics of Quantum Dots with Multiple Gates ..................... 159
  B.3  Distinguishing Double Dots ....................................................... 167

Appendix C  Etch-First Circuit Chips ................................................. 173

References ....................................................................................... 196
Chapters 3 and 4 are largely composed of work published in\(^1\), and performed in collaboration with Maayan Honig, Sharon Pecker, Avishai Benyamini, and Assaf Hamo.

Chapter 5 is based on work done in collaboration with Sharon Pecker, Ilanit Shammass, Avishai Benyamini, and Assaf Hamo.
### Listing of figures

1.1 Quantum simulation of condensed matter systems ........................................... 2
1.2 Local probes for quantum simulators .............................................................. 3
1.3 Towards quantum simulation in condensed matter systems .............................. 4
1.4 The geometrical advantage of 1D systems ....................................................... 5
1.5 One-dimensional electron physics. ................................................................. 6
1.6 Examples of 1D systems .................................................................................... 8
1.7 The carbon nanotube: a promising 1D system. ................................................ 9
1.8 Nanotube structure yields robust 1D electronic structure and resistance to disorder 10
1.9 Examples of existing scanned probes ............................................................. 16
1.10 Our goal: a nanotube scanning a nanotube ................................................... 17

2.1 The wide range of device regimes available to carbon nanotubes .................. 20
2.2 Disorder in nanotube devices ...................................................................... 22
2.3 Top-contact fabrication technique. .............................................................. 24
2.4 Growth-last device fabrication. ................................................................. 26
2.5 Increasing the number of gates in growth-last devices with oxide-sealed gates. 27
2.6 Stamping nanotube devices. ......................................................................... 28

3.1 Overview of the nano-assembly technique ..................................................... 32
3.2 Flow-aligned CVD growth of nanotubes. ...................................................... 34
3.3 Slack in suspended nanotubes. ................................................................... 35
3.4 Tensioned and suspended individual carbon nanotubes .............................. 36
3.5 Suspended nanotubes in the sparse and dense regimes. ............................... 38
3.6 Short-range bundling on long suspended nanotubes. ................................. 39
3.7 Nanotube and circuit chip design. ............................................................... 40
3.8 Two-step etching of silicon trenches. ............................................................ 43
3.9 Effect of nanotube growth conditions on platinum surface morphology .... 44
3.10 Controlling catalyst deposition with photolithography. .............................. 45
3.11 Importance of nanotube chip cleanliness before growth. ........................... 47
3.12 Fabrication of the circuit chip ................................................................. 51
3.13 Design of the device electrodes. ............................................................... 53
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.14</td>
<td>Silicon spikes on circuit chips after plasma etching</td>
<td>58</td>
</tr>
<tr>
<td>3.15</td>
<td>A fabricated circuit chip</td>
<td>59</td>
</tr>
<tr>
<td>3.16</td>
<td>Detecting the nanotube chip edge, distance, and trenches with capacitance measurements</td>
<td>61</td>
</tr>
<tr>
<td>3.17</td>
<td>Extracting relative and absolute chip angles using capacitance measurements</td>
<td>63</td>
</tr>
<tr>
<td>3.18</td>
<td>The nano-assembly circuit</td>
<td>64</td>
</tr>
<tr>
<td>3.19</td>
<td>Establishing contact to a nanotube suspended between two contacts.</td>
<td>65</td>
</tr>
<tr>
<td>3.20</td>
<td>The selective nanotube cutting process</td>
<td>67</td>
</tr>
<tr>
<td>3.21</td>
<td>Representative nano-assembled devices</td>
<td>69</td>
</tr>
<tr>
<td>3.22</td>
<td>Robustness of the nano-assembly technique</td>
<td>70</td>
</tr>
<tr>
<td>4.1</td>
<td>Localizing and moving electrons on a five-gated, small-bandgap nanotube device</td>
<td>75</td>
</tr>
<tr>
<td>4.2</td>
<td>Symmetry of transport characteristics in a five-gated device</td>
<td>78</td>
</tr>
<tr>
<td>4.3</td>
<td>Symmetry of transport characteristics in a 16-gated device</td>
<td>79</td>
</tr>
<tr>
<td>4.4</td>
<td>Smoothly shifting electrons along a 10-gated device</td>
<td>80</td>
</tr>
<tr>
<td>4.5</td>
<td>Measurement of a second five-gated small-bandgap nanotube device measured at dilution refrigerator temperatures</td>
<td>82</td>
</tr>
<tr>
<td>4.6</td>
<td>Electrostatic environment of the nanotube in a five-gated device</td>
<td>86</td>
</tr>
<tr>
<td>4.7</td>
<td>Electrostatic origin of the observed features in the two-gate conductance scans</td>
<td>87</td>
</tr>
<tr>
<td>4.8</td>
<td>Calculated bare electrostatic potential profiles along the nanotube for three different gating configurations</td>
<td>92</td>
</tr>
<tr>
<td>4.9</td>
<td>Extracting an upper bound on the disorder potential from the measured transport</td>
<td>94</td>
</tr>
<tr>
<td>4.10</td>
<td>Recovering the gate resolution with potential design</td>
<td>97</td>
</tr>
<tr>
<td>4.11</td>
<td>Designer quantum dots at any position along a nanotube with a best-fit method</td>
<td>98</td>
</tr>
<tr>
<td>4.12</td>
<td>Double quantum dot on a 16-gate device</td>
<td>101</td>
</tr>
<tr>
<td>4.13</td>
<td>Eliminating cross-capacitance in a double quantum dot with the capacitance matrix</td>
<td>102</td>
</tr>
<tr>
<td>4.14</td>
<td>Electrical detection of motion of a nanotube mechanical resonator using a second nanotube quantum dot detector</td>
<td>104</td>
</tr>
<tr>
<td>5.1</td>
<td>Operating principle of the scanning SET</td>
<td>109</td>
</tr>
<tr>
<td>5.2</td>
<td>Detection of quantum dot charging with an SET</td>
<td>110</td>
</tr>
<tr>
<td>5.3</td>
<td>Charging of a nanotube quantum dot measured with an SET</td>
<td>113</td>
</tr>
<tr>
<td>5.4</td>
<td>Spatially-resolved charging in a double dot</td>
<td>115</td>
</tr>
<tr>
<td>5.5</td>
<td>Scanning SET imaging of device gates in parallel and perpendicular configurations</td>
<td>117</td>
</tr>
<tr>
<td>A.1</td>
<td>Graphene Structure and Spectrum</td>
<td>123</td>
</tr>
<tr>
<td>A.2</td>
<td>Sublattice phase and pseudospin in graphene</td>
<td>126</td>
</tr>
<tr>
<td>A.3</td>
<td>Structure and spectrum of the triangular lattice</td>
<td>127</td>
</tr>
<tr>
<td>A.4</td>
<td>Wrapping graphene and getting a nanotube</td>
<td>129</td>
</tr>
<tr>
<td>A.5</td>
<td>Periodic Table of Nanotubes</td>
<td>132</td>
</tr>
<tr>
<td>A.6</td>
<td>Isospin and the orbital magnetic moment of carbon nanotubes</td>
<td>133</td>
</tr>
<tr>
<td>A.7</td>
<td>Pseudospin and backscattering in nanotubes</td>
<td>135</td>
</tr>
<tr>
<td>A.8</td>
<td>Nanotube curvature-induced gaps</td>
<td>137</td>
</tr>
<tr>
<td>A.9</td>
<td>Nanotube curvature and spin results in spin-orbit coupling</td>
<td>141</td>
</tr>
<tr>
<td>A.10</td>
<td>Effect of Spin-Orbit Coupling on the Nanotube Spectrum</td>
<td>142</td>
</tr>
<tr>
<td>A.11</td>
<td>Nanotubes subjected to axial magnetic fields.</td>
<td>146</td>
</tr>
<tr>
<td>A.12</td>
<td>Interactions in nanotubes</td>
<td>147</td>
</tr>
<tr>
<td>A.13</td>
<td>Describing 1D particles with a smooth labelling field</td>
<td>150</td>
</tr>
<tr>
<td>A.14</td>
<td>Momentum-space scattering in 1D</td>
<td>152</td>
</tr>
<tr>
<td>B.1</td>
<td>Coulomb blockade of an isolated conductor.</td>
<td>157</td>
</tr>
<tr>
<td>B.2</td>
<td>Quantum dot energetics and charging.</td>
<td>160</td>
</tr>
<tr>
<td>B.3</td>
<td>Quantum dot electrostatics from conductance measurements</td>
<td>165</td>
</tr>
<tr>
<td>B.4</td>
<td>Double quantum dot electrostatics</td>
<td>168</td>
</tr>
<tr>
<td>B.5</td>
<td>Double dot with tunneling</td>
<td>171</td>
</tr>
<tr>
<td>C.1</td>
<td>Process flow for etch-first circuit chips</td>
<td>177</td>
</tr>
<tr>
<td>C.2</td>
<td>Plasma etching of etch-first circuit chips and artifacts.</td>
<td>178</td>
</tr>
</tbody>
</table>
This thesis is dedicated to the family of my past, for their struggle, to the family of my present, for their infinite love, and to the family of my future, to whom I wish to give those both.
Acknowledgments

Like an electron in the Fermi sea, no human is ever really an island, and interacts with the surrounding world. These interactions are complex, taking place between many people over much time. The humble electron becomes a new object, different and better for all of the interactions it undergoes, buoyed and guided by the sea.

To the chief guide of this work, Shahal Ilani, I owe a deep debt of gratitude. His unbounded enthusiasm for physics, his endless energies as a teacher, and his gnawing need to understand ever deeper and push ever farther have set a very high bar. The first day I set foot in his office, after a long and fitful night on a plane, we conversed for nearly eight hours, cracking some of the first ideas that would later enter into the mix of a years-long project. Pushing the boundaries of experimental physics became a daily experience, as we imagined and then realized ideas that always seemed beyond the horizon. For pushing me, and us, to become better physicists, better teachers, better writers and presenters, and even better Hebrew speakers, I give my thanks.

A special thanks goes to Amir Yacoby, who humored an adventurous young graduate student and introduced me to Shahal.

To the group, a big thank you. It has been a privilege to work with a team of such creative individuals, who have entertained countless discussions from the deepest issues of physics to the minutiae of nano-fabrication, the nature of nanotubes and the intricacies of circuits, the sounds of music and the rhythm of Israeli culture, the plot twists of Jewish history and the nuances of Jewish law, and more. But it was not all talk; they lent a hand whenever asked, helped explain when the Hebrew was too complex, and helped understand when the culture shock was too great. Not to mention bossa jams, rock climbing trips, and movie nights. In the order I made their acquaintance, they are Maayan Honig, Sharon Pecker, Avishai Benyamini, Arjun Joshua, Assaf Hamo, Joey Sulpizio, Ilanit Shammass, Ilya Khrivich, and Lior Ella.

A thank you to the staff at the Weizmann Institute, for all of their assistance, insight, and professionalism. To the staff at Harvard, a thank you for patience and assistance even when I was oceans away.

Beyond the Institute walls, all the friends I have met throughout a long and winding journey, in Israel and at Harvard, who have sprinkled these years with adventures, laughs, good counsel and open ears, deserve thanks. To Bud and Dani, dear friends, tolerating the long disappearances of a graduate student and always feeling like family. With them I learned the unique release of cycling at
3 in the morning in Tel Aviv, possibly inebriated. To Susanne Pielawa, for her help and friendship in inumerous ways. To the apartment mates of Elkalai, Yuval Shagam and Hiro Inoue, it has been a long, long ride, and the discussions, movies, dinners and beers made it all better. And many more: Yechiel Hershenson, Tarek Anous, Dionysios Anninos, Mikhail Kats, Nick Litombe, Ferdinand Kuemmeth. I could go on, but I have to hand this in eventually.

To my family, words cannot express my gratitude, and far less these short ones here. Thank you, and much, much more.

To Alana, for standing by me, cheering and encouraging and counseling, and for holding my hand as we face a new horizon.
The study of many-body quantum physics is historically centered around condensed matter. Solid-state materials naturally possess a large number of interacting electrons. Combined with Fermi statistics, spin, and the lattice degrees of freedom, a wealth of interacting electron phases have been predicted and discovered in natural materials.
The use of materials for the study of interacting physics, however, places limitations on what states can be realized. Critical parameters, such as interaction strength and band filling, can only be slightly changed for each sample. Realizing different material Hamiltonians is thus limited by materials chemistry. The imperfections intrinsic to real materials further obscures the physics, by introducing random and uncontrollable disorder that can destroy or alter interesting electron phases.

To overcome these challenges, a different approach to the study of interacting many-body physics recently took root. Instead of hunting for imperfect materials, one can study many-body physics by design with a quantum simulation. Techniques for quantum simulation of condensed mat-
Figure 1.2: Local probes for quantum simulators. a) Laser probing atoms in an optical lattice. Adapted from\textsuperscript{9}. b) Quantum gas microscope for trapped atoms. Adapted from\textsuperscript{12}. c) Scanning transmon qubit. Adapted from\textsuperscript{10}.

ter have grown in number, and include optical lattices of trapped atoms\textsuperscript{6} (Figure 1.1a), ion traps\textsuperscript{3} (Figure 1.1b), photonic simulators\textsuperscript{4} (Figure 1.1c), and superconducting circuits\textsuperscript{5} (Figure 1.1d). These systems possess low disorder by design, and in some cases allow for the \textit{in situ} tuning of parameters such as interaction strength, filling, and sometimes even the lattice geometry itself\textsuperscript{7}.

Tools for obtaining spatial correlations have also been developed for quantum simulators. These include direct, \textit{in situ} imaging of the correlated particles\textsuperscript{8} (Figure 1.2b), and local probes for scanned imaging\textsuperscript{9,10} (Figure 1.2a,c). Techniques relying on time-dependent potentials to obtain correlations have also been realized\textsuperscript{11}. The combination of clean, tunable systems with local probes provides a powerful tool to study many-body systems.

Current quantum simulation techniques typically involve bosonic constituents with weak interactions. The absence of Fermi statistics and strong interactions precludes the study of some of the most interesting interacting phenomena in condensed matter. Experimental efforts to realize fermionic quantum simulators is ongoing, but this remains a challenging goal\textsuperscript{13}.

Ideally, one desires to fuse the flexibility and tunability of quantum simulators with the fermions and strong interactions of condensed matter. To do this, one may exploit low dimensional systems. For instance, in two dimensions, a large experimental effort has been made to realize clean 2D electron systems, such as in semiconductor heterostructures and graphene (Figure 1.3a,b). Theoretical
Figure 1.3: Towards quantum simulation in condensed matter systems. a) 2D electron systems in semiconductor heterostructures have been highly refined, but remain beholden to disorder. b) Graphene has rapidly approached the cleanliness of heterostructures, but is vulnerable to external sources of disorder, such as substrates and adsorbed atoms and molecules. c) Proposal for simulating high-$T_c$ materials with quantum dots using a large metallic gate above a 2D electron system. Adapted from 14. d) Creating a hexagonal superlattice in a 2D electron system with deposited metallic structures. Adapted from 16.

proposals to simulate specific problems, such as that of high-$T_c$ superconductivity, were put forward 14,15 (Figure 1.3c).

The challenge of realizing a 2D electron system with the versatility of a quantum simulator is twofold. On the one hand, it has yet to be demonstrated how the potential can be locally controlled across a 2D electron system. One recent approach relied on choosing predetermined potential landscapes and permanently incorporating them into a given sample without the possibility of tunable, local control 16 (Figure 1.3d).

More fundamentally, disorder remains an outstanding issue. In semiconductor heterostructures, dopants, while providing the 2D electron system, also inevitably impact the cleanliness of the system, with observable effects 17. In 2D crystals such graphene, the electron gas is exposed to the surrounding environment, which can mean a disordered substrate or adsorbed atoms and molecules.
which serve as scatterers for electrons.

1.1 A Way Forward in One Dimension

![Figure 1.4: The geometric advantage of 1D systems. A 1D system (lower panel, black) holds interacting electrons (red). Perpendicular electrodes (blue) allow the straightforward creation of electrostatic potentials with wells and barriers in tunable shapes (grey, upper and middle panels).](image)

A promising alternative to achieve the capabilities of quantum simulation in a condensed matter context are one-dimensional systems. 1D systems have a unique geometric advantage that allows the use of metallic electrodes to control the 1D potential. They can be placed at right angles to the system under study, as illustrated in Figure 1.4. This simplicity allows the straightforward realization of experimentally relevant potentials, such as potential wells and barriers, periodic potentials, and even the controlled introduction of disorder.

The geometry of 1D systems leads to a second important property, a degree of robustness against disorder. The second dimension of the electronic wavefunction is typically quantized into subbands. The resulting subband separation in momentum space helps prevent scattering by long-range disorder potentials.

These features are appealing for realizing a condensed matter quantum simulator, but what kind of physics can we study in 1D? Perhaps counterintuitively, 1D systems, rather than being simple
in comparison to their higher-dimensional siblings, in fact host a large repertoire of new physical phenomena. These surprisingly exotic properties, discussed below, are challenging to observe, and make the realization of a clean and tunable 1D system an enticing goal.

1.1.1 The Rich Physics of 1D Systems

One dimensional systems are distinctly different from their higher-dimensional cousins. Particles in 1D cannot avoid each other by moving in orthogonal directions, as is illustrated in Figure 1.5a. As a result, even the weakest interactions cause an electron in a 1D system to affect all other electrons. Local excitations of the system therefore propagate throughout the entire system, and well-defined fermionic quasiparticles do not exist. The resulting physics is no longer described by Landau’s Fermi
liquid theory, as in higher dimensions. Instead, the appropriate description is the Luttinger liquid theory, in which excitations in 1D are described by collective bosonic modes\textsuperscript{20,22}.

The Luttinger liquid framework predicts several unconventional electronic phenomena. Electrons injected into a Luttinger liquid break up into spin and charge excitation with different velocities\textsuperscript{20,24,25} (Figure 1.5b). The tunneling electron charge can also fractionalize\textsuperscript{26–29} (Figure 1.5c). Tunneling across a barrier in the middle of a Luttinger liquid results in suppressed conductance at low energies due to interaction of the injected electron with the many-body 1D state\textsuperscript{30,31}. With strong, long-range interactions, electrons crystallize into the Wigner crystal state, exhibiting a charge modulation periodic with the inter-electron spacing\textsuperscript{25,32–34} (Figure 1.5d). Under a periodic potential, a Mott insulating phase can arise if the period of the potential is commensurate with the inter-electron spacing. Electrons then lie in the potential minima, generating gaps in the spectrum and spin-wave-like degrees of freedom\textsuperscript{35–37}.

1.1.2 Realizations of 1D Systems

The wide range of exotic electronic behavior in 1D has driven the experimental realization of 1D systems with the goal of observing these 1D phenomena. Among the earliest realizations were bulk salts composed of coupled 1D chains\textsuperscript{38} (Figure 1.6a). These were used to measure bulk transport properties which revealed signatures of 1D behavior. Such signs were observed also in atomic chains self-assembled on the surfaces of crystals\textsuperscript{39} (Figure 1.6b). The individual 1D systems are difficult to access individually in these systems, and therefore they do not seem to be able to realize a tunable, locally-addressable 1D environment for electrons.

Another possibility is semiconducting nanowires, which have been extensively developed and used to form zero dimensional systems like single and double quantum dots\textsuperscript{42–45} (Figure 1.6d). Nanowires are typically limited to short lengths, which precludes the creation of extended 1D systems. They also typically possess relatively large diameters of \( \sim 100\, \text{nm} \), which indicates the presence of
Figure 1.6: Examples of 1D systems. a) Bechgaard salts, a bulk quasi-1D material. b) Atomic chains on a crystal surface. Adapted from 39. c) Cleaved-edge overgrowth quantum wires in semiconductor heterostructures. The 1D system is the lower blue wire. Adapted from 40. d) Epitaxial semiconductor nanowires. Adapted from 41.

closely-spaced subbands that can complicate electronic behavior and increase sensitivity to disorder. Furthermore, since they are typically formed through epitaxial growth, short-range lattice disorder and surface defects can scatter electrons, ultimately breaking the electron system into chains of zero-dimensional puddles.

An alternative strategy is to use an interface 2D electron system and deplete it to form a 1D system. This has been done in materials systems such as graphene 46,47, LAO/STO 48, quantum Hall edge states 29, and GaAs-based cleaved-edge overgrowth quantum wires 25,28,49–52 (Figure 1.6c). Cleaved-edge wires have been extensively used to investigate 1D physics, enabling the observation
of spin-charge separation and charge fractionalization. Due to the epitaxial growth of the host semiconductor heterostructure, which involves placing dopant atoms in the host crystal, disorder remains present, and can affect conductance quantization. The buried nature of the 1D system further puts a bound on the ability of local probes to access the system.

We seek a 1D system which is clean, possesses strong interactions, and is accessible by a local probe. We now show that these conditions are met by the carbon nanotube.

1.2 Carbon Nanotubes: A Promising 1D System

![Figure 1.7: The carbon nanotube.](image)

Since their discovery, carbon nanotubes have held promise as a unique nanomaterial. They can be thought of as a rolled up 2D sheet of carbon atoms in a hexagonal lattice, called graphene, illustrated in Figure 1.7a. Carbon nanotubes inherit the properties of graphene, including exceptionally high chemical stability, mechanical strength and electrical conductivity. Their diameters can reach less than 1nm, while their lengths can grow to be up to cm’s. These extreme aspect ratios bode well for long devices with multiple local, perpendicular electrodes, as shown schematically in Figure 1.7b.
Nanotubes possess a host of properties that make them highly appealing for realizing a clean, interacting 1D electron system. Their large subband spacing ensures robust 1D behavior, they are uniquely resistant to disorder, and they possess strong Coulomb interactions. These unusual properties can be easily motivated by use of a tight-binding model based on graphene. Graphene is a hexagonal lattice of carbon atoms with electronic properties that can be described with nearest-
neighbor hopping between atomic orbitals. This model is explored in more detail in appendix A, but even the simplest approach, described below, yields important conclusions.

The graphene lattice is hexagonal, with two atoms in the unit cell defining two triangular sublattices (Figure 1.8a). The spectrum therefore has two bands that originate in the two atoms of the unit cell, and touch at the vertices of the hexagonal Brillouin zone at the points $K$ and $K'$. The Fermi level is located at these points, which at low energies can be expanded to yield linear dispersions given by

$$E = \pm \hbar v_F|q|$$  \hspace{1cm} (1.1)

with $v_F \approx 10^6 m/s$ the Fermi velocity, $q$ the momentum wavevector with respect to the $K/K'$ points, and the two signs correspond to the two bands (Figure 1.8b). This linear dispersion describes two Dirac cones, centered about the $K/K'$ points, which govern the low-energy properties of graphene. The wavefunction at the Fermi level can be written as

$$\Psi \propto \begin{pmatrix} 1 \\ e^{i\phi} \end{pmatrix}$$  \hspace{1cm} (1.2)

where the two components correspond to the weight of the wavefunction on each of the two sublattices, and $\phi$ is the phase between the two components$^{11}$. This degree of freedom is known as the pseudospin. The pseudospin $\phi$ is a function of $q$ and can be calculated from the tight-binding model, with the result that $\phi$ revolves from 0 to $2\pi$ around the $K$ point, and winds in the opposite direction around the $K'$ point (Figure 1.8c). The pseudospin describes the relative phase between the two sublattices, while the wavefunction on an individual sublattice is found by solving the problem on one triangular sublattice. This reveals that the $K/K'$-point wavefunctions are anti-bonding states with threefold, $120^\circ$ rotation symmetry, so that the phase of individual sites in the sublattice changes by $2\pi/3$ from site to site, shown in Figure 1.8a with red, blue, and green colors (see section A.1 for
more details).

The carbon nanotube is constructed by imposing periodic boundary conditions on the graphene lattice. A vector \( \mathbf{C} \) is chosen to wrap graphene from one sublattice atom to another atom on the same sublattice, defining the nanotube circumference (Figure 1.8a). The periodic boundary condition is given by

\[
\exp ik_{\perp} \cdot \mathbf{C} = 1
\]

where \( k_{\perp} \) is the discrete momentum wavevector along the circumference of the tube, while a continuous wavevector, \( k_{||} \), points along the nanotube length (Figure 1.8b). The quantization condition slices the conical graphene spectrum into subbands spaced by \( 2\pi/|\mathbf{C}| \). If we wrap graphene by drawing \( \mathbf{C} \) between two atoms with the same phase in the anti-bonding \( K/K' \) wavefunction, then we have \( k_{\perp} \cdot \mathbf{C} = K \cdot \mathbf{C} \) and the slicing crosses through the Dirac point. In this case, the lowest subband has a linear dispersion and zero bandgap, making the nanotube metallic. Conversely, if we wrap with two atoms having different phase, then the slicing misses the Dirac point and the lowest subband has a bandgap. By choosing an atom with the extra \( 2\pi/3 \) phase shift in the \( K/K' \) anti-bonding state, the quantization condition needs this extra phase to complete the wrapping. This translates into a displacement of the lowest subband from the Dirac point, since

\[
k_{\perp} \cdot \mathbf{C} = K \cdot \mathbf{C} + \delta \phi = (K + \delta k) \cdot \mathbf{C},
\]

where the shift \( |\delta k| = 2\pi/3|\mathbf{C}| = 2/3d \), with \( d \) the nanotube diameter. Together with the linear dispersion, this gives the size of the bandgap

\[
E_g = 2\hbar v_F \times \frac{2}{3d} \approx \frac{0.8}{d[nm]} eV \approx 9300 \frac{d[nm]}{d[nm]} K
\]

We learn that even far above room temperature, only one subband remains energetically accessible for typical nanotube diameters of \( d \sim 1\text{nm} \). Nanotubes are thus robustly 1D with large subband spacing (see section A.2).

Wrapping graphene quantizes its energy spectrum, but also quantizes the pseudospin \( \phi \), with
important consequences on the effects of disorder. Each subband of a nanotube has an associated slice of the pseudospin spectrum. For a metallic nanotube, the lowest subband slices through the Dirac points, and as a result, each side of the linear dispersion has a different pseudospin: $\phi = 0$ for left-moving electrons and $\phi = \pi$ for right-moving electrons (we consider an armchair nanotube; see Figure 1.8d). The consequences of this fact are striking. A back-scattering process between left- and right-movers in the sublattice basis can be written as

$$\begin{pmatrix} 1 & -1 \\ S_{AA} & S_{AB} \\ S_{BA} & S_{BB} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = (S_{AA} - S_{BB}) + (S_{AB} - S_{BA})$$

(1.5)

where $S_{ij}$ are the elements of the scattering operator in the sublattice basis. Remarkably, the scattering expectation value vanishes for any scatterer that preserves sublattice symmetry. Breaking sublattice symmetry demands a potential variation on the scale of the carbon-carbon bond, effectively protecting carbon nanotubes from all but the shortest length-scale disorder. Semiconducting tubes are less protected, since they have some pseudospin overlap between left- and right-movers, but the strength of the scattering is still partially suppressed, and suppressed more strongly at higher energies$^{16,57}$.

Tight-binding shows that nanotubes are 1D and robust to disorder, but to understand the role of interactions we must go beyond this simple model. To get an intuitive understanding, consider fermions populating a 1D band, for which the 1D density $n$ is related to the Fermi wavevector $k_F$ by $n = \tau k_F / \pi$, where $\tau$ is the state degeneracy. The Coulomb interaction energy goes as $U_{int} \sim \frac{\epsilon}{\tau}$, with $\epsilon$ the dielectric constant and $r$ a typical interparticle spacing, which in terms of the density is $U_{int} \sim \frac{e^2}{\epsilon}$. For the linear band of a metallic carbon nanotube, the kinetic energy is $U_{kin} \sim \hbar v_F k$, and thus the ratio of interaction to kinetic energies is $U_{int} / U_{kin} = \tau \epsilon^2 / \pi \epsilon \hbar v_F \approx 3$, a constant independent of $n$ (we have assumed $\epsilon / \tau \approx \epsilon_0$). This is in marked contrast to free particles in a
parabolic band, for which the interaction vanishes at high energies and becomes important only near the band-bottom. Thus interactions in nanotubes are not only strong, but non-negligible at all energies. This conclusion holds also for semiconducting nanotubes, for which the dispersion is linear at high energies and has a parabolic band-bottom where interactions become even more dominant (see section A.7).

Detailed theoretical treatments predict that interactions have a decisive effect on determining electronic properties of carbon nanotubes. Short-range interactions are predicted to lead to Mott gaps. Strong Coulomb interactions are predicted to lead to Luttinger liquid behavior with a small Luttinger parameter \( g \approx 0.2 \) and distinct spin-charge separation effects, as well as modified tunneling exponents and spin correlations. Long-range Coulomb interactions may also lead to the formation of Wigner molecules via Wigner crystallization.

We thus expect that nanotubes are host to a variety of correlated, strongly-interacting electron phases, but we need a unique probe to identify these quantum states. In the next section, we examine probes that can confirm their existence by imaging them in real space.

1.3 Existing Local Probes and the Advantages of a Carbon Nanotube Charge Detector

How can one identify strongly-interacting 1D states of matter? Experiments have attempted to identify their existence with transport measurements, which measure the bulk transport of electrons through the 1D system. Transport has been used to study 1D behavior by measuring interaction-dependent power law behavior of the conductance, as predicted by Luttinger liquid theory. Transport spectroscopy of 1D electrons confined in carbon nanotube quantum dots was used to identify signatures of Wigner molecules and the Mott insulating state. Signatures of interactions in nanotubes have also been observed with shot-noise measurements and measurement...
of the non-equilibrium distribution function. Other techniques have been used in different material systems. For instance, to identify 1D spin-charge separation, momentum-resolved tunneling was measured between two parallel cleaved-edge overgrowth quantum wires. Charge fractionalization in semiconductor heterostructure-based wires was observed both with momentum-resolved tunneling and time-resolved measurements with fast voltage pulses.

Transport experiments are limited in that they measure a bulk quantity, and cannot resolve spatial structure. For more complex interacting states in an extended 1D geometry, the ability to map spatial structure would greatly enhance our ability to distinguish different correlated states. To achieve this, we can perform imaging with a scanning local probe.

Many local probes exist and have been used to study nanotubes. One such probe is the scanning tunneling microscope (STM), in which a metallic tip is brought in close contact to a conducting surface and the tunneling current between tip and sample is measured, giving information on the local tunneling density of states. The requirement of placing nanotubes on a conducting surface to perform this measurement, however, screens electron-electron interactions, resulting in single-particle behavior. By suspending the nanotube, screening due to the substrate can be avoided, but the large metallic tip itself screens interactions. Furthermore, the tip must be voltage biased with respect to the sample to fix the energy of tunneling electrons, implying that the work function difference between tip and nanotube cannot be independently nulled. This results in a position-dependent gating effect of the tip that cannot be corrected, and thus STM is a highly invasive measurement.

To allow nulling the potential difference between tip and sample, one can use a tip that does not require a bias voltage. This can be done with charge detection, in which the electrostatic potential of the sample is mapped. The Kelvin probe force microscope (KPFM) allows for this type of probing in an atomic force microscope (AFM) setup, by measuring local work function differences. However, this technique is not sufficiently sensitive, and uses a large conducting AFM tip which screens
electron-electron interactions. They are also invasive: AFM tips have been shown to gate scanned nanotube samples\cite{78,79}.

A more promising charge detection technique is the scanning single-electron transistor, or scanning SET\cite{75,80}. This measures electrostatic potential by monitoring current through the SET, which depends sensitively on local electric fields (see section 5.1). It has been successfully used to image edge states and localized excitations in the quantum Hall effect, and electron-hole puddles in graphene\cite{80-83}. Existing scanning SETs are formed by evaporating metallic islands onto pulled glass fiber tips, resulting in a typical SET size of $\sim 100\text{nm}$. Such a large metallic island in close proximity to a nanotube
screens electron-electron interactions, and can induce a local gating effect if work function differences are not properly nulled.

Recently, our group demonstrated a scanning SET defined in a suspended nanotube. The scanning nanotube SET was successfully used to image domain walls in LAO/STO samples. A scanning nanotube SET has distinct advantages for the imaging of correlated 1D states in carbon nanotubes. The SET can be scanned perpendicular to the nanotube under study, maximizing spatial resolution, as illustrated in Figure 1.10. Even more importantly, its nanometer diameter means that the SET will couple minimally to the nanotube under study, reducing the effects of screening and preserving strongly-interacting states. By suspending the nanotube SET far above its gates, and distancing the source and drain contacts from each other, the two nanotubes can be placed closer to each other than to any nearby metal, reducing the screening effects of the metallic electrodes. Furthermore, because the 1D electron system resides on the surfaces of the nanotubes, they can be brought in close proximity for exquisitely sensitive and local measurement, unlike buried systems.

![Figure 1.10](image)

**Figure 1.10:** Our goal: a pristine nanotube in a tunable potential, scanned by a second, perpendicular nanotube used as a high-resolution, non-invasive charge detector.

The geometric requirements for a non-invasive nanotube SET are thus similar to those for using nanotubes as a designer 1D electron system: a tunable device geometry that allows us to distance metallic electrodes to avoid screening and maintain a clean, interacting 1D system. In the next chapter...
ter, we review techniques for creating nanotube devices and identify the key challenges to realizing an ultra-clean and tunable 1D system for local probing.
Existing Techniques to Create Quantum Devices with Carbon Nanotubes

The promise of carbon nanotubes for 1D physics has led to many experiments harnessing nanotubes to probe their behavior. Signatures of strongly-interacting electrons have been observed, the most recent being the observation of a two-electron Wigner molecule \cite{31,33,34,37}. Exotic correlated states have been generated and probed, most notably the Kondo effect \cite{84-86}. Nanotubes have also been used
Figure 2.1: The wide range of device regimes available to carbon nanotubes. a) Conductance measured through several top-contacted nanotube devices, as a function of gate and bias voltages at a temperature of 1.5K. Top right inset shows a schematic of the nanotube device. Right insets: conductance averaged over gate voltage for the corresponding device. b) Schematics of two physical regimes of device behavior. Top: Coulomb blockade. Strong barriers cause Coulomb repulsion to prevent transport through the device at specific gate and bias voltages, useful for localizing electrons and spins (see main text and B). Bottom: Fabry-Perot regime. Electron waves traverse the 1D nanotube waveguide, and undergo destructive interference at specific gate and bias voltages. This allows the observation of open-system phenomena like the Kondo effect. Adapted from 106,107.

to localize and control individual electron spins, and could be used to realize a 1D quantum information chain 87–93. The coupling of electrons to nanotube mechanical motion has been extensively explored 94–100, as well as their interaction with correlated materials such as ferromagnets and superconductors 101–105.

The wide variety of experiments on nanotubes is driven by the range of physics to which they allow access. In section 1.2, we saw that electrons live on the surface of carbon nanotubes, allowing one direct access to the 1D electron system. The possibility of contacting a nanotube with different contact resistances then allows the realization of different physical regimes. Figure 2.1 shows how the low-temperature transport properties of a nanotube contacted by two metal leads allow one to realize different electronic behaviors. In the low resistance regime, where the contact resistance approaches the quantum limit, transport at low temperatures is described by the motion of waves.
traversing a one dimensional waveguide with two barriers of finite transmission $^{108,109}$. If the system is phase coherent, then interference effects will dominate the transport characteristics in this Fabry-Perot regime. Such an open-system regime allows one to realize correlations between electrons on the nanotube and in its metallic leads, leading to Kondo effects and enabling the use of cotunneling spectroscopy $^{84-86,110,111}$. On the other hand, in the high resistance regime, electrons spend enough time on the nanotube to interact, inducing Coulomb blockade and causing charging effects to dominate (see appendix B). This allows electrons to be localized, which is useful for transport spectroscopy and spin manipulation for quantum information science $^{34,88,89,97,112}$.

These properties show that carbon nanotubes are useful in a wide range of experiments, but experiments so far have mainly exploited nanotubes in the setting of zero-dimensional single or double quantum dots. What hinders the extension to longer one-dimensional settings? Below, we will see that the answer is disorder.

### 2.1 Observing the Effects of Disorder

Disorder is the bane of quantum devices, breaking up the electron liquid into disconnected puddles and obscuring their delicate interacting states. Disorder of the electrostatic potential can arise from imperfections of the underlying crystal lattice, the electrostatic environment, or of adsorbed contaminants. Historically, for many low-dimensional systems, quantum phenomena became observable only after the development of highly refined and clean versions of the material under study, for example as was the case for many discoveries relating to the quantum Hall effect in semiconductor heterostructures and graphene $^{113-115}$.

As we saw in section 1.2, some nanotubes are expected to be resistant to back-scattering thanks to pseudospin symmetry. The effect of disorder on different nanotubes can be observed in transport experiments, as in the example shown in Figure 2.2 $^{56}$. Here, in one device, a metallic nanotube
Figure 2.2: Metallic or small-bandgap nanotubes are less sensitive to disorder arising from substrates. a) Conductance, G, as a function of backgate voltage $V_g$. Upper left inset: the measured device, a metallic nanotube resting on two metallic terminals. At low temperatures, the conductance shows oscillations corresponding to the entire nanotube length of $8 \mu m$. Adapted from 56. b) Conductance, G, as a function of backgate voltage $V_g$ on a semiconducting nanotube in the same geometry. At low temperatures, the conductance exhibits peaks corresponding to a typical quantum dot size of $\sim 100 nm$. Semiconducting nanotubes are thus more susceptible to disorder. Adapted from 56. c) Left panel: a voltage-biased AFM is scanned across a nanotube resting on an insulating substrate and contacted by two metallic leads. Right panel: the conductance, G, is measured as a function of tip position, and shows rings centered at different positions along the nanotube. These correspond to charging of quantum dots localized along the disordered nanotube. Adapted from 79.
sits on top of two metallic contacts on a silicon substrate. At low temperatures, the conductance as a function of gate voltage shows Coulomb oscillations (see section B.2). The oscillations have a consistent periodicity, and therefore correspond to a single length scale. Here, that scale is 8 μm, the length of the entire nanotube extending beyond the underlying metal contacts. Thus, electrons in this quantum dot are in fact delocalized over the entire nanotube.

In another device made of a semiconducting nanotube, the behavior is strikingly different (Figure 2.2b). At low temperatures, the conductance as a function of gate voltage shows a series of peaks with different spacing and heights. The typical spacing of the peaks corresponds to an effective dot size of 100nm. In effect, disorder has broken up electrons on the nanotube into a series of low-conducting 'quantum dots'. The semiconducting nature of this nanotube renders it dramatically more susceptible to disorder in the electrostatic environment. For the creation of clean 1D systems, we want to avoid these nanotubes.

What is the origin of disorder that breaks up the 1D electron liquid? The spontaneous formation of quantum dots along nanotubes lying on a substrate was investigated in scanning experiments. By using a voltage-biased AFM tip, spontaneous quantum dots are charged and discharged as the nanotube is scanned, changing its conductance. As shown in Figure 2.2, the charging peaks show up as rings centered around the disorder-induced quantum dot. Potential modulations of the substrate, in addition to intrinsic disorder on the nanotube itself, thus induce potential barriers and wells along the tube that can break up the electron liquid.

These results show that nanotubes with small bandgaps can be robust to disorder, and that disorder arises from the electrostatic environment of the nanotube. Thus, by choosing the right nanotube and maintaining a clean environment, we can hope to realize a clean 1D electron system. Fabrication techniques must therefore fulfill two conditions: the deterministic choice of nanotubes and the creation of a clean environment. Below, we will see how different device fabrication techniques measure up against these requirements.
2.2 Fabrication Techniques

In this section, we review techniques used to fabricate quantum devices with nanotubes, with an eye towards understanding how they enable different experiments and how they give access to the clean 1D limit.

We assume here the use of chemical-vapor deposition (CVD)-grown nanotubes. This process will be described in detail in chapter 3. In brief, it involves the deposition of a catalyst and a high-temperature growth process which occurs under a gas flow, resulting in flow-aligned nanotube growth.

2.2.1 Top-contact

![Figure 2.3: Top-contact fabrication technique. a) Schematic of the method. Left panel: nanotubes are grown or deposited on an insulating substrate. Right panel: contacts are deposited on the nanotubes. If desired, an insulating layer is deposited and further gates patterned on top. Adapted from 116. b) A top-contact nanotube device imaged with an AFM and false-colored. Purple: insulating substrate. Yellow: metal contacts. Red: nanotube. Adapted from 106. c) A complex top-contact nanotube device imaged in SEM and false-colored. Red, numbered 1-4: metal contacts. All other lines: gates deposited after insulating oxide deposition. Adapted from 88.](image)

In this technique, nanotubes are grown or deposited on a substrate. Their locations are identified, typically with SEM or AFM, and contacts are designed and deposited for selected nanotubes,
as illustrated in Figure 2.3. It thus allows one to preselect nanotubes with AFM measurement of the nanotube height, giving a small degree of selectivity. Furthermore, top-contacting can give very low contact resistances, close to the quantum limit. The advantage of good contact has rendered this technique useful for studying processes like co-tunneling, which rely on open contacts and allow the observation of fine level structure in electron excitations. The deterministic nature of this fabrication technique also allows the creation of sophisticated device geometries, useful for quantum control.

The drawback of top-contacting is the fact that nanotubes sit directly on the substrate, exposing them to electrostatic disorder in the substrate and charge fluctuations from charge traps. More complex issues, like mechanical deformations of the nanotube due to forces acting on it from the deposited gates may also affect its behavior. The necessity of several chemical processing steps with the nanotube present can worsen disorder and chemically damage nanotubes or leave residues. These issues rule out the possibility of using this technique as a robust method to obtain long, clean nanotubes.

2.2.2 Growth-Last

To avoid damaging nanotubes with chemical post-processing, nanotubes can instead be grown on top of pre-fabricated device geometries. This is possible by defining selected regions to place growth catalyst with lithography (see section 3.2). Nanotube growth is then performed as the last stage in the fabrication process, preserving the pristine nature of the nanotubes. This method has been shown to give extremely clean devices and allowed the observation of spin-orbit coupling and Klein tunneling in nanotubes.

The need for the device to undergo CVD growth at high temperature puts strict limits on the materials and designs used. As a result, it has been challenging to implement increasing numbers of local gates with this technique, requiring clever solutions like those shown in Figure 2.4.
Figure 2.4: Growth-last device fabrication. a) Suspended nanotube grown over a silicon trench and Pt contacts. Top panel: schematic of the device. Grey: insulating substrated. Blue: metallic contacts. Green: recessed gate. Bottom panel: top-down SEM image of the device. Scale bar: 0.5 μm. Adapted from 117. b) Incorporating a split-gate geometry into a growth-last device using etching of a silicon-on-insulator (SOI) substrate. Purple: split gates formed from doped silicon. Green: thermally-grown oxide. Yellow: contacts deposited on the grown oxide. Bottom panel: SEM image of the fabricated device. Adapted from 112. c) A three-gate geometry for a growth-last device. The split-gate geometry is reproduced with a wider gap between the two split gates, allowing a global backgate to gate the device through the gap. Top panel: device schematic. Bottom panel: top-down SEM image of the device. Adapted from 118.

thermore, the technique is highly statistical with low yield, since only a small percentage of devices survive growth and simultaneously have a clean nanotube.

To achieve greater design flexibility together with the cleanliness of growth-last, one can bury multiple gates under a deposited oxide to seal them from melting and electrically shorting. Suspended nanotube devices have been created with this technique and used for photocurrent measurements of clean p-n junctions formed by local gates (see Figure 2.5)119,120. Ultra-clean devices at low temperature, however, have not yet been demonstrated.

One drawback of this method is the presence of the oxide layer itself, which can introduce noise and electrostatic disorder. If there are charge traps, they may be charged by the gate voltages, strongly affecting the electrostatic potential. But the low yield of functional and clean devices, due to the statistical nature of the fabrication and growth processes, remains a key obstacle.
2.2.3 Stamping

Ideally, one would like to avoid the limitations on design and materials inherent in a growth-last process, while maintaining the ability to make suspended nanotube devices. One way to achieve this is a stamping technique, developed in parallel to the nano-assembly technique we describe in this thesis (chapter 3)\(^ {92,121}\). Here, a device with high contacts is mechanically put into contact with a suspended nanotube, which is grown separately on a second chip (see Figure 2.6). The mechanical interactions cause the nanotube to stick to the measurement chip, sometimes resulting in transfer. With this technique, multi-gated devices have been demonstrated and used to form low-temperature quantum dots and mechanical resonators\(^ {92,93,98}\). So far, however, an ultra-clean low-temperature 1D system has not yet been demonstrated with a stamped device.

The drawbacks of this technique include the low yield of mechanical transfer, which remains a statistical process. A related issue is that the electrical lines to the measurement chip are floating during the transfer, and thus may experience electrical spikes or noise that can lead to the nanotube snapping to the gates or getting electrically cut. These problems are only avoided by obeying limitations on device geometry and materials\(^ {92,93,121}\).
Figure 2.6: Stamping method of device fabrication. a) Schematic of the process. Nanotubes are grown on an etched quartz sample. Separately, a device is fabricated on a pillar of silicon oxide. The two samples are then brought into contact, and the nanotube is mechanically transferred to the device. The gates are covered with oxide to prevent the nanotube from shorting to them. b) SEM image of a five-gated device fabricated with this technique, showing the nanotube touching the gate-covering oxide. Adapted from 92. c) Schematic of a two-gate device fabricated with stamping. d) SEM image of the two-gate device. Blue: contacts. Yellow: recessed gates. Adapted from 121.
2.3 The Need for a Deterministic Method

The two conditions necessary to realize 1D systems on nanotubes are, first, the ability to deterministically select nanotubes on the basis of their properties and second, the creation of a pristine electrostatic environment. Attempts to create nanotube devices by choosing them on a substrate have wrestled with environments that introduce disorder and chemical processing that damages nanotubes. Growth-last and stamping techniques avoid these issues, at the cost of being statistical with extremely low yield and with limitations on device geometry and materials. Creating long nanotube devices with many gates and design freedom seems to be a daunting task.

Ideally, we should be able to choose nanotubes on the basis of their clean electronic properties, while preserving a suspended device geometry to minimize disorder. In the next chapter, we describe a new method that addresses these issues using a deterministic fabrication technique. We will see that realizing these goals with our new methodology enables a vast array of new experiments.
A New Generation of Clean and Complex Carbon Nanotube Devices

To access the wealth of physics that nanotubes promise, we need a new approach to device creation. Ideally, we want to have complete flexibility in device design, both in geometry and materials. Just as importantly, we need a deterministic way to make nanotube devices, where we choose nanotubes based on their bandgap and their cleanliness. We now describe how we have realized this goal.
3.1 Overview of the Nano-Assembly Technique

The ideal device we seek to create consists of a single nanotube sitting on tall source and drain contacts, and suspended without slack over an arbitrary number of gates, as illustrated in Figure 3.1a. Our approach to realize this device follows two main principles. First, we separate nanotube growth and device fabrication to obtain complete design and material flexibility, and second, we use scanning probe microscope manipulation to achieve deterministic assembly.

On one chip, long nanotubes are grown suspended without slack over ~100μm wide trenches (Figure 3.1b; see section 3.2). On a second chip, we fabricate the electrical circuit on a narrow cantilever, ~10μm wide (Figure 3.1c; see section 3.3). To realize the device shown in Figure 3.1a, we fabricate an array of parallel electrodes where the outer electrodes are taller than an inner set. These two chips are entirely independent and neither one imposes any restrictions on the other.

The two chips are then inserted into a scanning probe microscope, schematically shown in Figure 3.1d. A capacitive detection scheme allows us to detect the edge of the nanotube chip and the positions of the trenches with ~1μm accuracy (see section 3.4). The circuit cantilever is then positioned above a selected trench, and slowly inserted into the trench. The taller contacts on the circuit touch a nanotube, which remains suspended over the gates.

Upon contact with a nanotube, electrical measurements detect the change of resistance between the two chips. Transport measurements can now be performed in situ, directly testing the electronic cleanliness and bandgap of the nanotube, at room or low (T=4K) temperatures. If a nanotube does not have the desired properties, we can detach from it using slow, gentle nanometer-scale motion with piezoelectric scanners, and move to a different trench or a different segment of the same nanotube. Importantly, touching and detaching from nanotubes can be done without damaging the nanotubes or the circuit. This allows us to directly select for bandgap and cleanliness before a nanotube is permanently attached to the device.
Figure 3.1: Illustration of the nano-assembly technique for creating clean and complex nanotube devices. a) An example of a device with desirable characteristics: a nanotube connected to source and drain electrodes (yellow) and suspended above multiple gates (blue). We assemble such a device from two independent chips: b) The "nanotube chip" with parallel nanotubes grown over wide trenches. See section 3.2. c) The "circuit chip" consisting of contact electrodes (yellow) and gate electrodes (blue) formed on a narrow cantilever. Typical dimensions are indicated. See section 3.3. d) The nano-assembly is achieved with a scanning probe microscope (illustrated), which controls the relative position of the two chips with high precision (arrows indicate directions of motion). e) A device is made by inserting the cantilever into a trench and "mating" the electrical circuit to several nanotubes until a desirable one is found. The nanotube touches the taller metallic contacts and remains suspended over the gates, allowing in-situ transport measurements (inset). Once a desirable nanotube is identified, it is locally cut by passing a large current between adjacent pairs of side contacts, without damaging the suspended segment, disconnecting the device from the nanotube chip. See subsection 3.4.3.

Once a nanotube is chosen, we permanently attach the nanotube to the circuit. This is done by passing a high current through adjacent pairs of contacts at either side of the suspended device, cut-
ting it between the chosen contacts (see subsection 3.4.3). This process does not damage the segment above the gates, which is attached to the device but is then separated from the two segments going to the nanotube chip. By retracting the circuit chip, these two nanotube segments fall away to the nanotube chip, leaving only the central segment. The circuit chip, with the assembled nanotube, can then be removed from the microscope and transferred to other setups for further measurements.

The nanotube chip, circuit chip, and assembly microscope are all specially designed to realize deterministic nano-assembly. Below, we describe in more detail each of these components. At the end of the chapter, we present devices made with this technique, including suspended devices with large numbers of gates, and a new type of device enabled by our technique with multiple nanotubes at selected positions.

3.2 The Nanotube Chip

The nanotube chip must fulfill several demanding goals for use in the nano-assembly process. It must support long suspended nanotubes, at least 50\(\mu\)m in length, oriented perpendicular to the trenches. The nanotubes should be straight, without slack, grown in the cleanest possible conditions, and be individual, single-walled tubes.

Chemical vapor deposition (CVD) is a growth method that naturally provides parallel, flow-aligned nanotubes. In this process, catalyst nanoparticles are deposited on a chip, which is placed inside a quartz tube sitting in a furnace. Gases, including a carbon-carrying gas, flow through the quartz tube and the furnace is heated. The catalyst nanoparticles absorb carbon from the dissociated carbon feedstock, and the absorbed carbon begins to form the nanotubes. Importantly, if the growth is done on a flat substrate such as a Si/SiO\(_2\) chip, the nanotubes grow up into the gas stream and settle permanently on the substrate after some time\(^{123,124}\).

The result of the CVD process on a Si/SiO\(_2\) chip is shown in Figure 3.2. The nanotubes sitting
Figure 3.2: Flow-aligned CVD growth of nanotubes. Nanotubes are grown from catalyst deposited on a Si/SiO$_2$ wafer, visible as the dark background. The bright rectangles are etched into the oxide and the underlying silicon. Nanotubes are visible as the bright lines crossing the oxide. They follow the direction of the flow, but also fluctuate along the flow direction. This additional length is the source of slack in suspended tubes. Scale bar: 100$\mu$m.

on the insulating oxide glow white in the SEM image. They are aligned with the gas flow direction, up to small deviations. Two main issues can be identified from this image. First, the nanotubes are not perfectly straight and ‘wiggle’, and thus the total nanotube length is longer than the end-to-end length. This means that suspended nanotubes will have slack and will be suspended in a non-flat way, as described below. Second, the nanotubes are dense, and thus can stick to each other to form bundles.

The first issue of slack is illustrated in the top left of Figure 3.3. Catalyst is deposited on a chip containing tall, steep pillars. Nanotubes grow from this catalyst upward into the gas stream, fluctuating as they become long. Eventually, they settle onto the next pillar. Van der Waals forces cause them to ‘zip’ down to the substrate, and the excess length is taken up by the bowed shape of the suspended nanotube. This outcome is visible in the bottom panel of Figure 3.3. Here, trenches have been formed by plasma-etching a silicon wafer to achieve deep trenches with steep sidewalls (visible
Figure 3.3: Slack in suspended nanotubes grown across steep trenches. Top left: schematic of CVD growth of nanotubes suspended between two steep pillars. Step 1: A nanotube grows from a catalyst up into the gas stream (blue arrows), fluctuating due to its long length. Step 2: The nanotube fluctuates or settles onto the opposite pillar. Step 3: Van der Waals forces cause the nanotube to ‘zip’ down to the substrate. The excess length contained between the two contact points causes the nanotube to become slacked, arching in its trajectory. Top right: A deep plasma-etched Si sample viewed edge-on. Scale bar: 100μm. Bottom: a nanotube suspended across one plasma-etch trench. The nanotube crosses the trench with a visible amount of slack, ~5μm, arching upwards in its trajectory. Scale bar: 10μm.

Bowing is a serious problem for the nano-assembly method. It can cause nanotubes to stick to the gates of the circuit chip during nano-assembly, instead of being suspended across them. It also makes it more difficult to attach and detach the circuit chip from slacked nanotubes, which have freedom of movement due to the slack. In addition, achieving zero tension for flat suspended devices requires deep insertion of the circuit into the trench, making the process far more dangerous for the circuit chip and less deterministic.

It is therefore imperative to eliminate slack, and we describe how to do this with a two-step etching process in subsection 3.2.2. This process creates triangular trenches with a high-angle sidewall,
Figure 3.4: Tensioned and suspended individual carbon nanotubes. Top right inset: schematic of tensioned suspended nanotube growth. Step 1: Growth occurs from catalyst deposited on the plateaus between the triangular trenches. Nanotubes grow upwards into the gas stream (blue arrows), fluctuating as they become long. Step 2: The nanotubes fluctuate or settle onto the next plateau. Step 3: Van der Waals forces cause the nanotube to 'zip' down onto the substrate. The shallow angle of the trench lip allows them to follow the surface down into the trenches. The extra vertical distance, $\sim 10\mu m$, eliminates the bowing shape, causing the suspended nanotube to lie flat across the trench. a) SEM image of the nanotube chip, comprising 30 trenches, ~100 $\mu m$-wide. Scale bar $50\mu m$. b) Zoom-in on a trench edge showing a single nanotube growing across the trench. The nanotube growth direction is aligned perpendicular to the trenches by the feedstock gas flow. The shallow slope at the trench edge allows the nanotubes to easily stick to the surface, removing their slack (angles of the two slopes are indicated). Scale bar $4\mu m$. c) Zoom-in on another trench edge with a different suspended nanotube. Scale bar: $3\mu m$.

and a trench lip with a low-angle. Nanotubes grow upwards across the trenches, and eventually settle, with Van der Waals forces causing them to 'zip' down onto the substrate (illustrated in the top left inset of Figure 3.4). As nanotubes settle, they have a high chance of coming into contact with the
low-angle trench lip, and thus follow it down into the trench. The extra vertical distance covered, \( \sim 10\mu m \), is enough to cancel the excess length.

We have imaged hundreds of nanotubes, and find that nanotubes settle following the trench lip in nearly every case. They become suspended once they reach the high-angle portion of the trench, as can be seen in Figure 3.4b and c. Any extra slack is less than the depth of field of the SEM. Thus, the excess nanotube length, which would otherwise lead to slack, is eliminated when the nanotube is forced to traverse the extra vertical distance of the trench lip. In a few cases, the nanotube touches the trench at different heights on either side. This does not pose a problem for the mating technique, since the circuit can simply be lowered until all contacts touch the nanotube (see section 3.4). Any fluctuation in the nanotube height is also not a problem for the same reason.

The second issue, high density, can cause nanotubes to stick to each other, forming nanotube bundles. We can see how this affects suspended nanotubes in Figure 3.5. Here, CVD growth is performed on wet-etch trenches, which have a triangular cross section (visible in the upper panel of Figure 3.5a). In Figure 3.5b, the growth is performed in a high density regime, causing the suspended nanotubes to stick to each other in bundles. The forces driving nanotubes to form bundles cause them to deviate from the gas flow direction during growth, and as a result these bundles are observed at many different angles. In the insets of Figure 3.5b, we can see the junction where two separate nanotubes stick to each other, forming the bundle that crosses the trench. In contrast, Figure 3.5a shows the low density regime, where nanotubes are grown isolated from each other, avoiding bundles and preserving flow alignment.

Even when grown sparsely, shorter nanotubes can stick to long suspended nanotubes, as shown in Figure 3.6. Here, in a plasma-etched nanotube chip, the long suspended nanotubes have shorter nanotubes stuck to them close to the trench sidewall. This is visible by the bright short regions close to the wall edge that become suddenly dim (at the points marked by arrows), indicating the end of a short, stuck segment. To avoid this short-range bundling contacting the circuit chip, the trenches
need to be sufficiently wide and the growth parameters chosen to keep such segments short.

Below, we discuss in more detail how these issues are overcome, through a combination of careful chip design and fabrication and ultra-clean growth.
3.2.1 Design

The nanotube chip consists of a periodic array of trenches, whose width is the most important design parameter. For the same growth conditions, wider trenches result in a lower yield of suspended nanotube growth. However, they are also easier to detect with capacitance measurements in the nano-assembly microscope, safer for insertion of the circuit cantilever, and allow a wider selection of nanotube segments (see section 3.4). On the other hand, narrower trenches allow for certain simplifications in fabrication that can help growth yield (see subsection 3.2.2), and allow for more trenches to fit in the same overall chip width, increasing the chances of finding clean nanotubes.

The trench width also determines the best length of the circuit cantilever, which is at least double the trench width, as illustrated in Figure 3.7. As discussed above, to attain individual suspended nan-
Figure 3.7: Nanotube and circuit chip design. To avoid bundling of suspended nanotubes the typical distance between suspended nanotubes, or the inverse of the nanotube density per unit trench length, $n^{-1}_\text{NT}$, should be greater than twice the trench width $W_{\text{trench}}$. The maximum fluctuation region of a suspended nanotube with a length equal to $W_{\text{trench}}$ is indicated by the dashed lines. To have a high probability of finding more than one nanotube per trench, the circuit cantilever should then be longer than $2 \times W_{\text{trench}}$.

Nanotubes we should avoid a density so high that it leads to nanotubes sticking to each other and forming bundles. The maximum density is determined by the fluctuation of nanotubes as they grow in the gas stream. The maximum allowed fluctuation for a nanotube suspended over a trench of width $W_{\text{trench}}$ is a half-circle centered on the growth point, as indicated in Figure 3.7. Thus, the density should be less than one suspended nanotube per two trench widths along the trench. If the density is greater than this, more bundles of nanotubes will be found than individual suspended tubes, since the nanotubes reaching the suspended length are likely to touch a neighboring nanotube. If the den-
sity is less than this, we have a chance of finding more than one single nanotube per trench only if the cantilever length is at least double the trench width. This is a useful rule of thumb to keep in mind when designing the two chips.

The width of the plateaus between trenches affects two aspects: the catalyst pads where catalyst is deposited, and the growth itself. The photolithography of the catalyst pads requires a plateau sufficiently wide such that the pad can be accurately defined. Since exposure on a non-flat surface is challenging (see subsection 3.2.2), the ratio of plateau to trench should not be too large or too small. We have worked with $50\mu m$ trench and $100\mu m$ plateau, and vice-versa, with good results. However, the yield of nanotubes is significantly lower in the case that the plateau is narrower than the trench width. One possible reason is the aerodynamics of the etched chip surface: when the plateau is wider than the trenches, the trenches are a small perturbation that do not affect the flow, whereas the opposite limit can affect the flow significantly. Further work on the aerodynamics of CVD nanotube growth may clarify this issue.

For nano-assembly of devices, the edge of the nanotube chip, where the circuit is inserted, must be pristine. Thus, the trenches are actually designed to about four times the desired chip length, and cleaved after catalyst deposition into the final chip size. This way, catalyst is not deposited on the front edge of the chip, which can lead to nanotubes that interfere with nano-assembly. The size of this multi-chip 'bar' is a free parameter, chosen for convenience. The total width of the chip is fixed by the free space on the chip holder for the microscope. A special feature, such as a wide central trench or plateau, is made as a position reference for the capacitance measurements.

### 3.2.2 Fabrication

Our nanotube chips are fabricated on standard silicon wafers with a silicon oxide surface layer. The three major steps in making them are etching, metallization, and catalyst deposition.
Etching

Silicon etching is a common fabrication process with many available techniques, but the type of etching performed can strongly affect the way nanotubes grow on the chip. We use chemical vapor deposition (CVD) to grow our nanotubes, a process that takes place at high temperature under a unidirectional gas flow. Nanotubes grown on a silicon substrate with this type of process typically grow upwards off of the chip, align with the gas flow, and settle onto the substrate after growth. While the gas flow alignment for long ($\gtrsim 10\mu m$) nanotubes is typically good to within a few degrees, each individual nanotube experiences fluctuations along its length during growth and settling, leading to slack (see above).

We have developed an etching process specifically designed to create a novel trench geometry that keeps our suspended nanotubes flat. We use a two-step etching process. The first is wet etching in KOH, which creates triangular trenches with a $\sim 50^\circ$ angle sidewall. To maintain reproducible results, we add IPA to the solution, a known method of reducing the violence of the etching. The etch mask is the silicon oxide layer of the wafer, which is patterned with photolithography and opened with buffered oxide etch (BOE).

The second step is etching in tetramethylammonium hydroxide (TMAH), a common ingredient in lithographic developer solution. TMAH etches silicon with a much lower sidewall angle of $\sim 23^\circ$. Crucially, the TMAH etches away the silicon at the lips of the KOH-formed trenches, due to a tendency to etch high index lattice planes. No mask is necessary for the TMAH etch, as it etches the plateaus and the KOH trench sidewalls extremely slowly. The oxide is therefore completely removed before the etching in order to expose the trench corners. After etching in TMAH, the trench consists of a double-angle structure, with the high angle of $50^\circ$ in the lower, deeper section, and opening up to $23^\circ$ at the trench lips, shown in Figure 3.8a. The temperature and concentration of the TMAH solution are important, and if incorrectly chosen, the etchant does not create the trench
lip and instead creates surface defects, shown in Figure 3.8b.

Metallization

The second step after etching is metallization. This step renders the nanotube chip highly conducting, enabling capacitance and resistance measurements in the microscope. The metal used must withstand the high temperature of CVD growth, and thus should have a melting temperature as high above \(1000^\circ C\) as possible. For us, this is platinum (with a titanium sticking layer).

Platinum has been successfully used in growth-last-style devices \(^{17,25}\) (see subsection 2.2.2). It has been shown to establish good electrical contact to nanotubes without completely melting.\(^{25}\) This is important since it means a voltage applied to the surface of the nanotube chip will also be transmitted to the suspended nanotubes, allowing electrical measurements through them.

For low temperature mating, at \(T = 4K\), the metal layer must maintain its conductivity. If it is too thin, it can become insulating at low enough temperature. This is made worse by nanotube
growth, during which the metal can ‘scar’, developing breaks in its continuity, if it is too thin or if the growth environment is too rich in oxygen. We have found that 150nm of Pt is sufficient to maintain conductivity at low temperatures. With this thickness, the internal stress of the material causes it to peel off of large 3in. wafers during evaporation. This can be avoided by interspersing the Pt with extra layers of the Ti sticking layer every ~ 50nm.

Catalyst Deposition

Catalyst for CVD growth of nanotubes consists of metal nanoparticles, which can be dispersed on a chip in a variety of different ways. The main criterion for any deposition method is the ability to control where the catalyst falls. If it is deposited on the entire chip, nanotubes will grow both from above and from inside the trenches. This leads to nanotubes in the trench sticking to suspended nanotubes and forming bundles, as shown in Figure 3.10.

To control where the nanotubes grow, we define catalyst pads with photolithography on the plateaus between trenches, illustrated in Figure 3.10b. This can be challenging because it involves
Figure 3.10: Controlling catalyst deposition with photolithography. a) SEM image of a nanotube suspended over an etched trench. Catalyst was deposited everywhere on the chip. In the center of the image, a nanotube growing from the flat bottom of the trench has stuck to the nanotube suspended across the trench. Scale bar: 2 μm. Top right inset: central region of the image with the focal point of the SEM beam on the trench bottom, showing the interfering nanotube growing up from the bottom of the trench. White spots are catalyst. Lower left inset: cut-away view of the imaged geometry, showing one nanotube crossing the flat-bottomed trench, and one nanotube growing up from the trench bottom. b) Using photolithography to define catalyst pads on plateaus. Cut-away views of trenches, grey: silicon, purple: photoresist. Top panel: If done improperly, the resist will open up at the corners of the trenches, leading to undesired deposition at a critical region. Bottom panel: With the right trench geometry and resist spinning, the resist will close at the trench corners, allowing well-defined catalyst pads.

performing photolithography on an etched surface, on which photoresist can flow and have a varying thickness. For typical photoresists with ~ 1 μm thickness, this is especially problematic at the trench lip, where the resist thins to the point of exposing the underlying substrate (top panel, Figure 3.10b). Catalyst can stick to this exposed substrate, causing nanotubes to grow from this corner region and stick to nanotubes arriving from the bigger pad on the plateau.

The TMAH rounding of the trench lip naturally provides a solution to this problem. With a thick enough resist, the combination of trench lip and viscous resist keeps the resist coverage smooth throughout the chip (bottom panel, Figure 3.10b). We use AZ4562, which can give up to 10 μm thick layers. This enables photolithography on the plateaus, where we define catalyst pads by opening the resist.

The catalyst pads can be as simple as rectangles, or more complicated patterns. For example, we

45
have tried checkerboard pads, but so far with no measurable advantage. The key parameter is the
distance of the pad edge from the trench lip: too far, and the yield of suspended nanotubes can be
significantly lower. Typically, one desires less than 10\textmu m distance.

Catalyst deposition can be performed using wet solution or metal evaporation. We have observed
growth with both methods. The simplest and most successful for us is a wet deposition recipe con-
taining alumina, which acts like a mechanical support to raise the catalyst off of the surface of the
chip, promoting suspended growth\textsuperscript{123,125}. The catalyst and alumina is dispersed in water, to keep
from dissolving the photoresist. Before deposition, the chip receives a plasma descumming, to re-
move photoresist residue from the pad area and to make the surface hydrophobic. A droplet of the
catalyst solution is then placed on the chip. During deposition, the chip should be kept in a closed
container, or the solution may evaporate, leaving behind a dirty surface. This should be prevented,
if necessary, by adding more solution to the droplet. After the deposition time, typically a few min-
utes, the droplet should be washed away with water and blow dried with nitrogen.

The circuit chip will enter the trenches at the edge of the nanotube chip, so it is crucial to keep
the edge of the chip clean and free of spurious growth. To guarantee a clean edge, the deposition is
performed on a long chip with no exposed trench edges. After deposition, the long chip is cleaved
into smaller pieces, with the photoresist layer still present. The cleaves cut across the trenches, leav-
ing an exposed trench edge. The surface of the chip remains protected by the thick photoresist, and
the edges are kept pristine. The chip can then be washed with solvents to remove the photoresist,
and plasma cleaned to remove residue.

It is important to emphasize that the application of photoresist means that organic residue will
be left on the chip. For CVD growth with a carbon-carrying process gas, this can encourage the
growth of amorphous carbon. Such small-scale 'dirt' can get stuck to the suspended nanotubes or
to the circuit, contaminating the process. This is shown in Figure 3.11, where amorphous carbon is
seen in side a trench and sitting on a suspended nanotube. It is therefore crucial for nano-assembly
to ensure that the plasma steps properly remove any organic residue. Depending on the degree of contamination, this can be checked optically, in SEM, or even AFM.

One strategy to maintain a clean catalyst pad region is to evaporate a second layer of Pt before deposition. The second Pt layer raises the catalyst to be higher than its surroundings, promoting suspended growth, and ensures a clean environment for deposition. This requires a lift-off compatible resist, which is typically much thinner than the thick resist we use for the widest trenches. Consequently, this technique is possible only for smaller trenches, not wider than $\sim 50 \mu m$, using slow spinning with gradual speed-up, and using extra exposure time to help keep the trench corners covered with resist before deposition.

3.2.3 Growth

Chemical vapor deposition (CVD) of carbon nanotubes proceeds along the following lines. The chip is inserted into a quartz tube sitting inside a furnace. An inert gas, such as argon, is used to flush...
the system of atmosphere. Hydrogen is then flowed through the tube to prepare the reducing environment for the catalyst. The furnace temperature is raised and the catalyst reduces. The carbon-carrying process gas is then allowed to flow for the growth time. At the end of the growth time, the carbon-carrying gas is shut off, the temperature of the furnace is lowered, and all gases are turned off.

This process sounds deceivingly simple. CVD growth of nanotubes remains a field of research with open questions. Starting a system from scratch can therefore be a daunting process, with a large number of variables to control for and understand. In the following, I’ll give a set of guidelines, observations, and techniques gleaned from extensive work on CVD growth, which can hopefully serve to simplify the process.

The CVD System

The most important point bears emphasizing: cleanliness is paramount. While CVD systems of varying cleanliness can give nanotube growth, one often runs into problems with no sense of which variable to modify. The best is to have a reliable starting point for work, which can be accomplished by following some central guidelines.

The first rule is to have a leak-tight system. On the atmospheric pressure side of the system, this can be checked with helium leak testing. On the high pressure side between the gas cylinders and the mass flow controllers, over-pressure tests can be performed to catch leaks. This is done by pressurizing the system with a pressure as high as the mass flow controllers can tolerate and waiting over time to watch for any drops. In our system, we use VCR connections to keep our leak rates very low. The seals of the quartz tube are also important, since they are typically the weak point in the system. Standard seals with a single o-ring do not form a reliable, reproducible seal. Double-o-ring seals with an intermediate vacuum space are more dependable. Choosing the right o-ring size is key.

The second rule is to keep as much of the system as possible isolated from the outside. Since it is only necessary to open the quartz tube to insert and remove samples from the downstream end, the
rest of the system can be isolated with valves.

A third way to ensure a reproducible system condition is to ‘reset’ the quartz tube condition after use by baking the tube at high temperature, $> 900^\circ C$. This can be done while the tube is open to air or closed with a chosen environment, like argon or hydrogen, which has been shown to help reduce disorder or contaminants in the case of single-crystal graphene CVD. If the tube is contaminated, it can be cleaned with a strong lab soap like Alconox. Solvents such as acetone and isopropanol should be avoided, as they leave a residue which is difficult to remove and without which we obtained more reliable growth results. One should also avoid scratching the tube with the insertion rod, which can be prevented by using a teflon or teflon-covered rod without any exposed metal.

The Growth Process

The first step in growth is to flush the system with inert gas. We use argon. This can be augmented with pump-purge steps for a cleaner system. Next, hydrogen is introduced for some time to allow mixing and to establish a consistent flow and gas proportions. The oven is then warmed up, with the argon and hydrogen. The warmup in hydrogen before growth is important because the catalyst is oxidized by the oxygen plasma used to clean the sample, rendering it inactive for nanotube growth. Reduction in a hot, hydrogen-rich atmosphere activates the catalyst, so it is important to give sufficient reduction time (the actual amount of time required remains an open question).

While the hydrogen plays an initial role as a catalyst activator, it also helps to maintain the growth environment free of oxygen or water. These contaminants can etch or burn nanotubes, or deactivate the catalyst. But hydrogen itself can also effectively etch nanotubes, and prevent them from growing. A rule of thumb is that the cleaner your system is, the less hydrogen you need. If you invest time in making a leak-free system with a clean environment, you should most likely reduce the amount of hydrogen in the growth process, or the growth yield can be negatively impacted.

The carbon-carrying gas used for growth can significantly impact growth parameters and results.
We typically use ethylene, which requires a very small flow rate, ~ 1% of the total gas flow rate. Another popular option is methane, which typically requires a much higher flow rate. All results discussed here pertain to ethylene. We have observed that higher proportions of ethylene yield more nanotube growth on all length scales.

After establishing a clean system, the most important parameter to define is the growth temperature. Too low, and no growth will take place. Too high, and only amorphous carbon will grow. Finding the window that gives reliable suspended nanotube growth is a first priority. This can be tricky because of dependence on other parameters. For instance, there is an interplay between growth temperature and the total flow rate. The faster the gas moves through the furnace, the less it gets heated. Conversely, a low flow rate will mean a hotter gas. The optimal temperature will therefore depend on the chosen flow rate, as well as on the position of the sample inside the furnace. The difference in temperature between the sample and the gas flowing directly above has been hypothesized to be responsible for a thermal buoyancy force which has been proposed to assist long nanotube growth. To maximize this temperature difference, one should in principle place the sample close to the upstream entrance of the quartz tube.

Other variations on sample placement can significantly affect growth. The standard is to place samples directly on the quartz tube. Another option is to use a carrier quartz slide which places the sample closer to the middle of the tube. One technique involves a smaller tube inserted into the larger quartz tube, which increases the flow velocity over the sample. Yet another possibility is to place the sample directly on the quartz tube, but upside down. These variations can give higher yield in some situations, but, like all parameters discussed here, can be different for different systems and can vary in time.

The growth time is an important knob to tune the yield. Whether the yield of suspended nanotubes is affected by the growth time, however, depends on whether the other parameters are in the right regime to begin with, so growth time isn’t a panacea.
Once the growth is done, the furnace is turned off and opened for cooling. At this stage, it is important to keep the hydrogen flow on until the sample cools to less than 200°C, to prevent the platinum surface from ‘curdling’ and causing damage to the nanotubes or becoming non-conductive. This may also prevent the possibility of damage to the nanotubes from the presence of oxygen at high temperatures.

3.3 The Circuit Chip

Figure 3.12: Fabrication of the circuit chip. a) Optical and electron-beam lithography are used to fabricate the circuit chip on a flat Si/SiO₂ wafer, defining both the nano-assembly components and the device itself. b) Deep plasma etching is used to place the device portion of the circuit on a tall and narrow pillar, designed for safe insertion into nanotube trenches.

The circuit chip consists of two main components: the geometry that enables nano-assembly, and the geometry of the device itself. The device geometry needs to sit on a cantilever-like protrusion that fits inside the nanotube trenches, as shown schematically in Figure 3.1c. This is accomplished by fabricating the circuit on a flat Si/SiO₂ wafer, as in Figure 3.12a, and then etching the silicon around the circuit so that it sits on a tall and narrow pillar, shown schematically in Figure 3.12b. Our goal for the circuit chip is to design a reliable, high-yield process with high design flexibility and robust behavior in the assembly microscope. Below we describe the design and fabrication techniques that
achieve this.

3.3.1 Design

There are three main components to the circuit chip design. The first is the optical lithography layer which defines bonding pads, lines to the cantilever, and capacitance pads for the assembly microscope. The second is the electron-beam lithography (EBL) portion on the cantilever, which defines the actual device geometry. The third is the etching, which defines the shape of the device ‘mesa’, including the cantilever portion.

Optical

The optical layer includes the line fanout and bonding pads, as well as the capacitance pads used for approach in the nano-assembly microscope. We first address the requirements for approach (see subsection 3.4.1). Three basic capacitance pads are required. Two large pads are used to find the distance between the circuit and nanotube samples, the edge of the nanotube chip and its angles relative to the circuit chip, and one to detect the trenches.

The two large capacitance pads are typically designed to have an area of \( \sim 75,000 \mu m^2 \). To find the edge with a reasonable resolution, these pads are designed with a narrow, \( \sim 100 \mu m \) width (perpendicular to the nanotube chip edge) and a much longer length (parallel to the nanotube chip edge) (see Figure 3.16a). With this size, one can obtain a large enough signal to work at reasonable starting distances, and obtain a good measurement of the trench edge and angles (see below). In order to avoid picking up the capacitance oscillation of the trench etching, the length of these pads should be an integer multiple of the trench pitch, or the sum of trench width and plateau width.

The trench-finding pad is a long and narrow rectangle designed to be smaller than the trench width (see Figure 3.16c). It is used after approaching the sample with the large pad, and thus has a
smaller typical area of $\sim 1000 \mu m^2$. It extends far beyond the large pads, so that the overlap of the large pads and the nanotube chip can be minimized and the signal of the trench-finding pad to the nanotube chip can dominate. To find trenches, its width must be much smaller than the trench width. For $100 \mu m$ wide trenches, a $10 \mu m$ wide pad works well. To increase signal, the length (parallel to the trenches) is made very long, several hundred $\mu m$'s.

Other pads can also be included for more angle measurements, such as a second pair of large pads recessed with respect to the first two. This allows all relative angles between the two chips to be measured and compensated (see subsection 3.4.1). Another option is a small pad to measure the trench chip edge with fine resolution.

The other lines are dedicated to device electrodes, such as gates and contacts. One important consideration is the capacitance of the lines to ground, which is typically determined by their capacitance to the conducting back gate, so that their area should be minimized to maintain measurement speed. The bonding pads, should be kept big enough to allow a few bonds, since bond pad damage is common.

**E-Beam**

![Diagram](image)

*Figure 3.13*: Design of the device electrodes in a simple device geometry. Important parameters include the ratio of device length to contact height, the contact width, and the length of the cutting segment (see main text).

The design of the device itself depends chiefly on the physics being pursued. The suspended device length, the contact material, number of gates and their pitch, and the height of the nanotube
above the gates all play a part in determining the observable physical phenomena (see Figure 3.13). Some parameters play important roles in the nano-assembly process. These are the width of the device contacts, the ratio between the suspended device length and the height above the gates, the distance between contacts used to cut nanotubes, and the length of exposed oxide segments.

The width of contacts affects the sticking force between them and the nanotubes contacted during nano-assembly. When touching a nanotube, very wide contacts, \( \geq 3 \mu m \), can induce the nanotube to snap to them at a distance far above the nanotube height in the trench, which can break the nanotube. Wide contacts also make it more difficult to gently detach from the nanotube, necessitating a large retraction distance to pull away from the nanotube that can cause it to break, or render it unusable for future mating.

The ratio of width to height of the contacts plays an important role. For very large aspect ratios, the nanotube may be flexible enough to stick to the gates during mating. We have successfully mated with aspect ratios greater than 10, but for aspect ratios beyond 45 the nanotube has a higher chance of touching the gates.

After choosing a nanotube, it is electrically cut between two pairs of contacts at either side of the suspended segment (see subsection 3.4.3). The current at which the cutting occurs depends on the length of the nanotube segment suspended between the cutting contacts. Longer segments of nanotube have less thermal sinking by the contacts, and thus can be cut at lower currents. To prevent accidental cutting, it is helpful to keep this length not longer than \( 1 \mu m \). On the other hand, a longer segment means easier cutting if the contact resistance is high, which may be the case when \textit{in situ} plasma treatment of the device is unavailable.

It is important to prevent nanotubes from touching the oxide of the chip. Nanotubes stick extremely well to \( SiO_2 \) and cannot be detached from it, causing them to break during retraction if they have touched the oxide. The oxide also acts as a thermal sink, rendering electrical cutting impossible. It is therefore crucial to design the chip to avoid the possibility of nanotubes sticking to oxide, by
maintaining limited spacing between any contacts with no gates in between. The distance from the last metal line to the edge of the etched pillar (‘oxide edge’ in Figure 3.13) should be kept small compared to the contact height. An oxide edge is necessary to prevent damage to the device lines due to misalignment of the etch mask photolithography and plasma etching.

3.3.2 Fabrication

The four main steps in circuit fabrication are optical layer, e-beam layer, deep etch, and final cleaning and cleaving.

Optical

The optical layer is a standard optical lithography and evaporation on a wafer. For a large yield of devices, we use at least 2 inch wafers, which means all optical lithography must be done with very small errors in the alignment between wafer and mask.

Clean processing is key in this step. Any dirt or particles causing the resist to open before evaporation will create a pillar after the deep etching of the wafer. This can sharply reduce yield, and should be avoided by ensuring the wafer is clean before resist spinning and before exposure.

As is typical for optical lithography processes, the wafer undergoes de-beading after spinning the resist. This is done by removing the raised resist at the edges with solvent. However, the exposed edge will then be covered in metal after evaporation, which will remain on the wafer for the etching process and can lead to contamination of the etching vacuum chamber. To avoid this metal ring, additional resist is added to the edge after resist exposure but before metal evaporation.

After metal evaporation, liftoff should be done rigorously, including time in heat and sonication. On a flat wafer, this does not damage devices.
**E-Beam**

The e-beam layer is a standard e-beam process on a flat wafer. All processing should be done as cleanly as possible. This includes plasma de-scumming of the developed wafer before evaporation.

E-beam-written lines sometimes develop spikes of metal at the edges of the evaporated pattern. For nano-assembly, it is important to maintain smooth contact lines. These spikes are caused by metal adhering to the resist overhang and connecting to the pattern. To avoid this, we perform the e-beam evaporation with the wafer positioned directly over the evaporation source. By minimizing the angular spread of the evaporation these features are minimized.

The e-beam pattern is typically hundreds of microns long but can be only tens or hundreds of nanometers wide. Liftoff after evaporation should therefore be done rigorously, as excess metal has a tendency to stick to these long and narrow line patterns. Fanouts to increase line width and spacing and reduce the length of the most narrow lines is useful to avoid unnecessarily difficult fabrication.

**Etch**

The etch step consists of two delicate processes: the optical lithography of the etch mask and the etching itself.

The optical lithography is demanding in that the region of the e-beam lines of the device, where the pillar will be etched, has to be accurately covered by the resist. This must be accomplished over the entire wafer. The angular alignment between mask and wafer must thus be very precise. It is key to have a properly aligned mask aligner stage.

The resist used for the etch mask has to withstand $> 100 \mu m$ of reactive ion etching, and so must be quite thick. We use AZ4562, with makes up to $10 \mu m$ thick layers. Before exposure, the resist undergoes de-beading with solvent, similar to the first optical exposure. After exposure and development, the wafer is dipped in buffered oxide etch (BOE) to remove the oxide and expose the under-
lying silicon for etching. However, because of the debeading, this will lead to etching of the wafer edges, leaving a thin wafer vulnerable to mechanical breakage. To avoid this, after exposure and development (but before final hard bake and BOE dip) additional resist is added to the wafer edge to maintain a ring of oxide around the wafer edge. This renders the wafer far more robust to handling and other mechanical stresses.

The etching itself takes place in a deep reactive ion etch (DRIE), inductively-coupled plasma (ICP) machine. We use a Bosch process to achieve high aspect ratio etching. This involves alternating between two processes: etch with SiF6 plasma, and passivation with C4F8. The process is sensitive to the cleanliness of the vacuum chamber, so before etching we perform a couple of processes to ensure a clean chamber. The first is a strong oxygen and argon plasma etch in the chamber with a blank carrier wafer, to clean and test the chamber. The second is a long passivation of the chamber with the C4F8 gas, with the same test carrier wafer. After this, a real etch process can be performed, which should be carefully monitored throughout the process to make sure nothing is going wrong. Issues that arise include a brownish residue, which often indicates contamination of the chamber, or burning of the resist, which could mean poor thermal coupling of the sample to the carrier wafer.

We achieve thermal coupling with a thermal grease between the sample and the carrier wafer.

**Clean and Cleave**

After deep etching, the wafer needs a serious clean to remove the hard-baked and plasma-etched resist. Before removing the resist, however, one important step should be performed. The deep etching can leave behind small pillars of silicon at the bottom of the etched wafer, or chips of loose silicon on the sidewalls of the etched chips. During solvent cleaning, these silicon pieces can come loose and settle on top of device layers, leading to unusable topography or damage to lines (see Figure 3.14). While the resist is still on the chips and protecting the devices, the wafer is dipped in water and gently agitated by hand for several minutes.
Figure 3.14: Silicon spikes on circuit chips after plasma etching. These can cause shorts or damage to the circuit, and move around during nano-assembly and measurement. They can be avoided by gently washing the sample in water before removing the resist etch mask (see main text). Scale bar 1μm.

The solvent cleaning to remove the resist should be thorough, including the usual solvents, acetone and methanol, and a more rigorous residue remover, like 1165. After solvent cleaning, a long, strong oxygen plasma should be performed to remove residues.

At this stage, the sample can be cleaved into chips. Optionally, it may be spun with another resist layer as a protection from cleaving damage and silicon shards. Cleaving can typically be performed cleanly enough that this is not always necessary. Usually, using a large mechanical scriber is unnecessary, and using a simple hand scriber and short scribe marks near the edge of the wafer is enough to perform cleaving, which helps avoid silicon shards. For asymmetric pieces smaller than half a wafer, the scribe should always be performed on the straight side and not the wafer edge. For pieces of three chips or less, the scribe can be performed on the sideface of the cut silicon, as opposed to top or bottom, and cleaved immediately with slight pressure from the corner of a tweezer.
3.4 Nano-Assembly with a Scanning Probe Microscope

The two chips, nanotube and circuit, are inserted inside a scanning probe microscope (SPM) specially built for the nano-assembly process. The heart of the system is a stack of piezoelectric nanopositioners (Attocube). This includes three piezoelectric motors for coarse, micron-scale motion along x, y, and z axes, and a piezoelectric scanner with nanometer resolution for motion in all three cardinal directions. The system is wired for access to all lines of the circuit chip and to the metallic surface of the nanotube chip. The piezo-stack is enclosed and attached to a surrounding titanium cage designed for high stiffness, to eliminate spurious vibrations.

The sample is inserted via a load-lock which contains an argon gun. This allows for in situ argon ion plasma cleaning of the contacts surface, which we have found helps to achieve good contact resistance to the nanotubes.
3.4.1 Approach

The first step is to bring the circuit chip into close proximity with the nanotube chip. This is done blindly using a capacitive detection scheme. Displacement current is measured between the nanotube chip and the capacitance pads of the circuit chip, with the circuit is shown in Figure 3.16. A reference signal is applied to the nanotube chip, with an amplitude of at least 1V and a frequency of $\sim 12\,kHz$. The capacitance pads are connected to the input of a lock-in amplifier, and the out-of-phase component of the measured signal constitutes the displacement current, from which is computed the capacitance.

To detect the nanotube chip, three scan types are used as illustrated in Figure 3.16. The first detects the edge of the nanotube chip, the second finds the distance between the two chips, and the final scan detects the trenches.

To detect the edge, the capacitance pad is positioned over the nanotube chip and scanned over the edge, moving away from the nanotube chip. The measured capacitance between the chip and the pad drops when the pad crosses the edge, and the inflection point reveals the position of the edge.

The second scan detects the vertical position of the chips by placing the pad over the nanotube chip and moving them closer together. The distance is determined using a parallel plate approximation, by writing the capacitance between the two chips as

$$\frac{1}{C} \propto z - z_0 \quad (3.1)$$

where the constant of proportionality depends on the effective area and the dielectric constant and $z_0$ is the point of contact between the two chips. The local slope of the capacitance versus vertical position gives a progressively better approximation to $z_0$ as the two chips approach one another.

The third scan type detects the nanotube trenches by scanning the narrow and long trench-finding capacitance pad over the trenches. The measured capacitance oscillates, reflecting the peri-
Figure 3.16: Detecting the nanotube chip edge, distance, and trenches with capacitance measurements. a) Edge detection. The nanotube chip is excited while the capacitance pad is connected to the input of a lock-in amplifier for measurement of the displacement current. Scanning over the nanotube chip edge results in a drop in capacitance, b), and the inflection point gives the edge position. c) Distance detection. The displacement current is measured as the two chips are brought closer together. The inverse capacitance approaches linearity with position, d), and the zero crossing gives the point of contact. e) Trench detection. The trench-detecting pad is now measured, and the scan is performed perpendicular to the trenches. The capacitance oscillates, f), revealing the trench positions as the capacitance minima.
odic structure of the nanotube chip, and allowing the trenches to be precisely located. The oscillations can fluctuate in amplitude due to angles and fluctuations of the etching. A feature on the nanotube chip, such as a large trench or plateau, can be used as an absolute position reference.

The trench detection scan requires the two chips to be close enough to obtain accurate trench locations, but this also increases the risk of a crash. To safely approach the nanotube chip, the relative angles of the two chips must be known in order to calculate the minimum approach distance. To obtain the relative angles, three scans are performed with pairs of two large capacitance pads, as illustrated in Figure 3.17a-c. An edge detection scan with two pads, displaced on the circuit chip along the edge direction as in Figure 3.17a, results in two inflection points, reflecting the relative x-y angle. A vertical scan with the same two pads provides two crash points, reflecting the relative x-z angle, shown in Figure 3.17b. A vertical scan with two pads positioned relative to each other along the trench direction reveals the z-y angle, shown in Figure 3.17c.

Once the trenches are found, the circuit cantilever can be inserted into the trenches to search for nanotubes, but there is one remaining danger. If the nanotube chip has an angle with respect to the axes of motion of the microscope, then the circuit will be inserted at different positions in different trenches. This is true for all three axes, and can lead to crashes. To avoid this risk, the angle of the nanotube chip can be measured and corrected. This procedure is illustrated in Figure 3.17d-f. In contrast to the relative angle scans, to extract the absolute angle of the nanotube chip requires one capacitance pad, but scanned at two different positions. The edge locations and contact points obtained at different pad positions then reflect only the nanotube chip angle. The scans to extract the nanotube x-y, x-z, and z-y angles are shown in Figure 3.17d, e and f, respectively. With these angles, the circuit can be inserted into all trenches at the same local position relative to the trench.
Figure 3.17: Extracting relative and absolute chip angles using capacitance measurements. To obtain the relative angles of the two chips, we measure with two capacitance pads during a single scan. a) A y-scan for pads offset along x gives the x-y angle. b) A z-scan for pads offset along x gives the z-x angle. c) A z-scan for pads offset along y gives the z-y angle. To extract the absolute angles of the nanotube chip relative to the motor axes, we measure with a single capacitance pad scanned at two different positions. d) Two y scans for pads offset along x gives the x-y angle. e) Two z scans for pads offset along x gives the z-x angle. f) Two z scans for pads offset along y gives the z-y angle.

3.4.2 Nano-assembly

For contacting nanotubes, the circuit is now changed to that shown in Figure 3.18. The goal is to measure the change in resistance between the nanotube and circuit chips when a nanotube touches the high contacts on the cantilever. The frequency of the excitation is lowered to < 1kHz, and the
contacts are connected to a lock-in amplifier, allowing the in-phase, resistive component of the signal on the contacts is monitored as the circuit is inserted into the trenches. The gates are all grounded, an important aspect of the process that reduces the chance of the nanotube snapping to the gates. A resistor with $R \sim 1M\Omega$ can be put in series with the voltage excitation as a precaution to prevent burning nanotubes with excessive current. Optionally, the backgate may be connected to a separate, simultaneous measurement, serving as a check for undesirable touches or crashes.

The circuit chip is now positioned above one of the trenches according to the positions determined during the capacitance scans. Only the narrow etched pillar overlaps the trench; the rest of the circuit chip lies safely outside the nanotube chip. At this point, the circuit is carefully lowered into the trench with the piezoelectric scanner. If the end of the scanner range is reached, the scan-
ner is retracted and a small step is made with the coarse motor. This process is repeated until the circuit either reaches a specified endpoint, or a nanotube is touched. If the end point is reached, we retract the circuit and increase the overlap of the pillar and the trench (i.e. a step in the y direction of Figure 3.16), repeating the process until the entire pillar overlaps the trench and no further steps are possible without crashing.

![Graph](image)

**Figure 3.19:** Establishing contact to a nanotube suspended between two contacts. One contact is sourced with voltage up to $3\,V$, while the neighboring contact is grounded. All other electrodes, including the backgate and the nanotube chip, are floating. A compliance of $1\,\mu A$ is used to limit the current. Blue: ramping the voltage. Red: reducing voltage. The current is observed to jump up to the compliance value around $2.5\,V$, returning to a lower-resistance state on decreasing the voltage. Inset: Measurement circuit overlaid on a SEM image of two neighboring contacts. A source-measure unit is connected to one contact while the neighbor is grounded. All other contacts are floating.

While the circuit is being lowered into the trench, the resistance between the two chips may drop below a threshold value, indicating possible contact with a nanotube. In this case, the scanner motion is immediately stopped, and the resistance between each contact electrode and the nanotube chip is measured to determine which contacts are touching the nanotube. In some cases, only one or a few lines are touching. This can be caused by a z-x angle of the nanotube or the circuit, or by softness of the nanotube causing one side of it to attach first. One way to achieve contact on all device lines is to continue slowly lowering the circuit into the trench until contact is measured on all lines;
this can also improve contact resistances. Alternatively, voltage can be applied to the contact touching the nanotube and grounding a neighboring line, and floating all other lines (and the nanotube chip), as shown in Figure 3.19. A current measurement between the two neighboring lines as the voltage between them is increased shows near-zero current, until a critical point where the current shoots up to a safe compliance value. On the downward voltage sweep, the nanotube is in a lower-resistance state, showing that it has established contact with the neighboring line. This process can be repeated to establish contact with all contact lines of the device, and is useful when contact resistances are high.

Now the nanotube may be checked in situ. Gate-dependence of nanotube conductance, measured with a lock-in, reveals whether the nanotube is metallic, small-bandgap, or semiconducting. If the mating process is performed at 4K, additional tests of cleanliness may be performed with the local gates (see chapter 4).

If the nanotube is dirty or not of the desired type, the circuit can be detached from it by slowly retracting the piezoelectric scanner. The resistance between the two chips increases sharply when contact is lost. The circuit can then be repositioned above other trenches, and the process repeated until a desirable nanotube is found.

Once a nanotube is found, the contact resistances may be high, and it may be necessary to improve contact resistance through annealing. This can be done by running current across neighboring contacts. By cycling the applied voltage, one can improve contact resistance, with results similar to the trace shown in Figure 3.19. This process is helpful in the absence of plasma treatment of the sample before assembly.

3.4.3 Cutting

After the circuit is mated to a nanotube and in situ transport measurements show the nanotube is clean and has the desired bandgap, we often want to separate the circuit chip from the nanotube
Figure 3.20: The selective nanotube cutting process. a) Main panel: Scanning electron micrograph of a seven-gated suspended nanotube device. Top-right and bottom-left insets: magnified top-down views of suspended segments that were cut using Joule heating with a flow of current (see text). The cut position is indicated by an arrow. Top right and bottom left traces: I-V curves of the burning process during voltage ramp-up. b) A cutting I-V curve during voltage ramp-up showing two abrupt current drops, attributed to the presence of a two-nanotube bundle or a double-walled nanotube.
chip to allow the transfer of the device to other measurement setups. We achieve this separation by controlled cutting of the nanotube at well-defined positions, as explained in detail below.

The process is demonstrated in subsection 3.4.3, which shows an SEM image of a seven-gated suspended nanotube device that has been selectively cut at two places (a magnified top-down view of these cuts is shown in the insets; the cuts are indicated by arrows). The relevant device segment is at the center, above the gates. To enable cutting at various locations we fabricate several contacts at each side of the device. The cutting process consists of applying a voltage between two adjacent contacts, which drives a large current through the short suspended nanotube segment between them. When the current passes a critical threshold the nanotube breaks at a single point, close to the center of the suspended segment. This cutting is believed to be due to Joule heating that leads to the highest temperature near the center of the suspended segment, which is farthest from the contacts that dissipate the heat. The line traces in the insets of Figure 3.20a show the current-voltage characteristics measured during the cutting. As a function of the applied voltage the current grows monotonically, until reaching the critical current (15μA-30μA) and then dropping abruptly to zero, indicating that the segment between the contacts was cut. Measurement of transport through suspended nanotube segments adjacent to the one that was cut before and after its cutting shows that they remain unaffected by this local process.

Another technique that was found to efficiently cut the nanotube locally is the application of fast voltage pulses (typically 10 V/0.5μs) to one contact while its neighbor contact is grounded and all other contacts are floating. However, we generally prefer to do the cutting using the first approach (DC current) since it further allows us to distinguish between an individual single-wall nanotube vs. bundles or multi-wall nanotubes. For the latter, the cutting does not happen in a single step, but often exhibits multiple steps that correspond to the multiple tubes or multiple shells breaking one at a time. An example of such a two-stepped cut is shown in Figure 3.20b. We observe the same pattern of steps when cutting the same nanotube at different junctions, demonstrating that these reflect
the intrinsic properties of the tube and not of the junctions. Bundles or multi-wall nanotubes also tend to be cut at higher currents, reaching > $100\mu A$ for the largest bundles. If upon the first cut we observe any indication of a bundle or multiple shells we detach the circuit from the nanotube before performing the second cut and move to mate with a different tube. In this case, both segments remain attached to the nanotube chip and fall away from the circuit chip. In general, we choose the growth parameters to yield sparse growth of suspended nanotubes, thereby avoiding bundles (see section 3.2). However, if we find, using the above measurements, that a specific chip has a high density of nanotubes or indications of bundles we discard it.

3.5 Resulting Devices

With this new nano-assembly process, we have succeeded to create unprecedented suspended nanotube devices. Figure 3.21a shows a representative seven-gate device made by our mating technique using nanotube and circuit chips similar to those shown in Figure 3.4 and Figure 3.15. The nanotube
is perpendicular to the gates, suspended without slack over a length of 1.2μm at a fixed height of 130nm above all gates, anchored over the entire length of the contacts and does not touch silicon oxide. These characteristics are achieved in the vast majority of mated devices, such as those shown in Figure 3.22, which have suspended nanotubes from 900nm to 2.6μm in length, with 5 to 16 gates, with different geometries and materials. Geometrically, these devices are the closest yet achieved to the ideal illustrated in Figure 3.1.

What is the limit on the number of gates? In methods such as growth-last or stamping, slack of the nanotube is unavoidable, and there is no control over gate voltages during the stamping/growing. This means the nanotube can be electrostatically attracted to the gates, causing it to snap to the gates when they are placed in close proximity. This severely limits how close the nanotube can be sus-
pended above the gates in these techniques. Typically, if the suspension height is smaller than 1/10 of the suspension length the nanotube would stick with good probability. Assuming that to get good spatial resolution one would like the distance between the gates to be comparable to the nanotube-gates distance, then this limits these techniques to ~10 gates. In our case, we start with straight tubes without slack and we can ground the gates during the assembly process, preventing any attraction. This allows us to achieve much more extreme suspension geometries. In fact, we succeeded to make devices that are 4.5\(\mu\)m long, suspended as close as 100nm to the gates, an aspect ratio of ~1:45. This implies that a device with ~45 gates whose pitch is equal to their distance to the nanotube can be made. We did not try yet to suspend nanotubes with even higher aspect ratios but considering the ease with which we fabricated the above mentioned devices we believe that they are still far from the full capabilities of our technique.

Our precision nano-assembly technique also allows the creation of an entirely new class of devices with multiple, independent nanotubes at chosen locations. To localize multiple nanotubes at specific places within a chip, we pattern contacts of different lengths on the circuit chip. During mating, we move the cantilever along the trench in small steps with the piezo-scanner. Once a nanotube is found, we successively touch and detach from it, finely stepping along the trench direction until resistance measurements indicate that it is touching contacts corresponding to a specific location. Figure 3.21d shows a double-nanotube device made with this technique. The first nanotube is positioned on a set of shorter contacts, with matching gates, and cut to electrically isolate it from all other contacts. The second nanotube is then positioned on a longer set of contacts, with a second set of gates wrapped around the longer contacts, allowing both nanotubes to be independently contacted and gated.

In this chapter we have shown that precision nano-assembly allows us to create devices with unprecedented freedom of design, materials, and nanotube placement. To realize this, we made new advances in nanotube growth, circuit fabrication, and nano-assembly in a microscope. How clean
are these devices, and how much freedom do we have in manipulating electrons on the suspended nanotube? In the next chapter, we investigate the electronic properties of these new devices.
The new technique described in chapter 3 is capable of realizing nanotube devices with the suspended, multi-gated geometry of Figure 1.7b, and new devices with multiple, nano-positioned nanotubes. But are these devices as ideal electronically as they are geometrically?

In this chapter, we answer this question using low temperature transport experiments on sus-
pended carbon nanotubes. We will show that these devices are pristine, exhibiting characteristics nearly free of disorder, as we will quantify. Instead, only electrostatics enters into their description, and we will experimentally extract the electrostatic description of the device, allowing precise design of the 1D electrostatic potential for electrons. We will also show how multi-nanotube devices, a novel new class of devices enabled by our technique, can lead to new functionality and a wide array of new experimental possibilities.

4.1 Localizing and Moving Electrons

We start with the simplest experiment on a five-gated, small-bandgap nanotube device, with all gates chained together, reproducing past single-gate transistor experiments. We use gold contacts that dope the nanotube segments above them with holes, and control the doping of the suspended nanotube segment electrostatically with the gates. As a function of the gate voltage the conductance measured at $T=4\text{K}$ shows three regimes (Figure 4.1a): At negative voltages the suspended segment is hole-doped, forming a continuous “nanotube wire” whose conductance is weakly gate-dependent. At intermediate voltages the nanotube is doped into its bandgap, determined to be $34\pm5\text{meV}$ from finite bias measurements, and the conductance is suppressed. For positive voltages the suspended segment is doped with electrons, forming a pair of p-n junctions near the contacts which confine a large quantum dot, whose charging by individual electrons generates Coulomb blockade oscillations in the conductance (see B). The oscillation periodicity, $\Delta V_g=31.5\pm1.5\text{mV}$, given by $\Delta V_g = e/C_g$ with $e$ the electron charge and $C_g$ the gate capacitance, agrees well with that expected from the capacitance of the length of the suspended segment, $L=880\text{nm}$, to all five gates. The corresponding charging energy of this large dot, obtained from finite-bias Coulomb diamonds, is $E_C=10\pm2\text{meV}$. The clean and regular spectrum of oscillations therefore signifies the formation of a single quantum dot over the entire suspended nanotube length, whose electronic cleanliness is comparable to the
best ultra-clean nanotube devices made to date.\textsuperscript{33,112,117,118} Figure 4.1: Localizing and moving electrons on a five-gated, small-bandgap nanotube device. a) Top: SEM image of a device similar to the one measured (gate numbers are indicated). The nanotube is locally coloured according to its doping: holes – blue, electrons – red. The suspended segment is electrostatically doped by the gates while the segments above the contacts are hole-doped by the metal. Bottom: conductance, $G$, measured as a function of a common voltage on all five gates, $V_g$. Coulomb oscillations are apparent at positive gate voltages due to the formation of a quantum dot extended over the entire suspended nanotube. Insets show position-dependent nanotube band diagrams in the three different conductance regimes: a hole-doped “nanotube wire”, the nanotube bandgap, and electron Coulomb oscillations (hole band – blue, electron band – red). b) Similar measurements as a function of voltages on five individual gates, $V_g(i)$ (the gate index), while the other gates maintain fixed hole-doping voltage, $V_g(j) = 0$. In each trace a small electron quantum dot is formed above the corresponding gate (side illustrations). c) Conductance, $G$ (colourmap), measured as a function of the voltages on two adjacent gates, $V_g1$ and $V_g2$ (top illustration). From the bottom-right to the top-left corner a dot is continuously shifted between the two adjacent gates. Corner overlays show schematic band diagrams for different applied voltages. d) Similar measurement for the mirror-symmetric experiment with gates 4 and 5. While the values of conductance differ between panels c and d due to different p-n junction barriers formed near the left and right contacts, the conductance patterns are remarkably similar (see main text).
The local gates now allow us to probe electronic behavior on finer spatial scales. By electron-doping the nanotube locally with a single gate and hole-doping the rest of it with all other gates, we form a smaller quantum dot localized above this gate. Accordingly, with five independent gates we can form, in principle, dots at five different locations along the nanotube, whose characteristics reflect the spatial dependence of the nanotube electronic properties. Figure 4.1b shows the corresponding five conductance traces, as a function of the individual gate voltages. Clearly, single quantum dots are formed at all positions. Their Coulomb blockade oscillations have periodicities of $\Delta V_g=280\pm 10\text{mV}$, indicating that the dots are well localized above a single gate (as we discuss below in section 4.4). The corresponding charging energy of these dots is $E_C=59\pm 8\text{meV}$. Moreover, all traces exhibit a single periodicity, showing that the dots are clean. The Coulomb peak heights, however, vary between dots at different locations, hinting at possible position-dependence in the electronic properties along the nanotube.

A more complete picture of the spatial dependence is obtained by using pairs of gates to continuously move a quantum dot along the nanotube. Figure 4.1c shows the conductance measured as a function of the voltages on gates 1 and 2, while all other gates are negatively biased. On the bottom left corner, both gates dope the nanotube with holes and no dot is formed. When gate 1 (gate 2) is positively biased, along the horizontal (vertical) axis, a dot forms above this gate. Biasing both gates together (upper-right corner) extends the dot above both gates. Thus, going from the bottom-right to the top-left of this figure the quantum dot shifts from one gate to its neighbor. In this measurement, the Coulomb charging peaks appear as charging lines, separating different charge states of the quantum dot. Their local slope corresponds to the relative capacitance of the dot to the two gates, and reflects the position of the centre-of-mass of the electronic charge. Notably, the slopes of all charging lines, down to that of the first electron, evolve smoothly and monotonically during the shift, reflecting the smooth transfer of the electronic confinement from site to site. The data, however, contain unexpected features: the charging lines exhibit periodic stripe modulation of the peak
heights and a band-like region where the conductance is suppressed (arrows, Figure 4.1c). These features may indicate the existence of disorder forming random barriers or dots. Below we show, however, that these features arise from intrinsic electrostatics, and not disorder.

A clear way to determine whether the observed features are due to disorder is to perform the mirror-symmetric version of the experiment depicted in Figure 4.1c. This measurement, shown in Figure 4.1d, is done with the opposite gates, 4 and 5, over a voltage range identical to that in Figure 4.1c. Comparing these mirror-symmetric measurements reveals a striking similarity: charging line slopes, positions, and spacing are all identical. Furthermore, the peak modulations and the conductance suppression are reproduced at the same gate voltages. The remarkable implication is that all the observed features are not the result of a random disorder potential, but rather arise from the intrinsic electrostatics of the device. These features, discussed below section 4.3, are due to gating of nanotube segments that are beyond the dot, such as Fabry-Perot-like oscillations in the hole-doped “nanotube leads”. In contrast to peak positions, which are identical in both experiments, peak heights are different. While peak positions are sensitive only to electrostatics, their heights also depend on the resistance to the metal contacts, which might vary for different contacts. We find, however, that this asymmetry does not originate in contact resistance but instead is also electrostatic in nature, coming from an inequivalence of the p-n junction barriers near the source and drain contacts, due to a slight lithographic misalignment (~15nm) of the gates toward the drain contact. The observation of nearly-perfect mirror symmetry thus demonstrates that, for electrons above the outer gates, electrostatics rather than random disorder determines the local electronic structure.

To check the effects of disorder in the bulk of the suspended nanotube, we generalize the above measurements to all pairs of gates in the device. Figure 4.2 shows a matrix of two-gate conductance measurements, whose columns and rows correspond to the gates scanned on the horizontal and vertical axes of each panel. In all panels the gate voltage ranges are identical, with all other gates maintaining a constant hole-doping voltage. On the main diagonal of this matrix, the scanned gates are
Figure 4.2: Symmetry of transport characteristics in a five-gated device. A matrix of conductance measurements where in each entry a different pair of gate voltages, $V_{gi}$ and $V_{gj}$, is scanned. The gate scanned along the horizontal (vertical) axis is indicated in the column (row) title, and the voltage scan range in all panels is identical to that in Figure 4.1. The detailed conductance features in all panels show symmetry with respect to mirror reflection around the nanotube centre (dashed black line). Colourmaps for all measurements are shown in the lower left corner of each entry, over the range $0 \ldots 80\text{nS}$ for all scans.

Clearly, all the scans along this diagonal feature a continuous bending of the charging lines, indicating the smooth movement of charge from any gate to nearest neighbors (as in Figure 4.1c and d).
Figure 4.3: Symmetry of transport characteristics at the edges of a 16-gated device. A very small-bandgap nanotube is suspended over \( \sim 2.6 \mu m \) above 16 gates of 75nm width and 75nm spacing, doped by electrons over the PdAu contacts, lying over 200nm above the PdAu gates, and measured at \( T = 4 K \). a) Conductance, G (colormap), as a function of two voltages, \( V_{g1,2} \) applied to gates 1 and 2, and \( V_{g3,4} \) applied to gates 3 and 4 (see top SEM image of the device, scale bar 1 \( \mu m \)). Top right corner: electron wire. Lower left corner: hole quantum dot above four gates. b) The same measurement as a function of \( V_{g16,15} \) and \( V_{g14,13} \), on the opposite end of the device.

While many features are observed in these experiments, the remarkable observation is that over the entire matrix all these features are symmetric among experiments with mirror symmetry around the nanotube centre (dashed black line). We conclude that, to the spatial resolution fixed by our gates and to the energy scale set by the temperature, disorder is playing a negligible role in determining the potential landscape along the entire device.

We have also observed symmetry of conductance scans in longer devices with up to 16 gates. One such device is shown in Figure 4.3, with a small-bandgap nanotube on PdAu contacts, suspended across \( \sim 2.6 \mu m \) at a height of 200nm above the PdAu gates of 75nm width and spacing. Conduc-
tance measurements are performed at $T = 4K$, and the nanotube has a measured bandgap of 8meV, obtained from finite bias measurements. In this device, the contacts dope the nanotube segments above them with electrons, and quantum dots are formed by applying negative voltages and forming p-n junction barriers that confine holes. In these measurements, the 16 gate voltages are chained in nearest-neighbour pairs, for a total of 8 effective gate voltages. Figure 4.3 shows a conductance scan at $T = 4K$ as a function of $V_1 = V_{g1,2}$ and $V_2 = V_{g3,4}$, and the symmetric scan for $V_7 = V_{g3,14}$ and $V_8 = V_{g5,16}$. Similar to Figure 4.1c and d, the color scale, which is fixed by the conductance and thus by the asymmetric contact resistance, has been modified to highlight the symmetry of features in the 2D scans. Nearly identical features are observed between the two scans. Here, unlike the above, a difference in the work functions of the source and drain contacts leads to a small position shift of the features observed on either edge of the device (see section 4.4). Interestingly, the extremely small bandgap of this nanotube means quantum dots are well-coupled to the leads, allowing us to observe the fourfold filling of nanotube electron shells in the charging line structure (see A).

**Figure 4.4:** Smoothly shifting electrons along a 10-gated device. A small-band gap nanotube on PdAu contacts is suspended over $\sim 2.3\mu m$, at a height of $\sim 400nm$ above the 10 PdAu gates of 120nm width and 80nm spacing. Conductance is measured as a function of nearest-neighbor voltages. Vertical pairs of color maps correspond to symmetry pairs (gates 1 and 2 above gates 9 and 10, etc.).
Smooth shifting of the confinement potential is also observed in longer devices. Figure 4.4 shows conductance measurements of a ten-gate device with a nanotube \( \sim 2.3 \mu m \) long suspended \( \sim 400 \text{nm} \) above the gates, as a function of nearest-neighbor gates, analogous to the lowest diagonal of Figure 4.2. The contacts dope the nanotube with electrons, resulting in the formation of hole quantum dots at negative voltages. The quantum dots are smoothly moved by the gate voltages at every point along the nanotube. Symmetry is observed for several features, including the positions of the hole dot charging lines and the electron band edge, as well as the number of charging lines and some modulations of the peak conductances.

4.2 Pristine Nanotubes At Dilution Temperatures

The measurements shown above had no observable indications of disorder at the energy scale of the measurement \( (T=4K) \). A natural question is whether at lower temperatures, smaller disorder scales would become observable, and we address this with measurements at dilution refrigerator temperatures. Figure 4.5 shows the conductance of a five-gated device, different than the one shown in the main text, measured in a dilution refrigerator with a base temperature of \( T=7\text{mK} \) (extracted electron temperature is \( T \approx 80\text{mK} \)). At these temperatures, the device is expected to be sensitive to smaller magnitude disorder; we show in the following that our observations on device cleanliness hold to these low temperatures. For this experiment, the two right gates are biased together along the horizontal axis, \( V_{g1} = V_{g2} = V_R \), the two left gates are biased together along the vertical axis, \( V_{g4} = V_{g5} = V_L \), and the center gate is biased with the average voltage, \( V_{g3} = (V_L + V_R)/2 \).

Overall, this device demonstrates almost perfectly clean behavior. In the upper right (lower left) corners, we observe the creation of a five-gate electron (hole) dot (see corresponding schematics) with a single Coulomb oscillation periodicity. In the lower right and upper left corners, p-n junctions are formed at the center of the suspended nanotube, leading to the creation of a hole-electron
Figure 4.5: A second five-gated small bandgap nanotube device measured at dilution refrigerator temperatures. The conductance, $G$, on a logarithmic scale, measured as a function of right and left gate voltages, $V_L$ and $V_R$. The right gate voltage is applied on the two right gates, $V_{g1} = V_{g2} = V_R$, the left gate voltage is applied on the two left gates, $V_{g4} = V_{g5} = V_L$ and on the center gate we apply the averaged value $V_{g3} = (V_L + V_R)/2$. The insets show schematic band diagrams corresponding to the four quadrants of the measurement: bottom left - a hole dot over all gates; top right - an electron dot above all gates; bottom right - electron/hole double dot; top left - hole/electron double dot.

and electron-hole double quantum dots respectively (see corresponding schematics). Notably, every vertical charging line in the lower right corner, corresponding to an electron localized on the right side of the device, evolves smoothly into a horizontal charging line in the upper left corner, corresponding to an electron localized on the left side. As explained in the main text this smooth evolution shows that individual electrons are smoothly shuttled from the right to the left side of the device without apparent effects of disorder. An almost perfectly symmetric behavior is observed for the hole charging lines that evolve smoothly from vertical in the top left corner to horizontal in
the bottom right corner. One deviation from the perfect behavior is observed for the first hole line, which is vertical even on the bottom part of the gate-gate scan and does not bend like the others. This means that this hole gets stuck on the right side and is not shuttled to the left side by the gates. By checking the relative coupling of this feature to the individual gates (not shown) we see that it almost exclusively gated by gate 5 and none of the others, demonstrating that this hole is localized between gate 5 and the contact. Such behavior could be due to a highly localized potential dip near the contact that binds only one carrier. Importantly, all the holes after this first localized one show the normal extended behavior and possess nearly perfect symmetry to the electrons.

Looking carefully on the data for the first few electrons and holes (excluding the first localized hole), we can observe small wiggles of the charging lines. These wiggles are an order of magnitude smaller than those observed in the best ultra-clean double-dot devices made to date, demonstrating that the underlying disorder potential in our devices is much smaller. As was clearly demonstrated, a potential hump or dip act differently on electron and holes, leading to different charging line structures for the two carrier types. The fact that we observe very similar wiggles for electrons and holes therefore emphasizes that disorder on the length scale of the gate spacing is probably not the mechanism leading to the observed wiggles. Instead, the effect must operate the same way on electrons and holes. One candidate is the attraction of the carriers in the nanotube to their image charges formed at the metallic leads, which leads to attractive potentials at the suspended nanotube edges for both electrons and holes, thereby leading to a double-dot-like effective potential which could explain the small wiggles. Another plausible mechanism is strong interactions between the carriers, which are predicted to lead to real-space separation of charge carriers and a similar modulation of the charging lines.
4.3 Extracting and Explaining Electrostatics

To establish our system as a controllable laboratory for 1D experiments, we must demonstrate a quantitative understanding of its electrostatics that permits 1D potential design. Naively, the simplest way to create a potential $\phi(x)$, where $x$ is the position along the nanotube, is to have gates as close as possible to the nanotube such that their gating is local and their voltages directly determine $\phi(x)$. However, close proximity to metallic gates screens the interactions between electrons in the nanotube, thus destroying this salient feature. We therefore intentionally choose to distance the gates from the nanotube, the price being that gating becomes non-local, and a gate influences not only the nanotube segment above it but also segments above other gates. Knowing what gate voltages are required to produce a certain $\phi(x)$ thus necessitates quantitative knowledge of the non-local capacitive coupling to the nanotube.

The influence of a specific gate on the nanotube is fully captured by a capacitance distribution function $C_i(x) = e n(x)/V_{gi}$, where $e$ is the electron charge, $i$ is the gate index and $n(x)$ is the charge distribution along the nanotube induced by a gate voltage $V_{gi}$, under the assumption that the nanotube is a perfect metallic conductor. We are interested in a discretized version of this function where the nanotube is partitioned to $N$ segments of equal length $l$ ($N$ is the number of gates) each segment being positioned above a corresponding gate. This partitioning reflects the “effective resolution” with which we can define the potentials with the gates. The capacitance of a gate $i$ to a segment $j$ is then given by $C_{ij} = \int_{\text{segment}-j} C_i(x) \approx C_i(x_j) \cdot l$, where $x_j$ is the center coordinate of the segment. Experimentally we can extract a closely related quantity, by measuring the charge response of a quantum dot localized at position $j$ to the voltage on gate $i$. The latter amounts to $\tilde{C}_{ij} = \int_{\text{dot}-j} C_i(x) \approx C_i(x_j) \cdot l_{\text{dot}-j}$, where $l_{\text{dot}-j}$ is the length of the dot formed at position $j$.

Clearly, the $C_{ij}$’s depend on the shape of the quantum dots. However, if we take only their ratios that measure the response of the same dot to two different gates, the details of the dot cancel out.
and we remain with the ratios of the quantities that we are seeking: $C_{ij}/C_{ii} = \tilde{C}_{ij}/\tilde{C}_{ii}$.

The information about how a gate at position $i$ influences the nanotube at a position above gate $j$ is embedded in the matrix of two-gate conductance measurements, such as that shown in Figure 4.2. At the bottom right corner of each $(i, j)$ panel a quantum dot is localized over gate $i$ and gated by gate $j$. The dot, acting as a local charge detector, allows us to directly measure the capacitance ratio from the slopes of the charging lines (see section B.2 where this is derived).

The ratios above give only relative capacitances, and thus do not provide the full information needed to determine all the absolute capacitance elements in the $C_{ij}$ matrix. To get the missing information we complement these data with measurements of the integrated capacitance of individual gates. To obtain these we form a large quantum dot extended over the entire suspended nanotube. We first measure the total capacitance of this dot to all five gates chained together. This quantity is directly extracted from the gate periodicity of the Coulomb oscillations. For the five-gate device in Figure 4.1, these are shown in Figure 4.1a. This capacitance gives the sum of all the $C_{ij}$ matrix elements: $C_{\text{total}} = \sum_{ij} C_{ij}$.

We then measure the relative contribution of each of the gates to this capacitance, giving the sum of one row in this matrix, $C_i/C_{\text{total}} = \sum_j C_{ij}/C_{\text{total}}$. We extract the sum by comparing the width of a Coulomb peak of the large dot when only one gate is scanned vs. the width of this peak when all gates are scanned. Together, all these quantities give us the full capacitance matrix without any free parameters.

The electrostatic simulations are performed with the finite-element calculation package COMSOL. For this, we use the real device dimensions extracted from SEM images to model the geometry. This includes the trapezoidal cross-section of the contacts, arising due to gradual closing of the e-beam resist window during the thick evaporation. The electrode and substrate geometry that go into the calculation are shown, for the five-gate device, in Figure 4.6. We model the nanotube as a metallic cylinder resting on the contacts and suspended over the gates. To extract the capacitance...
distribution of gate $i$ we set the voltage on this gate to $V_{gi}$ while keeping the other gates, the contacts and the nanotube grounded. We then calculate directly the capacitance distribution of this gate, $C_i(x) = en(x)/V_{gi}$.

The measured capacitance matrix and the calculated capacitance distribution functions for the individual gates of the five-gate device are shown in Figure 4.6. The simulations match the experimental values with no free parameters. These results reveal that the capacitance distributions of gates 2-4 are almost identical, demonstrating that the electrostatics in the “bulk” of the sample is translationally invariant. The edge gates (1 and 5) show reduced coupling due to screening by the contacts, and differ in their peak coupling due to the gate-contact misalignment asymmetry noted above and fully reproduced by the calculations. These results are not sensitive to mechanical displacement of the nanotube, which for this device, with 1V applied to all five gates, is estimated to be ~5nm out of
the 130 nm gate-nanotube distance, based on a comparison of the mechanical and electrostatic forces acting on the nanotube.

---

**Figure 4.7:** The electrostatic origin of the observed features in the two-gate conductance scans. 

- **a)** The main panel (duplicating Figure 4.1c of the main text) shows conductance, $G$, as a function of $V_{g1}$ and $V_{g2}$ while $V_{g3-5} = -0.8 \, V$. The dashed black lines and dashed white lines correspond to the secondary conductance features observed in this scan. These lines are drawn with slopes taken directly from the measured capacitances (see text and panels below). 
- **b)** The electrostatic potential along the nanotube length, $\phi(x)$, calculated for the points indicated in panel a, with circle colors corresponding to line colors. The contact and gate work function differences are included as described in the text. The device schematic is to scale in both dimensions. 
- **c)** Schematics of the electron (red) and hole (blue) population along the nanotube corresponding to the blue and red points in panel a. The top traces show the capacitance distributions of gates 1 and 2. The colored areas under the curves give the capacitances between these gates and the left nanotube hole “lead” to the dot. The ratio between these capacitance gives the slope of the dashed black lines in panel a. 
- **d)** Schematics of the electron and hole population along the nanotube corresponding to the green point in panel a. Here a dot forms only above gate 2 and the nanotube is in the gap over gate one. This leads to an extended right barrier that yields the suppression along the dashed white line in panel a. Its slope in the voltage-voltage plane is determined by the ration of the capacitances of the barrier region to gates 1 and 2.

We can now use this electrostatic understanding to explain the features observed in Figure 4.1c and d. In Figure 4.7a we reproduce Figure 4.1c and highlight the secondary conductance features that are observed on top of the Coulomb blockade features described in section 4.1. The first set of features, apparent in this scan and its symmetric partner (Figure 4.1d), are stripes of conductance modulation that are marked by dashed black lines. These stripes cross through the Coulomb charg-
ing lines and modulate their peak heights. The electronic configuration which corresponds to this region in the gate-gate diagram consists of an electron dot formed over gates 1 and 2 (red in device schematics) while above gates 3 to 5 the nanotube is populated with holes (blue in device schematics). The hole population in this segment is continuously connected to the holes above the left contact, forming a continuous “hole wire” that acts as a “nanotube lead” for the electronic quantum dot. We can confirm this picture by calculating the electrostatic potential induced by the gate voltages as shown in Figure 4.7b, where the nanotube lead and the electron dot correspond to the potential well and hill above their respective gates. By calculating the potentials at two points along the modulation stripe (indicated in the figure by red and blue circles), we see that the potential well corresponding to the nanotube lead remains identical while only the electron potential hill has changed. This indicates that going along a stripe preserves the charge density in the nanotube lead, whereas going perpendicular to it changes this charge density. Thus, the origin of the observed striped modulations of the Coulomb peak heights is reproducible Fabry-Perot-like modulation of the conductance of the “nanotube lead” (the nature of these conductance modulations is discussed further in the last paragraph of this section).

We now proceed to confirm this picture with a calculation of the stripe slopes. Although gates 1 and 2 are far from the nanotube lead, they still gate it by an amount that can be quantitatively determined from the capacitance distribution functions which were measured and calculated (Figure 4.6). The relative capacitance of the two gates to the nanotube lead amounts to the ratio of the areas under the capacitance distribution curves of these two gates integrated over the length of the nanotube lead, as shown in Figure 4.7c. The edge of the hole occupation is determined from the point where the electrostatic potential \( \phi(x) \) crosses zero, since this is where the p-n junction barrier will be centered. In the discretized version these capacitances are given by the elements of the capacitance matrix, \( C_{i}^{\text{p-lead}} = C_{i5} + C_{i4} + C_{i5} \) and \( C_{2}^{\text{p-lead}} = C_{23} + C_{24} + C_{25} \), all of which we measure directly. The dashed black lines in Figure 4.7a are drawn with a slope \( C_{1}^{\text{p-lead}} / C_{2}^{\text{p-lead}} \) taken from these mea-
surements, showing a good fit to the observed conductance modulation slopes.

Another clear feature observed in the two, mirror-symmetric, gate-gate scans is a band of suppressed conductance (dashed white lines, Figure 4.7a). We calculate the electrostatic potential in the middle of this suppressed band along the same line of fixed nanotube lead gating, shown in green in Figure 4.7b. From the electrostatic calculations we can identify that this feature corresponds to having the electron dot confined over only a single gate (gate 2) and having the nanotube bandgap pinned between gate 1 and the right contact. In this case, the right tunnel barrier of the electron dot is a p-n junction formed above an edge gate (gate 1 or gate 5), which has a longer depletion length than when formed above a center gate because the nearby contact is grounded (as opposed to the negatively-biased gates). The underlying origin of the longer barrier is the shallower slope of \( \phi(x) \) where intersects zero on the right. The existence of this extended edge barrier explains the observed conductance suppression. The slope of the dashed white line in the conductance map that follows the middle of the observed suppression band is \( \Delta V_2/\Delta V_1 \approx 2.5 \), corresponding reasonably well to that calculated with the capacitance distributions, \( C_{pn\text{-barrier}}^{1}/C_{pn\text{-barrier}}^{2} \approx C_{11}/C_{22} = 2.3 \). We note here that the actual length of the p-n junction will depend on electrostatic and quantum effects, and determining it requires a full solution of the Schrodinger-Poisson self-consistent equations. However, we are interested only in the positions of features in the voltage-voltage plane, which depend mostly on the position of the p-n junction (and less so its width), and this is captured well by our analysis.

Continuing to the other side of the suppression feature, the calculated electrostatic potential (gray line, Figure 4.7b) shows that the quantum dot remains over gate 2, but above gate 1 there is now a hole population, showing that the right “nanotube lead” has extended over gate 1. As a result, the right p-n junction is now shorter and the conductance is higher, comparable to that on the other side of the suppression feature.

We now comment on the nature of the hole-doped “nanotube leads” to the electronic dot. In the
main text we mentioned that when the nanotube is populated entirely with holes it behaves like a “nanotube wire”. In this regime we measure only weak gate modulation of the conductance, which phenomenologically resembles the measurements of nanotubes in the Fabry-Perot regime\textsuperscript{107}. In that regime the barriers between the nanotube and the contacts are highly transparent and the average conductance is comparable to $4e^2/h$. In our case, on the other hand, the measured conductance is significantly smaller than the quantum conductance and thus one expects the nanotube to behave as a quantum dot rather than a Fabry-Perot cavity. This dot, however, is unusual since its charging energy is strongly suppressed. This suppression results from the fact that the nanotube sits directly over the contacts and thus has an extremely large capacitance to them. At such short distances the geometrical capacitance to the contact, $C_{\text{geometric}}$, is much larger than the quantum capacitance of the nanotube segment above it, $C_{\text{quantum}}$, so that the latter dominates the total source capacitance, which in our case is $C_{\text{source}} = [(C_{\text{geometric}})^{-1} + (C_{\text{quantum}})^{-1}]^{-1} \approx 100aF$. The resulting charging energy of the nanotube leads, $U = e^2/(C_{\text{source}} + C_{\text{drain}} + \sum C_{\text{gates}})$, being dominated by $C_{\text{source}}$, thus roughly equals the level spacing of the nanotube above the contacts. In this respect, the system is similar to the Fabry-Perot cavity. The large suppression of the charging energy as compared to the quantum dots on the suspended part of the nanotube explains why at $T=4K$ the hole-doped nanotube behaves similarly to a Fabry-Perot cavity and shows weak Coulomb oscillations. This is the regime of the “nanotube wire” in our measurements, where the charging energy, $\sim$1mV, is small enough that the measurement temperature results in only weak gate-dependence of the conductance, and the series resistance to the metallic contacts nonetheless gives a small overall conductance. For a given overlap length of nanotube and contacts, when the temperature is low enough the leads would eventually show Coulomb blockade physics. For the device geometry shown in Figure 3.21a such Coulomb physics of the leads is indeed seen at dilution temperatures (see section 4.2). However, by making the overlap with the contacts long enough, and hence suppressing further the charging energy, it should be possible to make the leads behave as “wires” down to the lowest temperatures in
our measurements.

4.4 Quantifying Disorder

So far, we have shown how knowledge of the device electrostatics allows us to define the nanotube potential with gating. We can also use this knowledge to estimate the nanotube potential in the absence of gating, which corresponds to the uncontrolled disorder potential fluctuations in our device. This therefore provides an experimental estimate for the strength of disorder on length scales set by the gates’ resolution. Our goal is thus to measure the uncontrolled potential modulations, $\delta \phi(x)$, that exist in the nanotube in the absence of gating.

If the nanotube had the same work function as the gate metal, it would be un-gated when all the gates are un-biased with respect to the nanotube ($V_{gi; i=1...5} = 0$). However, since these workfunctions are generally different, the absence of electrochemical bias on the gates ($V_{gi; i=1...5} = 0$) actually means that there is a non-zero electrostatic potential difference between the gates and the nanotube. This difference, termed the “contact potential”, amounts to $\delta W_{\text{gate-nanotube}} = W_{\text{gate}} - W_{\text{nanotube}}$, and it gates the nanotube. To null this gating one must therefore apply a canceling electrochemical bias to the gates, $V_{gi; i=1...5} = -\delta W_{\text{gate-nanotube}}$. It is important to note that the contacts, which are by definition electrochemically shorted to the nanotube, produce a similar gating effect due to the difference between their workfunction and that of the nanotube, $\delta W_{\text{contact-nanotube}} = W_{\text{contact}} - W_{\text{nanotube}}$. This is the reason for the large hole doping of the nanotube segments that lie on top of the contacts, mentioned in the main text. Combining the above understanding with finite element simulations, which were shown in the previous section to describe our system quantitatively well, we can determine the bare electrostatic potential produced along the nanotube for any
combination of gate voltages:

\[
\phi(x) = \sum_i \alpha_i(x) \cdot (V_{gi} + \delta W_{\text{gate-nanotube}}) + (\alpha_S(x) + \alpha_D(x)) \cdot \delta W_{\text{contact-nanotube}} \tag{4.1}
\]

Here \(\alpha_i(x), \alpha_S(x)\) and \(\alpha_D(x)\) are unitless functions, determined from the finite elements simulations, that give the potential along the nanotube per unit of voltage applied on gate \(i\), the source and the drain respectively. The a priori unknown work function differences, \(\delta W_{\text{gate-nanotube}}\) and \(\delta W_{\text{contact-nanotube}}\), are the two free parameters of this equation that are determined from the experiments (see below).

![Figure 4.8: Calculated bare electrostatic potential profiles along the nanotube for three different gating configurations. a) Measured conductance trace for a dot formed above gate 1, equivalent to the top trace in Figure 4.1b. In this measurement \(V_{g1}\) is swept while the rest of the gates have a fixed potential \(V_{g2...5} = -0.8V\) that dope the segment above them with holes. b) Three calculated potential profiles calculated using finite elements and Equation 4.1, corresponding to three different transport regimes (the corresponding \(V_{g1}'s\) are shown in panel a as dots with similar colors). Workfunction differences are included (see text).](image)

To elucidate the relation between the potential profiles given by Equation 4.1 and the measured transport we take as an example the top conductance trace from Figure 4.1 that corresponds to the formation of a quantum dot above gate 1 (reproduced in Figure 4.8). In this scan \(V_{g1}\) is swept while all the other gates are kept at fixed voltages \(V_{g2...5} = -0.8V\). The calculated potential profiles, \(\phi(x)\), that correspond to three gate voltages along this scan (circles in Figure 4.8) are shown in Figure 4.8.
(work function differences are included; see below). In each of these plots the $\phi(x)$ in the suspended segment corresponds to the center of the nanotube bandgap as a function of position. Wherever it crosses from below to above the Fermi energy, $E_F$ (dashed horizontal line), the local occupation changes from holes to electrons. At low $V_{g1}$ the corresponding $\phi(x)$ (blue trace) is at all positions below $E_F$, implying that the nanotube is populated with holes over its entire length, thus forming a continuous “nanotube wire”. At high $V_{g1}$, the corresponding $\phi(x)$ (red trace) exceeds $E_F$ above gate 1, crossing it at two points. At these points p-n junctions form, confining the electrons above gate 1 to a quantum dot. In between these regimes there is a $\phi(x)$ that exactly reaches $E_F$ above gate 1 (purple trace). For this potential the center of the nanotube bandgap is at $E_F$ above gate 1, forming a single long barrier above this gate. This point corresponds to the center of the non-conducting regime in the transport (purple point Figure 4.8) where the conductance is maximally suppressed.

Looking at the corresponding transport traces obtained by scanning the other local gates (Figure 4.9a, reproduced from Figure 4.1) we see that the center of the “gap” appears at very different gate voltages for the different gate positions. Figure 4.9b plots the gap-center gate-voltage as a function of the gate position, showing that this value changes by $\Delta V_g \approx 325\text{mV}$ from the side gate to the center gate. This seemingly large potential modulation is in fact a direct result of the position dependence of the device electrostatics combined with the finite workfunction difference between the gates/contacts and the nanotube. Both these effects should be fully captured by Equation 4.1. Thus, if this equation is accurate, in the absence of disorder we should be able to reproduce the position of these five gap centers, with just two parameters, the metal workfunctions of the gates and contacts. This is demonstrated in Figure 4.9c, where we show the five $\phi(x)$’s that correspond to the gap centers in the five different conductance traces, calculated with $\delta W_{gate-nanotube} = -40\text{mV}$ and $\delta W_{contact-nanotube} = -170\text{mV}$.

These work function values are consistent with published values for gold, palladium, and carbon nanotubes. Gold work function values range between 4.8-5.5 eV, and palladium between 5.1-5.6 eV.
Figure 4.9: Extracting an upper bound on the disorder potential from the measured transport. a) The single-gate conductance scans reproduced from Figure 4.1b of the main text with the bandgap centers marked by colored arrows corresponding to the colored points in panel b and colored curves in panel c. b) The gap-center voltage as a function of the gate index, extracted from the graphs in panel a. c) The bare electrostatic potentials along the nanotube, $\phi(x)$, calculated using Equation 4.1 for the gate voltage configurations that correspond to the positions indicated by the respective arrows in panel a, with $\delta W_{\text{gate-nanotube}} = -40 \text{ mV}$ and $\delta W_{\text{contact-nanotube}} = -170 \text{ mV}$. Although the gate voltage at the gap center varies between dots at different locations by as much as ~$325 \text{ mV}$, when we take into account the electrostatics of the device through Equation 4.1 we see that within $\delta \phi \approx \pm 5 \text{ mV}$ all the potential profiles corresponding to the center of the gap are at the Fermi energy, giving an upper bound for the residual bare disorder potential fluctuations.

We do not have values for a palladium-gold alloy, used for our gates, but can assume that it is intermediate to the above values. Reported carbon nanotubes work functions are in the range $4.7$-$5.1 \text{ eV}$. The resulting work function differences range from $0.1$ to $0.8 \text{ eV}$ for gold-nanotube and $0$ to $0.9 \text{ eV}$ for palladium-nanotube. Our values, of $0.17 \text{ eV}$ for gold-nanotube and $0.04 \text{ eV}$ for PdAu-nanotube, are within this range of literature values. We want to emphasize, though, that the work function of metals is extremely sensitive to their surface properties, and can change significantly with changing the adsorbates on the electrodes. In the context of nanotubes it was shown in the past that adsorbates on the contacts dramatically affect the electrical contact to nanotubes due to work function changes, even changing them from p-type to n-type and vice-versa. We and other groups also observe this – as a function of the pump down time of the vacuum space of the cryostat in which the
device is mounted, prior to cooling it down, we consistently observe that devices change from n-type to p-type with increasing pump down time, demonstrating that the metal work function strongly depends on its surface adsorbates. This means that direct comparison to the literature values of ideal work functions might not reflect the physics in our measurements very well.

We can clearly see that in all cases the potential reaches $E_F$ above the corresponding gate with an accuracy of $\delta \phi \approx \pm 5\text{mV}$. These small fluctuations compared to those observed in Figure 4.9b ($\Delta V_g \approx 325\text{mV}$) show that most of the effect is a consequence of the device electrostatics, and once it is known quantitatively it can be taken into account and nulled out. The remaining small fluctuations give us an upper bound of $\sim 5\text{mV}$ on the magnitude of the bare potential disorder on the length scale set by the gate width. Using the lever-arm of the local gates ($\alpha \approx 0.3$) this is translated to $\sim 17\text{mV}$ on a local gate, and by comparing this to the measured single-gate Coulomb blockade periodicity ($\Delta V_g \approx 315\text{mV}$) necessary to introduce an electron charge above a gate we can obtain an upper bound on the local charge disorder of $\sim 5 \times 10^{-2}e$ on 100nm length scales, a small fraction of a single electron charge. Alternatively, we can consider the induced fluctuations in the self-consistent disorder potential, where screening will reduce the bare disorder potential seen by electrons. By factoring in the ratio between the geometric capacitance ($\sim 4aF/\mu\text{m}$) and the quantum capacitance ($\sim 400aF/\mu\text{m}$) of the nanotube, we estimate the self-consistent disorder potential at $\sim 50\mu\text{V}$. We note that this is a strict upper bound, since the potential fluctuations that we consider include all the errors in the measurements and calculations. The actual disorder is most likely significantly smaller. We also note that potential fluctuations on smaller length scales, which are too weak to form barriers for electron transport at the temperature of our measurements, would not be observed here, although short-length scale disorder such as adatoms can also generate long-length scale disorder that these measurements rule out.

These low values of disorder and the correspondence of the measured and calculated electrostatics show that potential profiles can be accurately designed, as we discuss in the next section.
4.5 Potential Design

In this section we demonstrate that with knowledge of the electrostatic coupling of the gates to the nanotube, we can design potential profiles along the nanotube with a spatial resolution given by the gate pitch and not smeared by the separation between the nanotube and the gates. As explained in section 4.3, by distancing the nanotube from the gates, we preserve electron-electron interactions. But at the same time we also spatially smear the effect of individual gates on the nanotube. While a close gate controls the potential in the nanotube along a length comparable to its width, a distant gate affects a longer section amounting to the convolution of the gate width and its distance to the nanotube. This distance therefore reduces the effective resolution with which we can design electrostatic potentials. However, by using our knowledge of the non-local gate coupling, we can deconvolve this spatial smearing and define potential features whose sharpness is determined by the gate pitch alone. Such deconvolution works as long as the nanotube is not too far from the gates compared to the gate separation.

To define the potential (or the charge) on the nanotube with gate pitch resolution means that if we partition the nanotube into $N$ segments of equal length, where $N$ is the number of gates, we can define the potential (or charge) in each one of these segments independently. However, the charge on the $i^{th}$ segment in the nanotube, $q_i$, due to a voltage on gate $j$, $V_{gj}$, is given by $q_i = C_{ij} V_{gj}$, where $C_{ij}$ is the capacitance coupling matrix element. Thus, the above equation shows that a gate does not only affect the local segment above it but also neighboring segments, reducing the effective resolution. To define the charge on each segment independently, we invert the equation, $V_{gj} = C_{ji}^{-1} q_i$, to obtain the linear combination of gate voltages $V_{gj}$ that is needed to control the charge in only a single segment of the nanotube, $q_i$. This inversion amounts to a discretized deconvolution of the capacitive smearing. Figure 4.10 illustrates how this deconvolution works for the dimensions of our devices (a nanotube-gate distance of 130nm and a gate pitch of 150nm), using a calculation of the po-
Figure 4.10: Recovering the gate resolution with potential design. a) The voltages applied on the gates and b) the corresponding calculated potential at the nanotube, $\phi(x)$, for the dimensions of our device (a nanotube-gate distance of 130nm and a gate pitch of 150nm). In this example a voltage is applied only on a single gate and the width of the potential feature along the nanotube is roughly the sum of the gate pitch and the nanotube-gate distance. c) Linear combination of voltages for producing localized potential above the central gate, obtained by inverting the measured capacitance matrix. d) The corresponding calculated potential along the nanotube showing that the non-locality of the gate coupling can be effectively deconvolved.

Potential at the nanotube, $\phi(x)$, with Equation 4.1. Applying a voltage on a single gate (Figure 4.10a) leads to a potential along the nanotube spread out over ~325nm (Figure 4.10b), roughly the sum of the gate pitch and the nanotube-gate distance. On the other hand, if we use instead the linear combination of gate voltages found using the inverse capacitance matrix (Figure 4.10c), we obtain a potential that is ~160nm wide (Figure 4.10d), comparable to the gate pitch. We note that the deconvolution becomes exponentially harder when the nanotube-gate distance becomes much larger than the gate pitch, $\eta \equiv d_{NT-gate}/d_{gate-gate}$, since in this regime the voltages necessary for producing the desired potentials increase exponentially in $\eta$, and any experimental error in determining $C_{ij}$ is exponentially amplified. However, as long as $\eta$ is not very large, as in our experiments where $\eta \sim 1$, the above deconvolution procedure works well.
The electrostatic potential at the nanotube may be designed by approximating the under-determined system of equations

\[ V_g = \alpha^{-1}(x) \phi_{\text{target}}(x), \]

where \( \alpha(x) \) is the set of dimensionless functions defined in Equation 4.1 and \( V_g \) is the set of gate voltages best approximating the desired potential \( \phi_{\text{target}}(x) \) according to a best-fit algorithm (see main text). a) For a 16-gated nanotube suspended 200nm above the gates of 150nm pitch, we choose \( \phi_{\text{target}}(x) \) to be a Gaussian with \( \sigma = 150 \text{nm} \) positioned near the center of the tube. The target potential and the actual potential, given by \( \alpha(x) V_g = \tilde{\phi}_{\text{target}}(x) \), where \( \tilde{\phi}_{\text{target}}(x) \) is the actual potential induced by the best-fit gate voltages \( V_g \). The two overlap nearly perfectly. b) The gate voltages \( V_g \) calculated in (a). c) The potentials \( \tilde{\phi}_{\text{target}}(x) \) for Gaussian target potentials at the same position as in (a), for different Gaussian width parameters \( \sigma \). The induced potentials reproduce the Gaussian for \( \sigma \geq 150 \text{nm} \), indicated by the red dotted line, below which oscillations in the wings of the potential are visible. d) Red: sum of squared errors, \( \sum_x (\phi_{\text{target}}(x) - \tilde{\phi}_{\text{target}}(x))^2 \), for Gaussians with \( \sigma = 150 \text{nm} \), at different positions along the nanotube. The error is negligibly small for Gaussians centered within the outermost gates of the device (red dotted lines). Blue: Gaussian potentials generated at different positions along the device.

Deconvolution on the basis of the measured capacitance matrix is limited by the resolution imposed by the gate geometry and by the nanotube bandgap, which together determine the size of quantum dots used to measure \( C_{ij} \). An alternative technique, exploiting the clean electrostatic environment of our devices, can be used to design potentials independent of measurement limitations. In section 4.4, we saw that each gate voltage \( V_{gi} \) imposes a bare electrostatic potential at the nanotube \( \phi_i(x) \), given by \( \phi_i(x) = \alpha_i(x) V_{gi} \) where \( \alpha_i(x) \) is the dimensionless function defined in Equation 4.1 (the work functions are set to zero). We can thus define a matrix \( \alpha(x) \) that...
together with the set of all gate voltages \( V_g \) defines the total electrostatic potential at the nanotube,
\[ \phi(x) = \alpha(x) V_g. \]
Suppose we desire a target bare potential \( \phi_{\text{target}}(x) \), and want to know how close a given device can approximate this target potential. This amounts to solving the underdetermined system of equations \( \tilde{V}_g = \alpha^{-1}(x) \phi_{\text{target}}(x) \), where \( \tilde{V}_g \) is the set of gate voltages that best approximates \( \phi_{\text{target}}(x) \) for the given device, and the matrix inversion is performed according to a least-squares algorithm (MATLAB). The actual potential realized at the nanotube is then given by computing \( \tilde{\phi}_{\text{target}}(x) = \alpha(x) \tilde{V}_g \).

An example of this procedure is shown in Figure 4.11a. For this example, \( \phi_{\text{target}}(x) \) is chosen to be a Gaussian function of \( \sigma = 150 \text{nm} \), positioned near the center of the nanotube, in a device geometry similar to the 16-gate device of Figure 4.3. The resulting best-fit potential realizable by the gates, \( \tilde{\phi}_{\text{target}}(x) \), is plotted along with \( \phi_{\text{target}}(x) \), and the two overlap nearly perfectly. The best-fit gate voltages, \( \tilde{V}_g \), are shown in Figure 4.11b.

What is the effective resolution with which we can design such Gaussian potentials? Figure 4.11c shows a plot of Gaussians at the same position as in Figure 4.11a, for a variety of different width parameters \( \sigma \). Below \( \sigma = 150 \text{nm} \) (dotted red line), oscillations become visible in the wings of the potential, indicating that this is an effective limit on the Gaussian width that can be effectively realized before the potential shape deviates from a Gaussian.

Can we make such Gaussian potentials at any position along the nanotube? We compute the sum of squared errors, \( \sum_x (\phi_{\text{target}}(x) - \tilde{\phi}_{\text{target}}(x))^2 \), for Gaussians centered at different points along the nanotube, shown in Figure 4.11d. The error is high at the edges of the device, but drops to a negligible value all along the nanotube for positions within the outermost gates (dotted red lines). We can thus use gates to accurately reproduce Gaussian potentials all along the ‘bulk’ of the suspended nanotube, as shown for three example locations in Figure 4.11d.

These results indicate that potential wells and quantum dots can be defined at any position along the nanotube with a tunable size, not limited by the number of gates or by the physical gate posi-
tions. The clean electrostatic environment of our devices and the possibility of introducing large numbers of gates allows such new experimental capabilities, and we are only scratching the surface of what potential design can accomplish.

4.6 Double Dots

An important application for multi-gated nanotubes is the creation of tunable double quantum dots, or double dots. Double dots have been used to identify Klein tunneling in nanotubes, for the creation of spin qubits, and for the measurement of the electron-nuclear interaction, among others. Here, we show how our devices allow us to create highly-tunable double dots in designer potentials.

In Figure 4.12a we show conductance measurements of a 16-gate, small-bandgap nanotube device (that shown in Figure 4.3). We measure as a function of two voltages \( V_{gR} \) and \( V_{gL} \), applied to gates 3-6 and 11-14, respectively, while applying a fixed voltage of \(-0.4\, V\) to gates 1-2 and 15-16, and a constant voltage of \(-0.3\, V\) to gates 7-10. The contacts dope the nanotube with electrons, such that quantum dots are created when \( V_{gR} \) and \( V_{gL} \) are at low voltages, creating p-n junctions that confine holes to two quantum dots. This is the case in the lower left hand corner of the 2D scan, where the creation of double dots shows up as a lattice of points where the conductance is nonzero (this is discussed in more detail in section B.3).

In Figure 4.12b, we exhibit a similar 2D conductance scan as a function of \( V_{gR} \) and \( V_{gL} \), but with the voltage on gates 7-10 lowered to \(-0.35\, V\). The lowered voltage increases the tunnel coupling between the two quantum dots, which enhances cotunneling between them at the lines connecting triple points. The lower voltage also increases the capacitance between the two dots, causing the points of nonzero conductance to separate into visible pairs of points. These pairs correspond to the degeneracy points of three charge states that satisfy the condition for transport through the double...
Figure 4.12: A double quantum dot of holes on a 16-gate, small bandgap device. The device shown is the same as that in Figure 4.3, measured at \( T = 4\,\text{K} \). a) Conductance scan as a function of two voltages \( V_{gR} \) and \( V_{gL} \), applied to gates 3-6 and 11-14, respectively. Gates 1-2 and 15-16 are held at \(-0.4\,\text{V}\), and gates 7-10 are held at \(-0.3\,\text{V}\). For low voltages, a double quantum dot of holes is formed with a characteristic lattice of points where conduction is allowed (see section B.3). Upper inset: the measured device and the applied voltages. Scale bar: 1\,\mu\text{m}. b) Conductance scan as a function of \( V_{gR} \) and \( V_{gL} \), with the volt on gates 7-10 now \(-0.35\,\text{V}\).
**Figure 4.13:** Eliminating cross-capacitance in a double quantum dot with the capacitance matrix. The five-gate device used here is the same as that shown in Figure 4.1, measured at $T = 4K$. Conductance is measured as a function of $q_2$ and $q_4$, with $q_1$, $q_3$ and $q_5$ held fixed in each scan. The gate voltage applied to the sample are defined by $V_{gj} = C_{ji}^{-1} q_i$, with $C_{ji}^{-1}$ the inverse of the capacitance matrix obtained in section 4.3 (see section 4.5). x- and y-axes are labelled according to the voltage on the $i^{th}$ gate. Panels (a)-(e) correspond to $q_3 = \{0.95, 0.85, 0.75, 0.65, 0.55\} V$, respectively.

dot system (see section B.3 for more details).

In this double dot, the lines dividing charge states of the dots have slopes that deviate from horizontal or vertical. This arises because of a combination of the interdot capacitance and the cross capacitance that couples a gate voltage to the neighboring dot (see section B.3). Eliminating this cross capacitance means compensating the gate voltages for the neighboring gates. This is precisely given by the capacitance matrix extracted in section 4.3. The inversion of the capacitance matrix, described in section 4.5, allows us to define the charge on the $i^{th}$ segment of nanotube $q_i$ via the relation $q_i = C_{ij} V_{gj}$ and its inverse, $V_{gj} = C_{ij}^{-1} q_i$.

We demonstrate this deconvolution for the five-gate device of Figure 4.1 in Figure 4.13. Here, we fix $q_1$, $q_3$ and $q_5$, and measure conductance as a function of $q_2$ and $q_4$, the effective 'plunger gates' for the two dots, for five values of $q_3$, corresponding to the central tunnel barrier between the two dots. For large values of $q_3$, the two dots are well-coupled, effectively forming a single dot for low electron numbers. As $q_3$ is lowered, the central segment of the nanotube is depleted, creating a stronger barrier between the two dots. The conductance scans reveal the transition from a barrier that allows measurable cotunneling and resolved triple point pairs (Figure 4.13b,c), to a de-
coupled double dot where the triple points are no longer resolved and no cotunneling is present (Figure 4.13d,e). The lattice of points sits on a square grid, revealing that the $q_i$ control knobs create independently-addressed quantum dots with no cross-coupling between the two effective plunger gates $q_2$ and $q_4$.

4.7 Novel Applications of Two-Nanotube Devices

In section 3.5, we demonstrated the ability to place two nanotubes in the same device at a controlled separation (see Figure 3.21). This device geometry enables new device functionality: the ability to use one nanotube as a quantum-dot detector to electrically sense the second tube. This circuit implementation, new to suspended nanotubes, has a large number of potential applications, including: charge detection in ultra-clean 1D systems, charge measurement in quantum information implementations, and measurements of mechanical motion of nanotube nano-mechanical resonators. A few works in the past have incorporated local detectors into nanotube circuits\textsuperscript{88,131}, however, so far these devices have been limited to nanotubes lying on a substrate, and local detection of an ultra-clean suspended nanotube has remained a challenging goal. Here, we demonstrate the possibility to perform such detection in a suspended device using the simplest example, in which we use a quantum dot on one tube to electrically sense the mechanical oscillations of the second tube. This approach is reminiscent to the single electron transistor (SET) motion detection performed on bulk silicon nano-beams\textsuperscript{132}, but here it is shown for the first time in the important context of multi-gated suspended nanotubes.

We note that due to the difficulty to make good SETs or quantum-dot detectors close to a nanotube mechanical resonator, to date, all studies of nanotube mechanical motion have used the gate-dependent transport through the moving nanotube itself to detect the motion. While this detection scheme has been very fruitful in past experiments\textsuperscript{94,96,122,133–135}, it puts important constraints on
which measurements can be performed. Since the detection needs the transport through the resonator to be gate-dependent, it cannot be used, for example, when the transport is blockaded (e.g. within a Coulomb blockade valley) or conversely when it is in a ‘metallic wire’ regime with no gate dependence. Using an external quantum-dot detector to detect the movement, as we demonstrate below, decouples the mechanical and the detection components, alleviating these constraints.

**Figure 4.14:** Electrical detection of a nanotube mechanical resonator motion using a second nanotube quantum dot detector. a) Measurement circuit, overlaid over the SEM image of the two-nanotube device from Figure 3.21 of the main text, that was used in these measurements. Yellow: contacts. Blue/red/white: gates. The nanotubes are colored in red (blue) to reflect regions that are electron (hole) doped. The red blob represents a quantum dot formed on the left side of the bottom nanotube, suspended between two contacts, using a voltage on the gate beneath it, \( V_{\text{det}} \). The second nanotube, doped with holes over its entire length, is driven into motion (indicated by blue arrows) by a frequency-modulated (FM) radiofrequency (RF) signal with a carrier frequency \( f \) applied on the left contact of the dot, that is capacitively coupled to it. The quantum dot acts as a non-linear mixing element, which mixes the RF signal on its source contact and the RF gating signal produced by the oscillating nanotube resonator, to a measurable low frequency signal measured at the drain of the dot (middle bottom contact). The right contact and gate are grounded during the measurement. The contacts and gates of the resonator tube are all kept at the same potential, \( V_{\text{res-det}} \), which we control. Scale bar 300nm. b) The out-of-phase component of the mixing current, \( I_{\text{y,mix}} \) (colormap), measured as a function of the bias between the resonator and detector circuits, \( V_{\text{res-det}} \), and the FM carrier frequency \( f \). The signal vanishes everywhere except at a mechanical resonance of the resonator, whose frequency increases with increasing \( V_{\text{res-det}} \) due to electrostatic tensioning of the nanotube resonator.

The measurement circuit is shown in Figure 4.14. The left segment of the bottom nanotube, contacting the left and middle contacts and suspended above a single gate is used as the quantum-dot detector. A DC voltage on this local gate, \( V_{\text{det}}^{\text{g}} \), creates a dot of electrons (marked red in the figure)
and brings its Coulomb blockade transport to a point that is sensitive to external gating. The top, longer nanotube forms the mechanical resonator, whose mechanical vibrations are measured using a simple adaptation of a standard mixing technique\textsuperscript{94}: A frequency-modulated (FM) radiofrequency (RF) signal with a carrier frequency $f$ is applied on the source contact of the detector quantum dot. This contact, being only 85nm away from the long nanotube, couples to it capacitively, actuating its mechanical motion when $f$ is resonant with one of its mechanical modes. This motion, in turn, produces a fluctuating gate potential on the quantum-dot detector. The size of the oscillating gate potential produced by the mechanical motion is $\delta V_{\text{dec}} = \delta z_{\text{mech}} \frac{1}{C} \frac{dC}{dz} V_{\text{res-dets}}$, where $\delta z_{\text{mech}}$ is amplitude of the mechanical vibration, $C$ and $dC/\,dz$ are the capacitance between the resonator and detector nanotubes and its derivative with respect to their mutual distance, and $V_{\text{res-dets}}$ is the (externally-controlled) bias between the resonator and detector circuits\textsuperscript{94}. This oscillating gate signal is mixed down with the FM signal transmitted directly to the quantum dot by its source contact, through the finite transconductance of the dot’s transport, $d^2 I/dV_g dV_{sd}$, producing a low-frequency mixing current detected at its drain (middle contact) using a lock-in amplifier operating at the FM modulation frequency.

Figure 4.14 shows the out-of-phase component of the mixing current measured as a function of $V_{\text{res-dets}}$ and $f$, at T=4K. In addition to controlling the amplitude of the detected signal, the voltage difference $V_{\text{res-dets}}$ applies also a mechanical force that tensions the long nanotube resonator. This tensioning leads to an increase of the frequencies of its mechanical modes. Figure 4.14 shows one such mechanical resonance, visible as a peak in the mixing current, exhibiting a parabolic dependence of its frequency on $V_{\text{res-dets}}$.

It is important to note that the above measurement was performed when the resonator nanotube was electrically tuned to be in the ‘hole wire’ regime, in which its transport is practically gate-independent. The absence of gate dependence would have not allowed the measurement of the mechanical resonance in the conventional way that uses transport through the resonator itself. How-
ever, such measurement becomes possible here by using the separate quantum-dot detector that can be tuned to a gate sensitive point, independent of the state of the resonator.

4.8 Conclusions

The ability to identify perfect nanotubes and selectively nano-assemble them at predefined positions in an electronic circuit makes possible devices that were previously inconceivable. Currently, we assemble electronically-pristine 1-3 μm-long multi-gated devices in a span of a few hours, suggesting that even far more complex devices are possible with this technique. These new devices constitute a novel laboratory for studying electronic phases of strongly-interacting electrons in 1D, subject to engineered potentials. They also act as clean mechanical resonators that can now be coupled to multiple quantum dots. Furthermore, we demonstrate a new class of devices involving multiple nanotubes positioned at chosen locations, heralding sensitive local charge detectors and coupled 1D systems. We expect these novel devices to lead to a new wave of experiments in nanotubes with applications ranging from fundamental condensed matter physics to nano-electromechanics and quantum information science.

To study interacting 1D phases, we now require a local probe to image spatial charge correlations. In the next chapter, we show how these devices can also be used as sensitive and non-invasive local charge detectors.
Carbon Nanotubes as Highly-Sensitive and Non-Invasive Charge Detectors

Identifying correlated, strongly-interacting phases of 1D electrons demands a sensitive and local charge detector. As discussed in chapter 1, existing detectors are invasive, destroying the fragile states we wish to study. Nanotube single-electron transistors (SETs) are a promising new avenue towards non-invasive, highly-sensitive charge detectors. In this chapter, we discuss their operation, the ad-
vantages of charge detection with our new nanotube devices, and the prospects for new experiments exploring 1D electrons.

5.1 Operation of a Nanotube SET

In Figure 1.9, we discussed the use of scanning SETs for charge imaging and potential measurement, and we now discuss the physical mechanism behind this device. A variety of different implementations of charge detection with SETs have been realized. A fixed SET was used to detect the charge configuration of a nearby double quantum dot by monitoring changes in its conductance\textsuperscript{88,136}. Alternatively, the current through an SET can be held fixed in a feedback configuration, allowing measurements of local potentials\textsuperscript{49}.

The configuration used here is illustrated in Figure 5.1. A bias voltage $V_{SD}$ is applied to the source contact of the nanotube SET. The current $I_{SET}$ is pre-amplified and measured with lock-ins on the drain contact. DC and AC voltage $V_g$ and $\delta V_g$, respectively, are applied to the SET gate. The current through the SET for $V_{SD} > 0$ is shown schematically in the top panel of Figure 5.1b, and exhibits Coulomb blockade peaks with a periodicity $\Delta V_g = e/C_g$, with $C_g$ the capacitance of the nanotube SET to its gate (see appendix B). Performing a lock-in measurement of the transconductance of the SET, $g_{SET}$, with respect to the SET gate excitation $\delta V_g$ results in a signal $g_{SET} = dI_{SET}/dV_g$, as illustrated schematically in the bottom panel of Figure 5.1b. We call this the SET gain, for reasons that will become clear below. Now, a metallic sample is introduced beneath the scanning nanotube SET. DC and AC voltages $V_{SMP}$ and $\delta V_{SMP}$ are applied to the sample. A similar lock-in measurement of the transconductance $g_{SET}$ with respect to $\delta V_{SMP}$ yields a signal $H_{SET\rightarrow S} = dI_{SET}/dV_{SMP}$, the transconductance of the SET with respect to the sample gate.

How do we relate the measured quantities, $dI_{SET}/dV_g$ and $dI_{SET}/dV_{SMP}$, to physical properties of the measured sample? We begin by recalling that the SET current is a function of its charge,
Figure 5.1: Operating principle of the scanning SET. a) Circuit diagram of the measurement. An SET is formed on a nanotube device with a local gate with a DC voltage $V_g$, and an AC excitation $\delta V_g$. A DC bias $V_{SD}$ is applied across the source and drain contacts. DC and AC excitations $V_{SMP}$ and $\delta V_{SMP}$ are applied to the sample. The measurement is performed on the drain contact with a current pre-amplifier and set to a lock-in amplifier. b) Top panel: The current through the SET, $I_{SET}$, as a function of the DC gate voltage $V_g$, exhibits Coulomb blockade peaks at low temperature and $V_{SD} > 0$. Bottom panel: The transconductance of the SET, $g_{SET}$, is measured as a function of $V_g$. By locking-in to the frequencies of the AC excitations, signals proportional to the gate and sample capacitances, $C_g$ and $C_{SMP-SET}$, are obtained.

$$I_{SET} = I(Q_{SET}) \text{ (see section B.2)}.$$ We can thus expand

$$\frac{dI_{SET}}{dV_i} = \frac{\partial I_{SET}}{\partial Q_{SET}} \frac{\partial Q_{SET}}{\partial V_i} = \frac{1}{C_{gSET}} \frac{\partial I_{SET}}{\partial V_{gSET}} \frac{\partial Q_{SET}}{\partial V_i} \quad (5.1)$$

where $V_i$ is a voltage acting on the SET. The quantity $\partial I_{SET}/\partial V_{gSET}$ is just the gain of the detector $g$. The quantity $\partial Q_{SET}/\partial V_i$ is the capacitance $C_i$ with which the voltage $V_i$ gates the SET. Thus we have

$$\frac{dI_{SET}}{dV_i} = g \frac{C_i}{C_{gSET}} \quad (5.2)$$

where $g$ is a function of the SET gate voltage $V_g$. As a result, the ratio between the two measured
quantities, \( g_{\text{SET}} = \frac{dI_{\text{SET}}}{dV_{\text{gSET}}} \) and \( H_{\text{SMP}} = \frac{dI_{\text{SMP}}}{dV_{\text{SMP}}} \), is entirely independent of the gain,

\[
\frac{H_{\text{SMP}}}{H_{\text{SET}}} = \frac{C_{\text{SMP}}}{C_{g\text{SET}}}
\]

which therefore allows the direct measurement of the sample capacitance divided normalized by the SET gate capacitance, independent of the voltage-dependence of the gain.

![Diagram showing detection of quantum dot charging with an SET.](image)

**Figure 5.2:** Detection of quantum dot charging with an SET. a) SET positioned above a metallic sample. The metal screens all fields, and thus as a function of the gate voltage \( V_g \), the potential at the SET \( \phi_{\text{SET}} \) remains zero. b) When the sample is insulating, all electric fields are transmitted through it, and \( \phi_{\text{SET}} \) increases linearly with \( V_g \), with a slope less than one (dotted line) that depends on the ratio of the capacitances, (see section B.2). c) When the sample is a quantum dot in the Coulomb blockade regime, \( \phi_{\text{SET}} \) increases linearly in the Coulomb valleys where the dot is insulating. When \( V_g \) lies on a Coulomb peak, the gate field is screened by the dot and \( \phi_{\text{SET}} \) experiences a sudden drop, which the transconductance picks up as a downward spike at the position of the step. The size of the steps depends on the capacitances as shown (this is simplified to neglect other capacitances; see section B.2).

When the sample probed by the SET is a gated, suspended nanotube, then the sample voltage
$V_{SMP}$ defined above is the potential of the nanotube, $\phi_{NT}$. Measurement of the capacitance that couples this nanotube potential to the SET reveals the state of the nanotube, as we now explain.

As illustrated in Figure 5.2, a gate voltage $V_g$ is applied to the device measured by SET. In the case that the device behaves like a perfect metal, grounded by the source and drain leads, the electric field created by the gate is perfectly screened. The potential at the SET, $\phi_{SET}$, therefore does not change as a function of $V_g$. In this case, the measured $H_g = dI_{SET}/dV_g$ would result in a zero measurement.

Conversely, if the nanotube is a perfect insulator, the gate voltage is entirely unscreened and the SET potential increases linearly with the gate voltage. The slope of this line is less than one, and is given by $C_{g-SET}/C_\Sigma$, where $C_{g-SET}$ is the capacitance of the lower gate to the SET and $C_\Sigma$ is the total capacitance of the SET.

The situation is markedly different when a quantum dot is formed on the suspended nanotube. In this case, the potential of the dot evolves in a sawtooth pattern with the gate voltage (see B). In Coulomb valleys of the quantum dot, the device is non-conducting and the potential increases linearly. At Coulomb peaks, electrons tunnel onto the quantum dot, allowing them to screen the gate-induced potential. If the capacitance of the nanotube to the SET, $C_{NT-SET}$, is less than the capacitance of the nanotube gate to the SET, $C_{g-SET}$, then the tunneling electrons screen the potential only partially, resulting in the jagged linear evolution illustrated in Figure 5.2c (see B).

How does quantum dot behavior show up in the SET measurement? We can express the measured transconductance in terms of the potential,

$$H_g = \frac{dI_{SET}}{dV_g} = \frac{\partial I_{SET}}{\partial Q_{SET}} \frac{\partial Q_{SET}}{\partial \phi_{SET}} \frac{\partial \phi_{SET}}{\partial V_g}$$

$$= g \frac{C_{\Sigma}}{C_{g-SET}} \frac{\partial \phi_{SET}}{\partial V_g}$$

(5.4)

where we have used $Q_{SET} = C_\Sigma \phi_{SET}$. Thus, the measured transconductance is proportional to the derivative of the evolution sketched in Figure 5.2c. As a result, charging of the lower quantum dot...
will show up as sharp downward spikes in a flat background.

5.2 Detection of Charging in a Nanotube Quantum Dot

To demonstrate the use of a nanotube SET, we insert two nanotube devices into a specially-designed scanning probe microscope. This microscope is similar in construction to that described in section 3.4, but designed for the insertion of two circuit chips and for operation in a dilution refrigerator.

Two small-bandgap nanotubes are chosen for each device, allowing us to use either one as an SET. The two devices are carefully brought into close proximity by using capacitance scans with the nanotube SETs. As described in subsection 3.4.1, capacitance reveals geometric features of the chips, allowing us to identify the circuit contacts and gates and approach them in a controlled fashion. The capacitance is obtained with the transconductance discussed above, rather than with measurement of the displacement current.

We demonstrate the operation of a nanotube SET with a pair of identical devices, arranged with the two nanotubes parallel to each other. The nanotube segment suspended over the gates is $\sim 1\mu m$ long in both devices, and lies $\sim 60nm$ above the seven gates. The configuration is shown schematically in the right panel of Figure 5.3a, with the SET at a distance of $\sim 1\mu m$ above the second nanotube. The measurement setup is shown on the left, and is similar to that described in section 5.1, with the difference that separate DC biases and AC excitations may be applied to the seven gates. For this measurement, the seven gates are chained together in both devices. The upper nanotube serves as the SET, which detects the second nanotube below.

To reveal charging of the second, lower nanotube, we measure the two transconductances, $H_{g\text{SET}} = \frac{dI_{\text{SET}}}{dV_{g\text{SET}}}$ and $H_g = \frac{dI_{\text{SET}}}{dV_g}$. Their ratio $H_g/H_{g\text{SET}}$ gives $C_g/C_{g\text{SET}}$, the normalized capacitance of the lower gate to the SET gate. At low voltages, the lower nanotube is doped into its bandgap, making it an insulator. In this regime, the transconductance ratio is a constant reflecting
Figure 5.3: Charging of a nanotube quantum dot measured with a nanotube SET at \( T = 4K \). a) Right panel: Schematic geometry of the setup. Two suspended nanotube devices are brought into close proximity inside a scanning probe microscope, here in a parallel configuration. Left panel: The real measurement circuit. DC and AC excitations \( V_{g_i} \) and \( \delta V_{g_i} \) are independently applied to the seven gates of the lower sample. The upper nanotube is operated as an SET positioned above the center of the lower sample. All of its gates are chained together for the transconductance measurement, and receive DC and AC excitations \( V_{g\text{SET}} \) and \( \delta V_{g\text{SET}} \). A source-drain bias \( V_{SD} \) is applied to the source contact, and measurement is locked-in to the drain. b) Transconductance of the SET with respect to all seven gates of the lower sample chained together, scaled by the transconductance of the SET gate, \( H_g/H_{g\text{SET}} \), measured as a function of the chained DC gate voltage \( V_g \) of the seven lower gates. At high voltages, the lower nanotube is doped into its bandgap (inset, upper right), and the measured value is the ratio of capacitances. At lower voltages, a hole quantum dot is formed (inset, upper left) and discrete charging of the quantum dot is detected as downward spikes in the transconductance.

the geometric capacitance of the lower gates to the SET. As the voltage is lowered, the transconductance ratio exhibits downward peaks, meaning a sudden lowering of the capacitance of the gates
to the SET at discrete values of the gate voltage. This indicates the formation of two p-n junctions
at the edges of the device, defining a hole quantum dot whose charging shows up as the spikes. At
these points, the quantum dot is allowed to charge, screening its gate from the SET.

5.3 Spatially Resolving Charge in a Multi-Gated Double Dot

We can use the unique multi-gated feature of our nanotube devices to demonstrate spatially-resolved
detection. To do this in the same configuration as described above, we separately gate the seven gates
of the lower device shown in Figure 5.3. Seven independent AC excitations are applied to all seven
gates. The left three gates, gates 1-3, are biased with a DC voltage $V + \varepsilon$, while the right three gates,
gates 5-7, are biased with a voltage $V - \varepsilon$, as illustrated in Figure 5.4. The central gate, gate 4, is biased
with a constant voltage that depletes the central region. As we saw in section 4.6, this voltage config-
uration can result in a double quantum dot, which has distinct signatures in transport measurements.
Below, we show how our nanotube SET can resolve the double-dot configuration.

We measure the transconductance of all seven gates independently for a fixed SET position
above the nanotube. This results in seven 2D scans of the scaled transconductance $H_{gi}/H_{gSET}$, with
$i = 1...7$, as a function of the common voltage $V$ and the detuning voltage $\varepsilon$. These are shown in
Figure 5.4 on the same color scale. When the dots are in their Coulomb valleys, they are insulating
and the measured values correspond to the scaled capacitances of the local gates to the SET. The ca-
pacitance is the highest for the central gate, and grows steadily lower for the edge gates. This reflects
the geometry of the devices, in which the edge gates are screened by the contacts and farther from
the center of the SET. The difference in coloring between symmetrically-positioned gates, such as
gates 1 and 7, reflects that the SET is not perfectly centered above the center of the device, resulting
in a measurable difference in the capacitance of the edge gates to the SET.

Charging of the two quantum dots appears on the flat background as stripes of low $H_{gi}/H_{gSET}$.
Figure 5.4: Spatially-resolved charging in a double dot at $T = 4\,\text{K}$. a) Voltage configuration for the double dot measurement. A DC voltage $V$ is applied to gates 1-3 and 5-7 of the lower device. A detuning voltage $\varepsilon$ is applied to gates 1-3 while $-\varepsilon$ is applied to gates 5-7. Gate 4 is maintained at a constant voltage. AC excitations are $50\,\mu\text{V}$. b) Scaled transconductance of the individual gates as a function of the common voltage $V$ and the detuning voltage $\varepsilon$. The SET is maintained at a fixed position above the center of the lower device. Charging of the left dot and right dot is visible by lines of different slopes. See main text.

These correspond to the expected downward spikes in the capacitance, now plotted as a function of the two voltages $V$ and $\varepsilon$. On the left, for gates 1-3, the stripes have a negative slope, reflecting the positive contribution of the detuning voltage to the left quantum dot (see section B.2). The lines
are visible in the transconductance of all three gates participating in the gating of the left quantum dot. Conversely, on the right side, for gates 5-7, the stripes have a positive slope, due to the negative contribution of the detuning to the gating of the right quantum dot. The stripes are also visible for all three of the participating gates. Interestingly, the transconductance with respect to gate 4 in the center reveals spikes corresponding to both the left and right quantum dots. Here, the screening of the gate due to either dot is approximately of equal magnitude, resulting in identically-colored lines for both the left and right dots. The signal of gate 4 therefore constitutes an experimental charge stability diagram of the entire double quantum dot system, allowing one to read out the relative number of localized charges in each dot.

This measurement demonstrates how a nanotube SET, in combination with multi-gated devices, can be used to spatially resolve charging along a suspended nanotube.

5.4 Mapping Gates and Enhancing Resolution with a Perpendicular Nanotube

For imaging of correlated states in nanotubes, we need high sensitivity, spatial resolution, and non-invasiveness. The nanotube SET allows us to achieve this by scanning perpendicular to the axis of the target nanotube. We now demonstrate preliminary results demonstrating how this allows us to exceed the resolution limit of the parallel configuration.

To benchmark the parallel configuration, we utilize the measurement scheme shown in Figure 5.3 to measure transconductances of the seven gates as a function of position along the lower sample. We choose a position far away from the second nanotube, allowing us to avoid screening from charging of the nanotube. For these measurements, the SET is at a distance of \( \sim 0.6\mu m \) from the second chip, as illustrated in Figure 5.5a. We measure the scaled transconductance with respect to each individual gate for the parallel configuration as a function of space. This is plotted in Figure 5.5b, for all of the gates. The seven gates positions are revealed by the peak positions, which are spaced by
Figure 5.5: Scanning SET imaging of device gates in parallel and perpendicular configurations. a) Scanned SET imaging of device gates in the parallel configuration. The SET is positioned ~0.6 μm above a seven-gated device, and scanned perpendicular to the gates. The gates have a 200 nm pitch. b) Scaled transconductance measured with respect to the seven gates of the lower device while the SET scans across it. The gates show up as peaks, whose width of ~1 μm is greater than the SET-sample distance and corresponds instead to the SET size of 1 μm. b) SET imaging of gates in a perpendicular configuration. The SET-sample distance is 1 μm. b) The measured transconductance with respect to two gates spaced by 1 μm and 450 nm in width. The peak widths are 1 μm, corresponding to the SET-sample distance.

200 nm as expected from the device design. This is a direct, scanning measurement of the capacitance matrix extracted in a discrete fashion in section 4.3. We can thus directly measure the gate capaci-
tance functions in their entirety, allowing us to design experimentally-validated potentials.

The width of the measured peaks is $\sim 1\mu m$. This is greater than the sample-to-sample distance of $0.6\mu m$, and originates in the size of the nanotube SET, whose suspended length is $1\mu m$. The parallel configuration causes the signal to be convolved with the large length of the scanning SET, resulting in broadened peaks. The resolution is thus apparently limited by the SET size.

Our scanning microscope allows us to orient the nanotube SET perpendicular to the target nanotube. The configuration is illustrated in Figure 5.5c. For a second pair of samples, we measure the scaled transconductance, here with respect to two gates, each $\sim 450nm$ in width and separated by $\sim 1\mu m$. The SET is at a distance of $\sim 1\mu m$ from the second chip, and is scanned perpendicular to the SET nanotube axis. The peaked measurements reveal the positions of the two measured gates and their distance from each other, corresponding to the design of the chip. The peak widths are $\sim 1\mu m$. This is greater than both the gate size and the SET size, which in this direction is merely $1nm$. Instead, this peak width corresponds to the distance of the SET from the gates. The resolution of imaging thus appears to be limited not by the detector size, but by the detector distance. By bringing the SET closer to the sample, we expect to achieve high spatial resolution limited only by the SET nanotube diameter. Doing this requires a scanning probe microscope with high mechanical stability, and work is ongoing to realize this goal.

5.5 Future Directions

The scanning nanotube SET constitutes a highly sensitive probe of the electrostatic landscape. We routinely measure devices with sensitivities of $< 10^{-5} e/\sqrt{Hz}$. By bringing the nanotube SET in close proximity to a second nanotube device, while keeping all metallic electrodes far from the SET and the second nanotube, we can realize a non-invasive, high-resolution imaging capability.

The ability to sensitively probe the local electrostatic potential of a 1D system makes possible a
wide array of new experiments. With a non-invasive detector, fragile interacting states can be probed by imaging the charge density in space. States like the Wigner crystal or the Mott insulator, with periodic charge density, will be directly identifiable by their periodic electrostatic signatures.

Knowledge of the electrostatic potential at all points along the 1D system also heralds a new form of measurement for out-of-equilibrium systems. By measuring the potential in space while a voltage biases the nanotube, we can measure where potential drops occur along the nanotube. In particular, this can be done in the presence of designer potential barriers and other potentials. Such a probe would constitute a generalization of a four-probe measurement, with the possibility to measure the potential drop at all points along the system. Because this measurement takes place without invasive contacts, we can extract information about potential drops without incoherent probes, allowing us to probe without disturbing coherent properties of the system.

The scanning nanotube can also serve as a local scatterer. By biasing it with respect to the second nanotube, it can serve as a local, tunable perturbation that can be used to probe interacting phases. This type of experiment has been proposed to identify unique 1D phenomena such as spin-charge separation and the 1D Kondo cloud.

We are only scratching the surface of the new experimental possibilities. In this thesis, we have realized a non-invasive local probe for an ultra-clean 1D system in designer potentials, and we expect this will lead to exciting developments for the study of strongly-interacting condensed matter systems, quantum information science, and beyond.
In this appendix we provide a more detailed derivation of the results quoted in section 1.2. Perturbations to the basic picture are explored, with the goal of developing an intuitive picture of the electronic properties of carbon nanotubes.
A.1 Tight-Binding on Graphene

To understand the electronic properties of carbon nanotubes, we begin with their parent material: graphene. Graphene is a hexagonal lattice of carbon atoms, or a Bravais triangular lattice with a two-atom basis. Expressing the sub-lattice basis vectors as \( a_1 = \frac{l}{2}(3, \sqrt{3}) \) and \( a_2 = \frac{l}{2}(3, -\sqrt{3}) \), with \( l \approx 1.42 \text{Å} \) the atomic bond length, gives reciprocal lattice vectors \( b_1 = \frac{2\pi}{3l}(1, \sqrt{3}) \) and \( b_2 = \frac{2\pi}{3l}(1, -\sqrt{3}) \). The nearest-neighbor sites for a given atom lie on the other sub-lattice, with positions given by \( \delta_1 = \frac{l}{2}(1, \sqrt{3}), \delta_2 = \frac{l}{2}(1, -\sqrt{3}) \) and \( \delta_3 = -l(1, 0) \). With these in mind, we can write down the tight-binding Hamiltonian in convenient second-quantized language as

\[
H = -t \sum_{<ij>} (a_i^\dagger b_j + a_i b_j^\dagger)
\]  

(A.1)

where \( a_i^\dagger \) and \( a_i \) are the creation and annihilation operators, respectively, of sub-lattice \( a \) at site \( i \) (spin is ignored here), \( t \) is the nearest-neighbor hopping matrix element, and the sum is over nearest-neighbors. We can cast this problem in a more transparent form by expressing the Hamiltonian as a matrix in the basis of sub-lattices:

\[
H = -t \sum_{<ij>} \begin{pmatrix} a_i^\dagger & b_j^\dagger \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_i \\ b_j \end{pmatrix}
\]  

(A.2)

This is more than notation, and later on will play a crucial role in elucidating the physics. We proceed according to tradition, by Fourier transforming the site operators to momentum space via the relation \( a_i = \sum_k e^{ik\cdot\mathbf{R}_i} \phi_k \), where the vector nature of the lattice site counting has been made explicit.
and the momentum sum is over the first Brillouin zone.

\[ H = -t \sum_k \sum_i \begin{pmatrix} a_k^\dagger & b_k^\dagger \end{pmatrix} \begin{pmatrix} 0 & e^{-i\delta_i k} \\ e^{i\delta_i k} & 0 \end{pmatrix} \begin{pmatrix} a_k \\ b_k \end{pmatrix} \]  

where \( \delta_i \) are the nearest-neighbor vectors given above. The nearest-neighbor sum can be directly reduced into sums of three phase factors, each a function of the reciprocal momentum components \( k_x \) and \( k_y \), allowing us to diagonalize the matrix. This reveals the eigenvalues

\[ E = \pm t \sqrt{3 + f(k)} \]

where

\[ f(k) = 2 \cos(\sqrt{3}k_y) + 4 \cos(\frac{3k_x}{2}) \cos(\frac{\sqrt{3}k_y}{2}) \]

The spectrum is plotted in Figure A.1. The two sub-lattices, which led to the matrix form of the hopping Hamiltonian, have resulted in two bands. Since we have one electron contributed per carbon atom (and two spin states), we expect to be at half-filling, so we identify these bands as our new hole and electron bands, with the Fermi energy lying precisely between them at zero energy. Remarkably, at zero energy the two bands are touching. For the typically-chosen hexagonal Brillouin zone, the touching happens at the six points at the zone corners, at the momenta \( K = \frac{2\pi}{3\sqrt{3}} (1, \frac{1}{\sqrt{3}}) \) and \( K' = \frac{2\pi}{3\sqrt{3}} (-1, \frac{1}{\sqrt{3}}) \). Alternatively, we can choose a Brillouin zone more clearly related to the original molecular basis, a parallelogram centered around the two atomic sites, and see that the touching happens at two points, shifted symmetrically away from (the newly-chosen) zero momentum. These are the \( K \) and \( K' \) points of lore. The two bands appear to touch in an exquisitely delicate manner, suggesting that we should expand the spectrum about the two points, obtaining

\[ E = \pm \frac{3}{2}v_F|q| = \pm h\nu_F|q| \]
Figure A.1: Structure and spectrum of graphene. a) Graphene lattice. b) Graphene Brillouin zone. c) Spectrum of graphene. d) States at the Fermi energy fall on linearly-dispersing Dirac cones. (a) and (b) adapted from 55.

where $\mathbf{q}$ is the momentum vector relative to the K or K’ points, and we defined the Fermi velocity $v_F$. With a hopping parameter $t \approx 2.8 \text{eV}$ and carbon-carbon distance of $l = 1.42 \text{Å}$, the Fermi velocity is roughly $10^6 \text{m/s}$, a healthy one percent of the speed of light $c$. The linear energy-momentum relation around the K/K’ points earns them the name of Dirac cones, after the theory describing ultrarelativistic particles with such a linear dispersion.

How does the wavefunction behave around the Dirac cones? From the matrix Hamiltonian description, we see that graphene wavefunctions are two-component vectors in the basis of the A and
B sublattices. There is a real-space (or momentum-space) function describing the spatial evolution of the state, accompanied by a spinor describing the distribution of the wavefunction between the two sublattices. A real electron spin is described by a spinor parametrized by two angles defining a Bloch sphere, and an analogous pseudospinor describes the graphene wavefunction. In the graphene problem, the tight-binding model as we have defined it assumes symmetry of the A and B sublattices. Thus the spinor can only be an eigenstate of the Hamiltonian if it respects this symmetry, which it does so by keeping the wavefunction always equally distributed between the two sublattices, up to a phase factor. Hence, our graphene spinor cares only about a single parameter, the complex phase angle between wavefunctions living on either sublattice, which corresponds to the complex equatorial angle of the pseudospin Bloch sphere. The pseudo-spinor then looks like

$$\Psi \propto \begin{pmatrix} 1 \\ e^{i\phi} \end{pmatrix} \quad (A.7)$$

and therefore we have to solve for the spatial wavefunction on a single sublattice, and then compute sub-lattice phase to stitch the two sublattices together.

To make this argument more rigorous, we construct the effective low-energy theory by linearizing the Hamiltonian at zero energy. At zero energy we have only two points to consider, K and K', and thus we expand the phase factors around them, obtaining

$$H \propto -\frac{3}{2} \times \begin{pmatrix} 0 & e^{i\xi} (q_x - i q_y) \\ e^{-i\xi} (q_x + i q_y) & 0 \end{pmatrix} \quad (A.8)$$

and

$$H' \propto -\frac{3}{2} \times \begin{pmatrix} 0 & e^{i\xi'} (q'_x + i q'_y) \\ e^{-i\xi'} (q'_x - i q'_y) & 0 \end{pmatrix} \quad (A.9)$$
where we’ve ignored the sums and operators, and the $q$’s are momentum wavevectors with respect to $K$ and $K’$. Noticing the factor of $-i$ between the $x$ and $y$ components we take creative license to rewrite this as

\[ H \propto \sigma_x k_x + \sigma_y k_y = \sigma \cdot k \]  \hspace{1cm} (A.10)

and

\[ H' \propto \sigma'_x k'_x - \sigma'_y k'_y = \sigma' \cdot k' \]  \hspace{1cm} (A.11)

where we’ve defined $\sigma = (\sigma_x, \sigma_y)$ and $\sigma' = (\sigma'_x, -\sigma'_y)$. Now we can see explicitly that there is no $\sigma_z$ term, justifying the general form of the pseudospinor conjectured above. In other words, the eigenstates of the (K-point) Hamiltonian take the form of Equation A.7, as can be seen by plugging in ansatz pseudospinors into the Schrödinger equation. Doing this explicitly reveals a key difference between $K$ and $K'$ points. At the $K$ point, the sub-lattice phase is related to the momentum wavevectors as

\[ q_x + iq_y = e^{i\phi} \]  \hspace{1cm} (A.12)

while around the $K'$ point the relation is

\[ q'_x - iq'_y = e^{i\phi'} = e^{-i\phi} \]  \hspace{1cm} (A.13)

so that an eigenspinor at $K$ looks like Equation A.7, while an eigenspinor at $K'$ has the opposite phase. That is, the phase winds around $K'$ in the opposite direction of $K$. We can calculate this phase from the Hamiltonian, complimentary to the eigenspectrum, as visualized in Figure A.2.

What does the wavefunction in an individual sublattice look like? In the tight-binding approximation, the phase accumulated in hopping from site to site of a single sublattice is given by $e^{i\mathbf{r}_i \cdot \mathbf{k}}$, with $\mathbf{r}_i$ the vector between the two sites. To see how this arises from the nearest-neighbor model of Equation A.1, where hopping is only between sublattices, one can trace out the second sublattice.
This will lead to cancellation of the $b_j$ terms on the same site, cancelling their phase. So a pair of terms looking like $e^{i\pi k}\hat{c}_i\hat{c}^\dagger_{i+1}b_j$ and $e^{-i\pi k}\hat{c}_i\hat{c}^\dagger_{i+1}b_j$ get multiplied, so that the $b$-sites cancel, leaving only $e^{i(\pi k-\pi k)}\hat{c}_i\hat{c}^\dagger_{i+1}$, with the expected phase.

A more elegant way to this result is to solve the tight-binding problem on a single-sublattice. In this case, the Hamiltonian is

$$\tilde{H} = -i \sum_{<i,j>} \hat{a}_i^\dagger \hat{a}_j$$  \hspace{1cm} (A.14)

and the Fourier substitution defined above gives

$$\tilde{H} = -i \sum_{k} \sum_{i} e^{i\delta \cdot k} \hat{a}_k^\dagger \hat{a}_k$$  \hspace{1cm} (A.15)
Figure A.3: Structure and spectrum of the triangular lattice: a single sublattice of graphene. a) Structure of the lattice. Tight-binding hopping parameters $t$ are defined for nearest-neighbor sites. b) The spectrum, revealing $K$ and $K'$ points at the vertices of the Brillouin zone. c) Tight-binding wavefunctions around the $K$ (left panel) and $K'$ (right panel) points. The two wavefunctions are related by a reflection around the horizontal axis.

with $\vec{d}_1$ the nearest-neighbor vectors of the triangular lattice and the sums over the first Brillouin zone and nearest-neighbors, respectively. We compute the phase factor sum, here $-2(\cos(k \cdot d_1) + \cos(k \cdot d_2) + \cos(k \cdot (d_1 - d_2)))$, which immediately gives the spectrum, shown in Figure A.3. Note that the $K$ and $K'$ points are present here, identical to the graphene case. The $K$ and $K'$ points...
therefore originate in the structure of one sublattice; the consequence of the additional atom in the unit cell of graphene is the two bands of electrons and holes.

The single-sublattice wavefunction can now be obtained by evaluating the phase factors $\tilde{\delta}_j \cdot \mathbf{k}$. The result around the K point, for nearest neighbors starting at $\mathbf{a}_1$, is $\phi = \frac{2\pi}{3}, \frac{4\pi}{3}, ...$ while around K', it is $\phi' = 4\pi, \frac{2\pi}{3}, ...$, as shown in Figure A.3. The wavefunctions around the K and K' points are a form of anti-bonding wavefunction respecting the lattice symmetry, related to each other by a rotation of $\frac{2\pi}{3}$. We can combine these using the sublattice phase (the pseudospin) to get the full graphene wavefunction at low energies.

A.2 Nanotubes: Quantized Graphene

To obtain a nanotube, we begin with the simplest approach: apply periodic boundary conditions to graphene. First, we choose a vector to ‘wrap’ the graphene. To make a legal Bravais lattice, this vector must be drawn between two atoms of the same sublattice, such that it can be written in terms of the basis vectors as $\mathbf{C} = n\mathbf{a}_1 + m\mathbf{a}_2$, where $n$ and $m$ are called the chiral indices. The hexagonal symmetry of the graphene lattice implies that only 30 degrees of the plane are unique, so we have only to consider $\mathbf{C}$ between and including $(n, 0)$ and $(n, n)$, with $0 \leq m \leq n$. These indices are known as the chirality, with $(n, 0)$ evocatively named ‘zigzag’ and $(n, n)$ named ‘armchair’. The length of $\mathbf{C}$ is then the circumference of the tube, which is $|\mathbf{C}| = \sqrt{3l^2 + m^2 + nm} = 2.46\sqrt{n^2 + m^2 + nm} \, [\text{Å}]$. For a zigzag tube, this gives a diameter of $d = 0.78 \times n \, [\text{Å}]$, so taking a typical value of 1nm diameter for a zigzag tube implies at least 10 hexagons around the circumference. For the zigzag tube, $\mathbf{C}$ and $\mathbf{a}_1$ are parallel, so we define the chiral angle as the angle between $\mathbf{C}$ and $\mathbf{a}_1$, or $\cos \theta = \mathbf{C} \cdot \mathbf{a}_1 / |\mathbf{C}| |\mathbf{a}_1|$, which takes values between 0 and $\pi/6$.

To define the nanotube unit cell, we need a second vector perpendicular to the wrapping vector, call it $\mathbf{L}$, defined by $\mathbf{C} \cdot \mathbf{L} = 0$. Its length is given by finding the first atom of the same sublattice...
Figure A.4: Wrapping graphene and getting a nanotube. a) The structure of a nanotube is obtained by choosing a wrapping vector $C$ between two atoms of one sublattice. See main text for definitions. Wrapping of a (5,3) nanotube is shown as an example. Adapted from 138. b) Top panel: Wrapping quantizes the momentum along the circumference, $k_{\perp}$, leaving the axial wavevector, $k_{||}$, the only continuous degree of freedom. Bottom panel: Quantization slices the graphene spectrum into subbands. c) The K/K’ wavefunctions predict whether a subband will the Dirac points on the nanotube. If the two atoms on either end of the wrapping vector have the same phase, then the subband crosses the Dirac point and the nanotube is metallic. d) If the two atoms have a different phase, the lowest subband does not cross the Dirac point, and the nanotube is a gapped semiconductor.

that it intersects. For example, a zigzag tube has $L$ such that one hexagon is included in the unit cell length. The rectangular area covered by the wrapping and length vectors defines the unit cell of the
nanotube. To form the nanotube, the unit cell is tesselated along the nanotube length parallel to \( \mathbf{L} \). Thus, there is one continuous wavevector running along the nanotube length, and a quantized vector along the circumference. The number of carbon atoms inside the unit cell, \( 2N \), tells us how many subbands \( N \) arise in the 1D Brillouin zone, with each an electron-hole pair.

To treat the quantization condition imposed by the wrapping vector, we define a momentum wavevector parallel, \( k_\parallel \), and perpendicular, \( k_\perp \), to the nanotube axis. The quantization condition for wrapping is

\[
\exp(ik_\perp |C|) = 1
\]  

(A.16)

giving

\[
k_\perp = \frac{2\pi \nu}{|C|}
\]  

(A.17)

with \( \nu \) an integer taking \( N \) possible values corresponding to the \( N \) subbands. As we mentioned, \( k_\parallel \) remains continuous. This implies that the spectrum of the nanotube will be a periodic slicing of the graphene bands into discretized lines, each with a length \( 2\pi/|L| \) (the size of the Brillouin zone) and spaced from each other by \( 2\pi/|C| \). In the reduced zone scheme, we fold the \( N \) bands back into the first Brillouin zone, obtaining the \( 2\pi/|C| \)-periodic slicing shown in Figure A.4.

Where does the 1D nanotube slice the graphene spectrum? We seek the low-energy properties around \( E = 0 \), so we are interested in whether the subbands touch the K/K' points. At a K point, the wavefunction in a single sublattice looks like that shown in Figure A.4. Wrapping a nanotube takes one atom and makes it equivalent to another in the same sublattice, but not all atoms are equivalent. If the phase of one atom is the same as that at the end of the wrapping vector \( C \), then the K/K' wavefunction is still valid. But if it is any other atom, then in order for Equation A.16 to hold, an additional phase must be acquired by the wavefunction as it wraps around the circumference. We
can express this state of affairs as follows

\[ k_\perp |C| = \mathbf{K} \cdot \mathbf{C} + \phi_p \]  

(A.18)

with a corresponding expression around the K’ point. Here, \( \phi_p \) is the extra phase that takes the K/K’ wavefunction phase (either 0, \( \frac{2\pi}{3} \), or \( \frac{4\pi}{3} \)) back to 0, with one of three values \( (0, \frac{-2\pi}{3}, \text{or} \frac{-4\pi}{3}) \). It is convenient to replace \( \frac{4\pi}{3} \) with \( \frac{-2\pi}{3} \), so that we can express the phase as \( \phi_p = \frac{2\pi}{3} \times p \) where \( p = \{0, 1, -1\} \). Whether \( \phi_p \) is nonzero then depends on which \( C \) we choose, and the K/K’ wavefunction tells us precisely which ones. We can express that knowledge in terms of the chiral indices as

\[ n - m = 3q - p \]  

(A.19)

where \( p \) is the phase multiplier defined above, equal to 0, 1, or -1, and \( q \) is an integer. When \( n - m \) is an integer multiple of three, as it is for a zigzag or armchair tube or for any two blue points in Figure A.5, then \( p = 0 \) and a 1D subband exists which slices through the K and K’ points. In this case, the nanotube inherits graphene’s semimetal nature and is a ‘zero-bandgap’ conductor. In all other cases, the 1D subband misses the K and K’ points and the lowest subband has a gap, making the nanotube an intrinsic semiconductor.

From this analysis we learn that the K/K’ wavefunction shown in Figure A.3 is really a nanotube periodic table. Starting at a blue point, any wrapping vector that ends at another blue point yields a metallic nanotube. Wrapping vectors that end at any other point yield a semiconductor (see Figure A.5). In this picture, roughly one third of all nanotubes are metallic, and the rest are semiconducting.

How big is the gap in a semiconducting nanotube? We can estimate this using the Fermi velocity of graphene we derived in section A.1 and the change in \( k_\perp \) from wrapping. The extra phase
acquired by the wavefunction due to wrapping of a non-K/K' atom effectively shifts the lowest subband away from the K/K' point, by an amount which we find to be

$$k_{\perp} = K + \frac{\phi_p}{|C|} = K + \frac{2p}{3d}$$  \hspace{1cm} (A.20)

where $d$ is the nanotube diameter and $p$ is the wrapping parameter defined above. The gap that opens up for a semiconducting nanotube is then

$$E_g = 2\hbar v_F \times \frac{2}{3d} \approx \frac{0.8}{d[\text{nm}]} eV \approx \frac{9300}{d[\text{nm}]} K$$  \hspace{1cm} (A.21)
The size of the gap indicates that even far above room temperature only the lowest subband is energetically accessible, rendering nanotubes strongly one dimensional. For many experimentally-available nanotubes, this is still sufficiently small to call them a bonafide semiconductor, with valence and conduction bands energetically accessible by doping (electrostatic or chemical).

What is the dispersion for gapped nanotubes? The lowest subband slices the Dirac cone away from the Dirac point, thus at low energies the spectrum should be well-approximated by a hyperbola, whose form is

\[ E \approx h v_F \sqrt{k_\parallel^2 + (\Delta k_\perp)^2} = \sqrt{(h v_F k_\parallel)^2 + E_g^2 / 4} \quad (A.22) \]

with \( \Delta k_\perp \) the perpendicular distance from the Dirac point.

The quantized 1D subbands tell us about motion along the axis of the nanotube via the longitudinal velocity, given by \( v_{NT} = \partial E / \partial k_\parallel \). For metallic nanotubes, this velocity is just the Fermi velocity of graphene. For semiconductors, the slicing of the Dirac cone gives a hyperbola, and the velocity reaches the graphene Fermi velocity asymptotically at large momenta/high energy. At the subband bottom, the longitudinal velocity approaches zero.

---

\[ \text{Figure A.6: Isospin and the orbital magnetic moment of carbon nanotubes. a) Nanotube subbands that don’t intersect the Dirac point possess a fixed component of } k_\perp, \text{ and a corresponding transverse velocity } v_{\text{orb}} = \partial E / \partial k_\perp. \text{ This velocity is equal and opposite for the corresponding subband on the K’ cone. b) The transverse velocity means electrons travel in a loop around the nanotube circumference, generating a magnetic moment known as the isospin, in analogy to the real electron spin. Adapted from}^{112,139} \]
There is an additional velocity in the transverse direction given by $v_{\text{orb}} = \partial E / \partial k_\perp$. This quantity is zero for the lowest subband of a metallic nanotube, since all points of the metallic dispersion are at the apex of hyperbolae sliced in the perpendicular direction. For semiconductors, and for higher subbands of metallic nanotubes, the subband is positioned away from the Dirac point and all points of the dispersion have a nonzero orbital velocity, in the direction of the nanotube circumference (Figure A.6). This does not break time-reversal symmetry, since the $k = 0$ solution is always valid, and thus all subbands are positioned symmetrically with respect to $k = 0$. Thus every subband with $v_{\text{orb}}$ has a degenerate partner found by reflection across $k = 0$ with $v'_{\text{orb}} = -v_{\text{orb}}$ (Figure A.6).

At the subband bottom, $v_{\text{orb}} = v_F$, dropping asymptotically to zero at high energies.

What is the physical meaning of this transverse velocity? It defines a circulating current around the nanotube circumference, which therefore implies a magnetic moment of size

$$m_{\text{orb}} = I_{\text{orb}} A_{\text{NT}} = e v_{\text{orb}} \frac{d}{\pi d^2} = \frac{e}{4} v_F(E) d$$

where $v_F(E)$ is the energy-dependent Fermi velocity. Assuming $v_F(E) = v_F$, the graphene value, and $d = 1\text{ nm}$ nanotube, this gives a magnetic moment of $\lesssim 4\mu_B$, with $\mu_B$ the bare electron spin magnetic moment. This degree of freedom is termed the isospin. Combined with the real spin, this gives a fourfold state degeneracy. Like the spin, it is quantized (thanks to the quantized subbands), but unlike the spin, it will couple only to magnetic fields pointing along the nanotube axis.

We have seen how subbands lead to isospin, but how does the pseudospin degree of freedom affect this picture? The sublattice phase that we calculated (shown in Figure A.2), like the energy spectrum, will be sliced by the wrapping quantization of the nanotube. The 1D subbands will therefore inherit specific values of the phase for different $k_\perp$. If the nanotube is metallic, the slice passes through the K and K’ points. This means that the metallic subband around the K point has a right-moving branch with a sublattice spinor of phase 0 and a left-moving branch with phase $\pi$, but these states are degenerate because of the time-reversal symmetry.
Figure A.7: Pseudospin and backscattering in an armchair nanotube. a) Sublattice phase around the K and K’ points, reproduced from Figure A.2. Black circular arrows: pseudospin evolution around the K/K’ points. Small arrows: pseudospin direction around the K/K’ points. Gray line: lowest subband of an armchair nanotube, slicing the pseudospin at the highlighted points (blue and yellow arrows). b) The lowest subband of an armchair nanotube. Pseudospin of each branch indicated (blue and yellow; see panel (a)). In the K valley, forward-moving states have pseudospin pointing towards $k_{\parallel}$ (blue), while backward-moving states have opposite pseudospin pointing towards $-k_{\parallel}$ (yellow). Backscattering to the K’ valley requires a large momentum of $\sim K$, and as a result, backscattering is suppressed. Lower panel: structure of an armchair nanotube.

while around the K’ point, right-movers have a spinor of phase $\pi$ and left-movers have phase $0$ (see Figure A.7).

Consider a backscattering process between right- and left-movers in the K valley of this armchair tube. The expectation value of the scattering will be the contraction of the scattering operator with the two initial and final state pseudospinors. In the pseudospin basis, this looks like

$$\begin{pmatrix} 1 & -1 \\ S_{AA} & S_{AB} \\ S_{BA} & S_{BB} \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = (S_{AA} - S_{BB}) + (S_{AB} - S_{BA})$$

(A.24)

where $S_{AA}$ terms are the scattering elements in the sublattice basis, and the two vectors represent the K-point right- and left-mover. We can plainly see that this scattering has a nonzero expectation value.

135
only if the scatterer breaks sublattice symmetry. To break this symmetry, the scatterer must have a sufficiently sharp potential variation on the scale of the carbon-carbon bond. The only other way to backscatter would be to send the $K$-right mover to the $K'$-left mover, which has the same pseudospin, but this requires a change in momentum of order $K$ (see Figure A.7). Between breaking the symmetry or supplying a large momentum change, any scatterer must be either exceptionally strong or extremely sharp to cause backscattering. The pseudospin, in effect, protects metallic nanotubes from the effects of anything but the strongest disorder\textsuperscript{56,57}.

For semiconducting tubes, the quantization occurs away from the $K$-point, which means that the phase of right-movers and left-movers won’t differ precisely by $\pi$. In other words, the pseudospinors overlap, which means backscattering can occur without the protection afforded metallic tubes. Near the subband bottom the states are most vulnerable to disorder, while at high energies the symmetry is effectively re-established, and backscattering can be suppressed. The pseudospin distinction between metallic and semiconducting nanotubes will play a key role in experiments.

A.3 Curvature

One question looms large: nanotubes are tubes, not flat graphene, so how does their curvature affect the simple arguments above? To understand the effect, we must look back at the bonding structure of carbon in graphene. The electronic states we included in the tight-binding model are the $\pi$-orbitals of the $sp^2$-bonded carbon, which consist of a raw $p_z$ atomic orbital taking part in the ‘resonant’ states that result in electrical conductivity. The low-energy valence bonds between carbon atoms consist of $\sigma$-orbitals, which guarantee the atomic structure of the lattice. These $\sigma$-states are a hybridized mix of $s$, $p_x$, and $p_y$ atomic orbitals. In flat graphene, the $\pi$ states are completely orthogonal to the $\sigma$ states, granting us our elegantly simple tight-binding theory\textsuperscript{55}.

Curvature breaks this orthogonality: the $p_z$ of $\pi$ can now mix with the $p_x$ and $p_y$ of $\sigma$, and by
extension the $s$. To see this, consider the introduction of curvature into a zigzag nanotube, as illustrated in Figure A.8a. The $p_z$ orbitals in flat graphene are oriented perpendicular to both the local lattice plane at the atomic site and the bond directions to all nearest-neighbors. With curvature, the $p_z$ orbital, here labeled $p_r$, points in the radial direction at the position of the atomic site but is no longer perpendicular to all of the nearest-neighbor bond directions. This implies that a complete theory must include additional tunneling terms from $\pi$ to $\sigma$ states\textsuperscript{140–143}.

This modification of nearest-neighbor tight-binding can be expected to lead to changes in the spectrum. The full treatment is outside the scope of the discussion here, but nonetheless we can achieve a functional understanding. First, we observe that the perturbation to nearest-neighbor tunneling depends on the angle that the bond makes with its old, flat version. This angle is roughly the bond length, $l$, divided by the nanotube radius, $R$, giving us a small expansion parameter $l/R$. Corrections to the flat case should then go roughly as $\cos(l/R) \approx 1 - \frac{1}{2} \left( \frac{l}{R} \right)^2 + \ldots$ Now consider that the form of the graphene Hamiltonian in Equation A.3 gives Dirac points whose positions are determined by
the relation

\[ \sum_i t_i e^{i \delta_{ij} \mathbf{k}} = 0 \]  

(A.25)

where we’ve generalized the nearest-neighbor tunneling \( t_i \) to be different for each of the three bonds\(^{142}\). We expect that in the presence of curvature these three parameters will, in general, be different from each other, receiving a small curvature correction. As a result, the \( \mathbf{k} \) at which Equation A.25 holds will change, which means that the Dirac points \( \mathbf{K} \) and \( \mathbf{K}' \) will move with respect to the quantization condition. The correction can be written as \( \mathbf{K}^c = \mathbf{K} + \Delta \mathbf{k}^c \), and we can obtain the shift \( \Delta \mathbf{k}^c \) by expanding Equation A.25.

The correction to the energy gap is obtained by considering how \( \Delta \mathbf{k}^c \) shifts the Dirac points away from the quantization lines, in the direction of \( k_\perp \). Shifts along the quantization lines, in the direction of \( k_\parallel \) have no effect. The result for the gap correction is

\[ \Delta E_g = \frac{43 \, 2 \, meV}{d^2 \, [nm]} \cos^3 \theta \]  

(A.26)

where the shift in \( k \)-space is opposite for the \( \mathbf{K} \) and \( \mathbf{K}' \) points\(^{40}\). This result changes our previous picture substantially: all nanotubes now have a gap, excluding only those with \( \theta = \pi / 6 \). These are armchair tubes for which the momentum shift \( \Delta \mathbf{k}^c \) is along the quantization line (and therefore makes no substantial change to the low-energy picture).

These results show that nearly all nanotubes will have a gap, and true metallic tubes are rare. Only armchair tubes do not inherit the curvature-induced gap. For every other nanotube, a gap correction of this size means we can expect small-bandgap behavior for \( T \leq 4K \).
A.4 Curvature with Spin: Spin-Orbit Coupling

Nanotube curvature is also responsible for a far more subtle but enormously interesting and useful effect, the spin-orbit coupling. So far, we have ignored the spin of the electron in all of our calculations. Presumably, it would enter the physics through a simple two-fold degeneracy of the energy spectrum; with the isospin, this makes a total fourfold degeneracy. But this is, in fact, a simplification. Even on the atomic level, the spin degeneracy is not exact: the carbon atom, like every atom, possesses a coupling between its spin and orbital degrees of freedom. This is fundamental: an electron moves inside the electric field of the nucleus, and thus experiences a magnetic field proportional to $v \times E$, and this magnetic field couples to the electron spin via its magnetic moment $m$ through the coupling $m \cdot B$.

As elegant as this story is, it isn’t complete: relativity breaks the symmetry between the reference frames (nuclear or electron at rest) and forces us to use a full relativistic expansion to derive the correct result. Skipping the derivations, a purely radial confinement potential results in a spin-orbit Hamiltonian

$$H_{SO} = \frac{\hbar^2}{4m^2c^2} \left( \nabla \cdot V \times \mathbf{p} \right) \cdot \mathbf{\sigma}$$

$$= \frac{1}{2m^2c^2} \frac{1}{r} \frac{dV}{dr} \left[ \frac{L_+ \sigma_- + L_- \sigma_+}{2} + L_z \sigma_z \right]$$

$$\equiv \Delta \left[ \frac{L_+ \sigma_- + L_- \sigma_+}{2} + L_z \sigma_z \right]$$

(A.27)

with $m$ the electron mass, $c$ the speed of light, $V$ the atomic potential, $\mathbf{p}$ the electron momentum, $\mathbf{\sigma}$ the electron spin, and $L$ a component of the electron angular momentum operator. The atomic confinement potential is

$$V(r) = \frac{Ze^2}{4\pi\epsilon_0 r}$$

(A.28)
with \( Z \) the atomic number and \( \epsilon_0 \) the vacuum permitivity. We use the hydrogen wavefunctions to compute

\[
\left\langle \frac{1}{r^3} \right\rangle = \frac{Z^3}{a_B^3} \frac{1}{n^3} \frac{1}{\ell(\ell + \frac{1}{2})(\ell + 1)}
\]

(A.29)

with \( a_B \) the Bohr radius, \( n \) the principal quantum number and \( \ell \) the eigenvalues of the orbital angular momentum operator. With all of this, we can get the atomic spin-orbit energy scale

\[
\Delta = \frac{Z^4 m^2 c^2}{4} \frac{1}{n^3} \frac{j(j + 1) - \ell(\ell + 1) - \sigma(\sigma + 1)}{\ell(\ell + \frac{1}{2})(\ell + 1)}
\]

(A.30)

with \( \alpha \) the fine structure constant and \( j \) the eigenvalues of the total angular momentum operator, \( J = L + \sigma \). For carbon, we are interested in the \( p \)-orbital electrons, for which \( n = 2, \ell = 1 \), and due to screening of the nuclear charge by the two core \( s \) electrons, \( Z \approx 6 - 2 = 4 \). Considering the splitting between the \( j = 1/2 \) and \( j = 3/2 \) states, this gives an atomic splitting of \( \Delta_{SO} \approx 12\text{meV} \).

Comparing this with the typical level spacing of atoms, or even the spin-orbit coupling of heavier metals in the eV range, we see it is very small. This is principally due to the sensitive dependence on the atomic number \( Z \). With carbon a relatively light atom, the spin-orbit coupling in the bare atom is therefore small.

The situation in flat graphene is not very different, as the symmetries of the problem don’t introduce any extra coupling beyond the atomic case. Rewriting the spin-orbit Hamiltonian of Equation A.27 using second-quantized operators gives

\[
H_{SO} = \Delta \left[ c^\dagger_{x\uparrow} (c_{x\downarrow} + ic_{y\downarrow}) - c^\dagger_{x\downarrow} (c_{x\uparrow} + ic_{y\uparrow}) + i(c^\dagger_{y\downarrow} c_{y\uparrow} - c^\dagger_{x\downarrow} c_{x\uparrow}) + h.c. \right]
\]

(A.31)

where the creation and annihilation operator convention is now for atomic orbitals \( x, y, \) and \( z \), with spin states \( \uparrow \) and \( \downarrow \), and \( h.c. \) is the hermitian conjugate. Now we can see that spin-orbit coupling allows a \( p_z \) electron to hop to the \( p_x \) and \( p_y \) orbitals at the cost of a spin flip, at which point it is free
to mix among $p_x$ and $p_y$. The only way that such a process can affect nearest-neighbor hopping in the graphene lattice is as follows: one spin-orbit hop ($p_z$ to $p_x$ and $p_y$), two orbital hops (from $p_x$ to $p_y$), and a second spin-orbit hop to get back to $p_z$, as illustrated in Figure A.9a. Thus, spin-orbit affects hopping only at second order\textsuperscript{144}. We saw that the energy scale is small, which means a second-order process is highly unlikely and plays a tiny role, on the $\mu eV$ scale.

![Figure A.9: Nanotube curvature and spin results in spin-orbit coupling. a) Hopping process generating spin-orbit coupling in the flat graphene lattice. It is second order in the atomic spin-orbit operator, keeping it at $\mu eV$ scale and negligible. b) Curvature in the nanotube allows spin-orbit to enter at first order due to allowed mixing between $p_z$ and $\sigma_{xy}$ orbitals ($p_r$ and $p_t$ orbitals, respectively, in terms of the radial and transverse directions of the tube). Adapted from\textsuperscript{140,144}.](image)

In the nanotube, curvature changes the picture. Curvature breaks the reflection symmetry of the graphene sheet and allows an electron to hop directly from $p_z$ to $p_x$ and $p_y$, as discussed in section A.3. This means that just one spin-orbit hop is necessary to complete a full nearest-neighbor $p_z$-to-$p_z$ process, thereby changing the tight-binding picture to first order in the energy scale $\Delta$ (illustrated in Figure A.9b). The additional tunneling modifies the nanotube spectrum by shifting the K and K’ points, as described by Equation A.25, but this time in a spin- and isospin-dependent way.

How does spin-orbit coupling modify the spectrum? The coupling between the electron spin and the circulating isospin can be seen in two ways. An electron moving around the tube sees a magnetic flux due to its spin, resulting in an extra, Aharonov-Bohm-like phase shift that modifies the
quantization condition. This shifts $k_\perp$, causing the K cone to shift left (right) for spin up (down), and opposite at the K’ cone (shown schematically in Figure A.10c). On the other hand, as the electron spin moves around the tube, it sees the magnetic field generated by the circulating charge motion. This acts like an isospin-dependent Zeeman field, splitting spin up and spin down for K electrons, but splitting in the opposite direction for K’ electrons (shown schematically in Figure A.10d). The Zeeman shift therefore shifts spin up (down) electrons down (up) in energy.

From these two perspectives, we expect that spin-orbit will contribute two Hamiltonian terms: one which is diagonal in the sublattice basis (Zeeman-like), and one which is off-diagonal (orbital-like). These terms are calculated to give an effective spin-orbit Hamiltonian (in the A-B sublattice basis)

\[
H_{SO} = \begin{pmatrix}
\sigma_\eta k_{SO}^Z & \sigma_\eta k_{SO}^O \\
\sigma_\eta h_{SO}^O & \sigma_\eta h_{SO}^O
\end{pmatrix}
\]  

(A.32)
where $\sigma = \pm 1$ is the spin eigenvalue along the nanotube axis, $\eta = \pm 1$ is the isospin eigenvalue, and the Zeeman spin-orbit term is

$$\hbar Z_{SO} = \frac{\alpha Z}{d} \Delta \cos 3\theta \quad (A.33)$$

while the orbital spin-orbit term is

$$\hbar O_{SO} = \frac{\alpha O}{d} \Delta \quad (A.34)$$

where $\alpha_Z$ and $\alpha_O$ are length scales which depend on microscopic details (and are allowed to be negative). The curvature-induced origin of the spin-orbit coupling survives as the inverse dependence on nanotube diameter $d$, and both are proportional to the atomic spin-orbit energy scale $\Delta$. The orbital term depends only on these parameters. But the Zeeman term inherits a dependence on chirality, similar to that of the curvature-induced gap in Equation A.26. This dependence means it can be larger or smaller than the orbital term. Thus we can expect different nanotubes to have very different spectroscopic signatures arising from the spin-orbit coupling.

In addition to depending on chirality, the spin-orbit coupling depends on energy. We saw in section A.2 that a semiconducting nanotube possesses an orbital velocity that decreases with increasing energy as the hyperbolic dispersion approaches linearity and $\nu_{orb} = \partial E/\partial k_{\perp}$ vanishes. The associated magnetic moment vanishes with increasing energy, so the spin-orbit coupling also vanishes, but this is the case only for the orbital term. The Zeeman term enters the spectrum in an energy-independent way. This can be seen by diagonalizing the nanotube Hamiltonian with spin-orbit, as explicitly computed in the supplementary information of.

**A.5 External Perturbations: Mechanical Strain**

Like other solid state systems, nanotubes can be subjected to external mechanical forces. An applied stress leads to a strain which modifies the geometry of the atomic lattice. The bond lengths change
in response to the applied mechanical force, and as a result the tunneling parameters in the tight-binding model change. The zero-energy condition expressed in Equation A.25 implies that the Dirac points will shift in response, and if that shift is in the direction of \( k_\perp \), the bandgap will change. Indeed, the curvature described in section A.3 can be thought of as a circumferential mechanical stress.

There are several mechanical deformations to consider, but here we mention one experimentally relevant type of external stress applied along the nanotube axis. The resulting shift of the wavevector is

\[
\Delta k_\perp = \zeta (1 + \nu) \cos \theta
\]  

(A.35)

where \( \zeta \) is the applied axial stress and \( \nu \) is the Poisson’s ratio for the nanotube\(^{149} \). The resulting gap correction is

\[
\Delta E_g = \begin{cases} 
2 \hbar v_F \zeta (1 + \nu) \cos \theta & \text{if } p = 0, 1 \\
-2 \hbar v_F \zeta (1 + \nu) \cos \theta & \text{if } p = -1
\end{cases}
\]  

(A.36)

where \( p \) is the wrapping parameter. The dependence on \( p \) can be understood as arising from the different positions of the lowest subbands relative to the Dirac point tubes. A metallic \( p = 0 \) tube always gains a gap from the applied stress, while the semiconducting tubes with \( p = \pm 1 \) either grow or shrink the gap depending on whether the lowest subbands moves toward or away from the Dirac point.

Like the curvature, the gap correction depends on chirality: for zigzag tubes it is maximal, while for armchair tubes it vanishes. This means we can expect any residual axial strain in a real nanotube device to introduce gaps for most nanotubes, another reason to expect most tubes to be gapful at low temperatures.
A.6 External Perturbations: Magnetic Fields

To probe the magnetic structure of charge carriers in nanotubes, we can apply a magnetic field and measure how various physical properties evolve with the field. In nanotubes a magnetic field has two main effects: the spin Zeeman effect and an orbital Aharonov-Bohm effect.

The electron spin couples to the external magnetic field following the well-known Zeeman Hamiltonian

$$H_Z = -\frac{1}{2} g \mu_B B \cdot \sigma$$  \hspace{1cm} (A.37)

with $g$ the gyromagnetic ratio of the electron, $B$ the external magnetic field, and $\sigma$ the spin of the electron. No surprises here.

For magnetic fields parallel to the nanotube axis, electrons in the nanotube orbit a magnetic flux. This leads to a nanoscopic Aharonov-Bohm interference effect. Over the course of an orbit around the tube, an electron will acquire a phase $\phi_{AB}$, which modifies the total phase of the wavefunction around the tube. We should then include this extra phase in the quantization condition for wrapping Equation A.18 as follows

$$k \parallel |C| = K \cdot C + \phi \parallel + \phi_{AB}$$  \hspace{1cm} (A.38)

The $AB$ phase is related to the flux through the tube by

$$\phi_{AB} = 2\pi \frac{\Phi}{\Phi_0}$$  \hspace{1cm} (A.39)

with $\Phi = B \parallel A$ the magnetic flux through the tube and $\Phi_0 = h/e$ the flux quantum. The shift in wavevector is then given by

$$\Delta k_{\perp, B\parallel} = \frac{ed}{4\hbar} B\parallel$$  \hspace{1cm} (A.40)

The parallel magnetic field effectively shifts all of the quantization lines in k-space along the perpen-
Nanotubes subjected to axial magnetic fields. a) An axial magnetic field pierces the nanotube. b) Spectrum of a semiconducting nanotube. Left: the two lowest subbands slicing the two Dirac cones of graphene. Right: The resulting lowest subbands. c) Spectrum of the semiconducting nanotube in (b) subject to an axial magnetic field, which shifts the two subbands, resulting in the spectrum shown on the right. Adapted from 150.

dicular direction. We can expect therefore that the nanotube bandgap will change with a magnitude

$$|\Delta E_g| = \hbar v_F \Delta k_{\perp, B} = \frac{e v_F d}{4} B_\parallel \approx 250 \text{\mu eV} \times d[\text{nm}] B_\parallel [\text{T}] \tag{\text{A.41}}$$

Metallic nanotubes will develop a gap, while in semiconducting tubes the gap may grow or shrink, depending on the wrapping parameter $\rho$ and the direction of the magnetic field. Importantly, this behavior is modified for finite nanotubes where spatial quantization leads to a level spacing, in which case the magnetic field always first closes the gap.

Figure A.11: Nanotubes subjected to axial magnetic fields. a) An axial magnetic field pierces the nanotube. b) Spectrum of a semiconducting nanotube. Left: the two lowest subbands slicing the two Dirac cones of graphene. Right: The resulting lowest subbands. c) Spectrum of the semiconducting nanotube in (b) subject to an axial magnetic field, which shifts the two subbands, resulting in the spectrum shown on the right. Adapted from 150.
A.7 Interactions

Why are interactions expected to be strong in nanotubes? We can motivate this prediction with a simple picture, which we develop in more detail from that discussed in section 1.2. For free, massive particles, the kinetic energy scales as $U_{\text{kin}} \sim \frac{\hbar^2 k^2}{2m}$. We populate such a band with fermions sporting a Fermi-Dirac distribution with a Fermi wavevector $k_F$ (see Figure A.12a). We assume a system of size $L$, which implies a mode separation in 1D momentum space of $2\pi/L$. Assuming a state degeneracy of $\tau$, we can relate the density to the Fermi wavevector by simply counting the number of states $N$, via $N = \tau 2k_F/(2\pi/L)$, giving a density $n = \tau k_F/\pi$ in 1D, which we can use to relate $U_{\text{kin}}$ of an electron near the Fermi level to the density.
Now we add a bare Coulomb interaction to these 1D fermions. The interaction energy scales like $U_{\text{int}} \sim \frac{e^2}{\epsilon r}$, with $\epsilon$ the dielectric constant and $r$ the separation between two particles in the system. The typical particle spacing in 1D is just the inverse density, so that $U_{\text{int}} \sim \frac{e^2}{\epsilon n}$. The ratio of interaction to kinetic energy is then

$$U_{\text{int}} / U_{\text{kin}} = 1 / na_B'$$

with the effective Bohr radius given by $d_B = \frac{\pi}{2} \frac{\hbar^2 \epsilon}{2me^2}$. At high densities, the kinetic energy is much larger than the interaction, since the Fermi-Dirac distribution automatically promotes electrons to high energies. Since most thermodynamic properties are determined by electrons around the Fermi level, this means the kinetic energy dominates the physics, which is effectively of non-interacting particles.

At low densities, the argument is reversed: interaction dominates over the kinetic energy. Indeed, if the density is sufficiently low, the system is well-described by only the bare Coulomb interaction. The energy of the system is then minimized by reducing the interaction, which, in a semi-classical sense, we obtain by separating the electrons as far as possible from one other. In a system of finite size, this immediately implies crystallization. The notion of an electron crystal, a true ‘electron solid’, was so fascinating when it was theoretically predicted in 1934 as to acquire a name: the Wigner crystal, a long-sought state of matter that remains to be directly observed.

The story is dramatically altered when we consider a linear band for which the kinetic energy scales as $U_{\text{kin}} \sim \hbar v_F k$ (see Figure A.12b). In this case, $U_{\text{int}} / U_{\text{kin}} = \frac{\tau e^2}{\pi \epsilon \hbar v_F}$, a constant independent of density. For linear bands, therefore, we cannot drop any terms to make simplifying assumptions, and must consider the full interacting Hamiltonian.

In nanotubes, the single-particle picture leads us to the result that most nanotubes should possess a hyperbolic dispersion, where $U_{\text{kin}} \sim \hbar v_F \sqrt{k_\parallel^2 + k_\perp^2}$ (Figure A.12c; see A). At low energies close to the band bottom, this is a roughly parabolic dispersion, while at higher energies the dispersion
approaches linearity. In the linear regime, the constant $U_{\text{int}}/U_{\text{kin}} \sim 3$, indicating that interactions play a significant role in nanotubes at all energies.

Detailed theoretical treatments predict that interactions have a decisive effect on determining electronic properties of carbon nanotubes\textsuperscript{24,58–60}. Short-range interactions are predicted to lead to Mott gaps\textsuperscript{58,61,62}. Long-range Coulomb interactions are predicted to lead to Luttinger liquid behavior with a small Luttinger parameter $g \approx 0.2$ and spin-charge separation effects\textsuperscript{24}, as well as modified tunneling exponents and spin correlations\textsuperscript{59}. In confinement potentials, interactions may also lead to the formation of Wigner molecules via Wigner crystallization\textsuperscript{60}.

The observation that linear bands prevent the reduction of the Hamiltonian into purely kinetic or interaction energies points to a severe issue. Importantly, this is a special property of one-dimensional systems: the density-dependence in higher dimensions never leads to this peculiar situation. This may seem surprising: shouldn’t a one-dimensional system be simpler than its higher-dimensional cousins? Indeed, early in the 20th century, it was believed that this should be the case, and the pioneers of many-body quantum mechanics resorted to one dimension to simplify the theory and more easily obtain insight into higher-dimensional problems. It was realized, however, that one-dimensional systems are separate creatures, possessing unique physical properties.

To clarify the picture, we will make a small but decisive observation. In one-dimension, particles are no longer precisely indistinguishable. This may sound like blasphemy, but is formally true, as we’ll motivate following\textsuperscript{20}. Consider defining a 'labeling function' $\phi_l$ which takes the value $\phi_l(x_i) = 2\pi i$ at the position of the $i$th particle $x_i$. Such a function can be uniquely defined to be smooth and continuous in one dimension. In contrast, in higher dimensions such a function is ill-defined, non-local and multi-valued. This amounts to labeling our previously indistinguishable particles, which makes the notion of fermions and bosons superfluous: our particles can now be described by either type of quantum creature, with the transformation between them a function of $\phi_l$. The meaning of this will hopefully become clearer below. The density of our 1D system of particles can be written as
Figure A.13: Describing 1D particles with a smooth labelling field. a) The labelling field for a crystalline 1D arrangement of particles is linear. b) For any non-crystalline arrangement, the labelling field can be smoothly adapted to describe the density (see main text). Adapted from 20.

$$\rho(x) = \sum_i \delta(x - x_i)$$  \hspace{1cm} (A.43)

To express this in terms of $\phi_l$, we use the rule for composite delta functions and form a new sum

$$\rho(x) = \sum_n \nabla \phi_l(x) \delta(\phi_l(x) - 2\pi n)$$  \hspace{1cm} (A.44)

where we have assumed (legally) that $\phi_l$ is monotonically increasing. This is a Dirac comb, for which the Fourier series allows us to write this as

$$\rho(x) = \frac{\nabla \phi_l(x)}{2\pi} \sum_m e^{im\phi_l(x)}$$  \hspace{1cm} (A.45)

A linear $\phi_l$, corresponding to evenly distributed particles (i.e. a periodic crystal), gives a constant times a series of harmonics. To retrieve a constant density from this crystal, we can integrate over interparticle distances greater than the typical inverse density, eliminating higher harmonics. Taking
this to the extreme, we can integrate out all harmonics (i.e. all $m \neq 0$) to obtain

$$\rho(x) \sim \nabla \phi_l$$  \hspace{1cm} (A.46)

In other words, the density is entirely determined by the labelling field, whose gradient is the new dynamic variable. Higher harmonics will introduce terms non-linear in $\phi_l$, as one sees by expanding the exponential. Discarding higher harmonics therefore corresponds to constructing a linear theory. Linear dispersions are typically found for collective degrees of freedom, such as phonons or hydrodynamic waves, which are well-described by quantized harmonic oscillators. How does a theory of 1D electrons connect to fluids? In one dimension particles cannot avoid interacting with their neighbors, so any movement of an individual particle necessarily involves motion of the entire group of particles. In higher dimensions, this isn’t the case: other particles simply ’get out of the way’, illustrated in Figure 1.5a. This is the screening that takes place in theories of simple metals or semiconductors, as described by the Fermi liquid theory, which connects the resulting quasiparticle with the original screened electron via renormalized parameters. A full discussion of the breakdown of Fermi liquid theory in one dimension is beyond the scope of this document. Here, we will note the key results and what this means for nanotubes.

The 1D theory linear in the field gradient is the Tomonaga-Luttinger liquid. A linear theory consists of a Hamiltonian with two terms quadratic in each of two conjugate variables. For a harmonic oscillator, this is momentum and position. For the one dimensional Luttinger theory (we now use the common abbreviation), the two conjugate variables are one field describing the density fluctuations (related to the labelling field defined above), and one field describing currents of the density. The Hamiltonian (without interactions) looks like

$$H_{1D} = \frac{\nu_F}{2\pi} \int dx \left[ (\partial_x \phi)^2 + (\partial_x \theta)^2 \right]$$  \hspace{1cm} (A.47)
We won’t derive this Hamiltonian, nor the proper commutation relations that prove the excitations are bosonic. We only observe that the harmonic modes here are collective excitations of the 1D electron system, in some sense like plasmonic oscillations of the electron density. To include interactions, one can exploit the fact that in 1D, the two branches of the Fermi “surface” are decoupled. Excitations can occur only with small momentum transfer \( q \sim 0 \) or large momentum transfer \( q \sim 2k_F \) Figure A.14. That is, only forward-scattering or back-scattering are allowed. If we discard the \( q \sim 2k_F \) contribution and assume the \( q \sim 0 \) component is finite (consistent with a screened Coulomb interaction), one can derive an elegant form of the interacting Hamiltonian

\[
H_{LL} = \frac{v}{2\pi} \int dx \left[ g (\partial_x \phi)^2 + \frac{1}{g} (\partial_x \theta)^2 \right]
\]  

(A.48)

where \( v \) is a renormalized Fermi velocity and \( g \) is a new parameter, where both are determined by the strength of the interaction \( 20 \). Here, \( g < 1 \) for repulsive interactions and \( g > 1 \) for attractive interactions (\( g > 0 \) always). Again, the commutation relations are bosonic. This is the Luttinger model in its full glory, which gives a simple way to obtain correlation functions even in the presence of interactions. It is instructive to note the density-density correlation function for this system, which looks
\[ \langle \rho(x)\rho(0) \rangle \sim A_0 + A_1 \frac{g}{x^2} + A_2 \cos(2k_F x) \left( \frac{1}{x} \right)^2 g + A_3 \cos(4k_F x) \left( \frac{1}{x} \right)^8 g + \ldots \] (A.49)

where \( \alpha \) is a short-length-scale cutoff and \( A_i \) are dimensionful and non-universal prefactors. The first two terms are comparable to that obtained for Fermi liquids, renormalized by the interaction parameter \( g \). The \( A_2 \) term, the \( 2k_F \) oscillation, inherits a modified decay compared to that of a Fermi liquid. For repulsive interactions, it decays with a power slower than 2 by a factor of \( g \). The \( A_4 \) term, a \( 4k_F \) oscillation, decays much faster than the rest. Such a term arises when including higher-order terms in the expression for the density, as in Equation A.44.

While there are numerous insights to be gleaned from the Luttinger model, let us pause to ask is it sensible to use this modified form of the interaction? The Coulomb interaction has a Fourier transform which goes as \( \ln(q) \) for small \( q \), so small momentum dominates. However, at \( q \to 0 \), the Fourier components diverge, so it is inappropriate to assume a \( q \)-independent interaction. In the nanotube, the construction of a 1D theory involves integrating out the small-scale, high-energy structure. Electrons on the nanotube are smeared over the circumference, and therefore have a lower probability of interacting at short range. If we take an on-site atomic interaction of \( U \), the effective 1D interaction this induces is \( u = U/N \), where \( N \) is the number of atoms in the circumference of the tube. In other words, long-range interactions, if present, will dominate the physics. The precise form at long-range, i.e. the small \( q \)-dependence, is therefore crucial.

To see the effect of the unscreened Coulomb interaction, we can keep the \( q \)-dependence by taking \( v \to v(q) \) and \( g \to g(q) \). The resulting density-density correlations now look like

\[ \langle \rho(x)\rho(0) \rangle \sim \ldots + A'_2 \cos(2k_F x) \frac{1}{x} e^{-C\sqrt{\log x}} + A'_4 \cos(4k_F x) e^{-C\sqrt{\log x}} + \ldots \] (A.50)
where the prefactors are modified and $C$ is a parameter depending on $\nu$ and $g^{20,32}$. Note that the factor of 2 for spin is included, but we ignore all effects of this besides the degeneracy. The slowest-decaying term is now the $4k_F$ term, which decays slower than any power law. This $4k_F$ periodicity is a modulation at the wavelength of the typical interparticle spacing. Thus, with the full power of the Coulomb repulsion at long range, the electrons indeed try to form a Wigner crystal. The slow decay preserves the absence of a true long-range order, avoiding a phase transition, which is outlawed in 1D.

In real nanotube devices, the interaction will have a short-range cutoff at the scale of the nanotube diameter, as well as a long-range cutoff at the scale of the distance to a nearby metal which screens the interaction. We have not yet included spin, besides for a simple factor-of-two degeneracy. In 1D, spin-charge separation leads to a separate set of excitations with their own ordering. In nanotubes, we must take care of the additional isospin as well. Furthermore, additional predictions and point to the possible existence of a Mott-insulator like gap $^{24,49}$. All of this is subject to the assumption of a 1D system, which must be tested in a real device at finite temperature, with coupling to the 3D environment of the bulk metallic leads.

We have now covered much of the basic groundwork for understanding the behavior of carbon nanotubes.
Overview of Electrons in Quantum Dots

In this thesis, we probe properties of interacting electrons by confining them to a suspended carbon nanotube. We use two main techniques: measuring the transport properties of the suspended nanotube device and measuring the local charge distribution in the nanotube with an external charge detector. To understand these experiments, we have to know how electrons behave when confined at low temperatures.
B.1 Coulomb Blockade and Quantum Dots

Real electrons are quantized packets of charge, and interact with each other via the Coulomb potential. Yet classical electrodynamics and Ohm’s law ignore this. When does charge quantization become apparent?

Consider a conducting island connected to the outside world via a lumped resistance $R$ and a lumped capacitance $C$. There could be more than one external connection, but we can always compute these net lumped elements. Together, these define the time it takes to charge or discharge the island, the timescale $\tau = RC$. To charge the island with charge $Q$, we require a voltage $V = Q/C$. For a single electron, this is $V = e/C$. The energy it takes to charge the island with a single electron is therefore $U = \int eV dN = e^2/2C$, the charging energy. In classical electrodynamics, the charging energy is assumed to be negligible, since Maxwell requires only a continuous charge density for which we ignore the granular nature of the charge.

The quantization of charge into units of $e$ should therefore become apparent when the charging energy is in some sense large. Quantum mechanically, this happens when the uncertainty relation is satisfied such that $U\tau > \hbar$, or the time an electron spends on the island should be long enough that the uncertainty in energy is less than $U$, such that $U$ is large and well-defined. Substituting the definitions of $\tau$ and $U$, this gives us a condition on $R$: $R > \hbar/e^2$, or equivalently for the conductance, $G < e^2/\hbar$. The fundamental quantity given by $\hbar/e^2$ is the quantum of resistance, expressing the resistance of a single quantum channel in the Landauer sense. That is, for resistances larger than that of a single quantum channel, we can trap an electron long enough that the charging energy is sharply defined quantum mechanically. This also tells us when the opposite limit of non-interacting waves in a Fabry-Perot-like cavity is applicable, namely when $R \gtrsim \hbar/e^2$.

The average time an electron spends on the island $\tau$ immediately gives us the average rate at which electrons pass through the island $1/\tau$. This frequency defines an energy scale $\Gamma = \hbar/\tau$. The
uncertainty relation then gives us $\Gamma < U$. In other words, the condition on resistance being greater than the resistance quantum is equivalent to the charging energy being larger than the coupling to the outside world.

Figure B.1: Coulomb blockade of an isolated conductor. a) A conducting island is connected via a lumped resistance $R$ and lumped capacitance $C$ to external conducting leads. b) The classical charge on the island, $Q(t)$, charges for $t < 0$ and discharges for $t > 0$. For discrete charges, we instead write the probability $P_e(t)$ for an electron to be on the island at $t = 0$. The time constant $\tau = RC$ shows up as the decay constant of the exponentials. c) Fourier transform of the charging-discharging double-sided exponential gives a Lorentzian, whose width is proportional to $\Gamma = 1/\tau$.

What is the meaning of $\Gamma$? For a classical capacitor, $\tau = RC$ is the time for the classical charge $Q$ on the capacitor plates to be reduced by a factor of $1/e$ (the number, not the charge). For quantized charges, this exponential decay relates the probability that an individual electron remains on the capacitor after time $t$, with $P(t) = \exp(-t/\tau)$. Fourier transforming this distribution into energy space gives a Lorentzian of width $h/\tau = h\Gamma$ (technical note: the Lorentzian results only when Fourier transforming the two-sided exponential $\exp(-|t/\tau|)$, which arises when correctly including the physical process of charging the capacitor before discharging it). The lifetime therefore determines the width of a resonance corresponding to the addition of an extra electron to the capacitor formed by the island.

The inequality $\Gamma < U$ therefore means that when the charging energy is larger than the linewidth, then we expect the emergence of resonances corresponding to the addition of individual extra electrons to the island, called Coulomb peaks. In other words, when the spacing between peaks (the
charging energy) is larger than their width (the lifetime), then we observe so-called Coulomb oscillations of current. If $\Gamma \ll U$, then the peaks are distant from each other and in between the peaks we enter the fully-developed Coulomb blockade regime, where individual electrons can be added to the island only at the specific points in energy space given by the charging peaks. If the inequality is not obeyed, then the resonances overlap and the resonances are not distinguishable.

At a Coulomb peak, we can also compute the current through the island, $I = N e / \tau$, where $N$ is the number of electrons allowed on the dot. $N$ is determined by the ratio of two energy scales. The first is given by the energy bandwidth of states allowed to enter the dot, which experimentally is the bias voltage $V$. This is divided by a second energy scale, fixing the number of states within the energy bandwidth. If we consider only the charging energy $U$, then $N = eV / U$ and using $G = dI / dV < e^2 / h$, we obtain the relation $\Gamma < U$, as above.

The limit we have discussed is called the single electron transistor (SET), which is a metallic island for which the resistance and capacitance obey the above conditions. The charging behavior of a SET can be observed so long as the temperature of the measurement $T$ is small enough, $k_B T < U$. This is analogous to the condition on the lifetime, or intrinsic linewidth given by $\Gamma$, in the sense that the Gaussian broadening due to temperature should be less than the distance between the resonant peaks.

In the case that the island is very small or made of a semiconductor, then we should also consider its level spacing $\Delta \epsilon$. Now we must take care of the four energy scales $U$, $\Delta \epsilon$, $\Gamma$ and $k_B T$. If $\Delta \epsilon, k_B T > U$, we can drop $U$ and write $N = eV / \Delta \epsilon$. Then the condition on $G$ gives $\Gamma < \Delta \epsilon$, a regime where the level spacing dominates transport. This is a difficult regime to access for our devices, since it implies a very small island which is simultaneously very well-coupled to the environment, and is more accessible by atomic-scale objects.

On the other hand, if $k_B T < U$, such that we are in the Coulomb blockade regime, and simultaneously $U \lesssim \Delta \epsilon$, then we enter a regime where single-electron charging determines the transport,
but each Coulomb resonance has access to an individual quantum level of the island. This is the quantum dot regime, and we now turn to see how the features of a quantum dot show up in transport and detector experiments.

B.2 Electrostatics of Quantum Dots with Multiple Gates

We want to understand how a set of voltages applied to different metallic electrodes affects a quantum dot, so that we can extract physical parameters of the dot from experiments. We assume that the quantum dot is defined independently of the applied voltages, so that all capacitances and level spacings are independent of the voltages.

We begin by considering a system of \( N \) conductors with a capacitance matrix \( C_{ij} \), to which we can apply voltages \( V_j \), to obtain a charge \( Q_i \) on each conductor given by

\[
Q_i = \sum_{j=0}^{N} C_{ij} V_j \tag{B.1}
\]

One of these conductors is our quantum dot (say \( j = 0 \)), which can possess a discrete charge \( Q_D \) along with the induced charge from the surrounding conductors. The potential of the dot \( V_D \) is

\[
V_D = \frac{1}{C_\Sigma} \left( Q_D - \sum_{j=1}^{N} C_{Dj} V_j \right) \tag{B.2}
\]

where the total capacitance of the dot \( C_\Sigma \) is given by

\[
C_\Sigma = - \sum_{j=1}^{N} C_{Dj} \tag{B.3}
\]

and we remember that off-diagonal elements of the capacitance matrix, the mutual capacitances, are negative thanks to Ben Franklin.
We calculate the energy of the dot \( U(N) \), where \( N \) is the discrete number of charges on the dot defined through \( Q_D = -eN \), by integrating

\[
U(N) = \int_0^{-eN} V_D(Q_D) dQ_D = \frac{e^2 N^2}{2C_\Sigma} + \frac{eN}{C_\Sigma} \left( \sum_{j=1}^{N_f} C_D j V_j \right) 
\]  

(B.4)
Let us denote the external charge induced by the external voltage sources (the sum on the right hand side) as $-Q_{\text{ext}}$, which is a continuous number reflecting the amount of charge that would be induced in the dot if it were a perfect metal with no charging energy. We can express the energy as

$$U(N) = \frac{e^2 N^2}{2C_\Sigma} - \frac{eN}{C_\Sigma} Q_{\text{ext}} = \frac{(eN - Q_{\text{ext}})^2}{2C_\Sigma} - \frac{Q_{\text{ext}}^2}{2C_\Sigma}$$  \hspace{1cm} (B.5)

The two terms of the right-hand side can be understood as the electrostatic energy of the dot, $U_D$, and the electrostatic energy of the electrode used to charge it. A sketch of $U_D(Q_{\text{ext}} = C_{Dj} V_j)$ is shown in Figure B.3a. For every value of the discrete quantity $eN$, the island has the energy dependence of a capacitor whose charge is $Q_{\text{ext}}$, which is a smoothly varying quantity called the 'control charge'. When $Q_{\text{ext}}$ lies inside a parabola, then the ground state of the quantum dot has a fixed number of charges $eN$, and the dot is insulating. When $Q_{\text{ext}}$ is at a value where $U(N) = U(N+1)$, then two parabolas cross, and the number of charges on the dot is degenerate between $N$ and $N+1$ charges. The charge on the dot can fluctuate and current is allowed to flow across the dot. Each parabola corresponds to a ground state of the island with a different value of $N$, the discrete charge, which increases in a stepwise fashion with $Q_{\text{ext}}$ (Figure B.3b). (The total energy $U(Q_{\text{ext}})$, of course, is not composed of degenerate parabolas, but rather of a set of lines with a parabolic envelope which define non-degenerate ground states for every value of $V_j$).

The energy of the dot $U_D$ helps clarify the evolution of its potential, $V_D$ (shown in Figure B.3c). At $Q_{\text{ext}} = V_j = 0$, we have $U_D = V_D = 0$. As $Q_{\text{ext}}$ increases, $V_D$ increases along with it. When the degeneracy condition is fulfilled, a discrete charge tunnels onto the dot, changing $N$ and decreasing the potential. This happens when $Q_{\text{ext}} = 0.5$, but $N$ is allowed to change only by 1. Thus, the tunneling charge screens more than the applied field, and the potential becomes negative. This is called overscreening, and happens because of the fundamentally discrete nature of the charge. The potential returns to zero only once $Q_{\text{ext}} = 1$, compensating precisely the potential induced by the
gate voltage and that of the discrete charge

In experiments, we measure the chemical potential, a relative energy scale, and not the total energy. With discrete charges, it is simple to express this using Equation B.4 as

\[
\mu = U(N) - U(N-1) = \frac{e^2}{C\Sigma} \left( N - \frac{1}{2} \right) + \frac{e}{C\Sigma} \left( \sum_{j=1}^{N_f} C_{D_j} V_j \right)
\]  

(B.6)

A schematic of \( \mu \) is shown in Figure B.3d. It evolves linearly until a discrete charge is added, and it jumps up to equilibrate with the chemical potential of the particle reservoir (usually, metallic Fermi leads, for which we’ve assumed a chemical potential of zero).

So far we have ignored the level spacing, as for an SET. To include it for a quantum dot, we express a new energy \( E(N) \) which includes the energy levels \( \epsilon_i \)

\[
E(N) = \sum_{i=1}^{N} \epsilon_i + U(N)
\]  

(B.7)

For \( N \) electrons, we pay both for charging and for the first \( N \) energy levels. The chemical potential is now

\[
\mu_N = E(N) - E(N-1) = \epsilon_N + \frac{e^2}{C\Sigma} \left( N - \frac{1}{2} \right) - e \left( \sum_{j=1}^{N_f} \alpha_j V_j \right)
\]  

(B.8)

where we define the lever arm \( \alpha_i = -C_{D_i}/C\Sigma > 0 \). We see that, unlike the total energy, the addition energy depends only on the \( N^{th} \) energy level \( \epsilon_N \).

One important complication is the work functions of the conductors. Even if the electrostatic potential is zero on a given conductor, it may still have a different electrochemical potential than the dot due to a difference between their work functions. We define the \( i^{th} \) work function difference as
\[ \phi_i, \text{ inserting it as} \]

\[
\mu_N = E(N) - E(N-1) = \epsilon_N + \frac{\epsilon^2}{C_\Sigma} (N - \frac{1}{2}) - e \left[ \sum_{j=1}^{N_f} \alpha_j (V_j + \phi_j) \right] \quad (B.9)
\]

This extra term is just a constant offset, so we’ll ignore it for this section where we want just the physical principles of measurement. We will return to it later.

When the \( N \) and \( N - 1 \) states are degenerate and \( \mu = 0 \) we have a peak in conductance through the dot. If we apply a voltage \( V_g \) to \( m \) gates, then \( \mu = 0 \) happens when

\[
eV_g = \frac{1}{\alpha_m} \left[ \epsilon_N + \frac{\epsilon^2}{C_\Sigma} (N - \frac{1}{2}) - e \sum_{j} \alpha_j V_j \right] \quad (B.10)
\]

where \( \alpha_m = \sum_i \alpha_i = -\frac{1}{C_\Sigma} \sum_i C_{Di} \) the combined lever arm of the \( m \) active gates. The spacing between different points of conductance is then

\[
e\Delta V_g = \frac{1}{\alpha_m} (\mu_N - \mu_{N-1}) = \frac{1}{\alpha_m} ((\epsilon_N - \epsilon_{N-1}) + \frac{\epsilon^2}{C_\Sigma})
= \Delta \epsilon_N \frac{1}{\alpha_m} + \epsilon^2 C_m \quad (B.11)
\]

where \( \Delta \epsilon_N \) is the \( N^{th} \) level spacing and \( C_m = -\sum_i C_{Di} \) the total capacitance of the active gates to the dot. This simple expression is one of the reasons quantum dots are so useful for physics: transport experiments through them allow us to extract directly the energy spectrum of the dot. We need only to subtract the charging energy with the right lever arm factor. Notice that if the level spacing goes to zero, then the spacing depends only on the capacitance of the active gates, and thus the peak spacing can also be used to extract the capacitance of different gates to the dot.

In a real device, two of the conductors, the source and drain, are in contact with the quantum dot.
(the rest we call gates). This means they are allowed to give and take particles from the quantum dot. A current can then flow through the quantum dot when \( \mu_D < \mu < \mu_S \) (for \( \mu_S > \mu_D \)), that is, when the dot chemical potential lies between that of the source and drain. In the lab, the drain is always connected to ground and \( \mu_D = 0 \). A bias voltage \( V_b \) is applied therefore only to the source and \( \mu_S = -eV_b \) (that is, the bias is not symmetrically applied, as is often assumed theoretically; the sign is Franklin’s fault). For \( V_b > 0 \), a given charge state \( N \) is therefore stable when its associated chemical potential \( \mu_N \) lies below \( \mu_S \), or \( \mu_N < -eV_b \), and the next charge state \( N + 1 \) has \( \mu_{N+1} > 0 \). When these are equalities, they give us a line relating \( V_g \) and \( V_b \) where the dot is at the edge of the blockaded region. For a voltage \( V_g \) applied to \( m \) gates, these two lines are given by

\[
V_b = \left( \frac{\alpha_m}{1 - \alpha_S} \right) V_g + \frac{1}{e\alpha_S} \left[ \epsilon_N + \frac{e^2}{C_{\Sigma}} (N - \frac{1}{2}) - e \sum_j \alpha_j V_j \right]
\]

\[
V_b = -\left( \frac{\alpha_m}{\alpha_S} \right) V_g + \frac{1}{e\alpha_S} \left[ \epsilon_{N+1} + \frac{e^2}{C_{\Sigma}} (N + \frac{1}{2}) - e \sum_j \alpha_j V_j \right]
\]

(B.13)

where \( \alpha_S \) is the lever arm of the source contact. In terms of the capacitances,

\[
V_b \propto \left( \frac{C_m}{C_{\Sigma} - C_S} \right) V_g
\]

\[
V_b \propto -\left( \frac{C_m}{C_S} \right) V_g
\]

(B.14)

with \( C_m \), the total capacitance of the active gates to the dot, defined above. Together with the two conditions for \( V_b < 0 \) (not shown), these four lines define Coulomb diamonds where transport is blockaded and charge is fixed. Note that because the bias is applied only to the source, the diamonds are not symmetric (even if the system is otherwise perfectly symmetric). The two lines intersect at
the tip of the diamond when \( V_g = V_g \) and

\[
e V_g^{\text{max}} = \Delta \epsilon_{N+1} + \frac{e^2}{C_S}
\]

Together with Equation B.11, this gives a direct measure of the lever arm \( \alpha_m \) of the active gates. The information contained in the Coulomb diamond picture studied here is summarized in Figure B.3.

**Figure B.3:** Quantum dot electrostatics from conductance measurements. a) Coulomb diamonds of zero conductance \( (G=0) \). See main text for definitions. b) Gate-gate measurements of conductance at zero bias. The slope reflects the ratio of capacitance of the two active gates. For multiple active gates, the capacitance is the sum of individual capacitances.

Another type of experiment involves measuring the relative coupling of two gates to the quantum dot at zero bias. We apply a voltage \( V_g^m \) to \( m \) gates and \( V_g^n \) to \( n \) gates, and the \( N^{th} \) conductance peak at \( \mu = 0 \) happens when

\[
V_g^m = \frac{\alpha_n}{\alpha_m} V_g^n + \frac{1}{e \alpha_m} \left[ \epsilon_N + \frac{e^2}{C_S} \left( N - \frac{1}{2} \right) - e \sum_j \alpha_j V_j \right]
\]

where \( \alpha_m \) and \( \alpha_n \), the total lever arms of each set of active gates, are defined analogously as above. In
terms of the capacitances, this is a line with

\[ V_m' \propto -\frac{C_m}{C_m} V_g' \]  

(B.17)

where \( C_m \) and \( C_n \), the total capacitance of each gate set to the dot, are defined analogously as above. As shown in Figure B.3, lines of constant charge have a slope relating the relative capacitances of the two gate voltages (or gate voltage groups) acting on the dot.

What if the capacitances change with gate voltages? If the quantum dot is defined by the gate voltages themselves, which is the case for our devices (see section 4.1), then the shape of the potential well defining the quantum dot can change. This means that the charge density changes shape with gate voltage, and so does the capacitance. The slopes therefore change, with their local derivative defining the capacitance ratio for the charge configuration defined for that set of gate voltages. This is the situation exemplified in Figure 4.1c and d.

What if one of the conductors is itself a quantum dot? This is the situation described in section 5.1. We want to know how the potential on the first dot changes with respect to a gate voltage that also charges a second dot, with total capacitance \( C_{\Sigma 2} \). From Equation B.2, we know the potential induced on the dot from a gate voltage \( V_g \) is

\[ V_D = \frac{C_g}{C_{\Sigma}} V_g \]  

(B.18)

while the voltage induced by a nearby quantum dot with potential \( V_{dot2} \) is

\[ V_D^{dot2} = \frac{C_{dot2}}{C_{\Sigma}} V_{dot2} \]  

(B.19)

(ignoring signs). We define the gate voltage \( V_g \) to have a capacitance \( C_{g}^{dot2} \) to the second quantum dot. When \( V_g = e / C_{g}^{dot2} \), the gate charges the second dot with an electron, and its potential changes
to be $V_{\text{dot2}} = e/C_{\Sigma 2}$. Substituting,

$$V_D^g = \frac{C_g}{C_{\Sigma}} \frac{e}{C_{\text{dot2}}}$$  \hspace{1cm} (B.20)

and

$$V_{\text{dot2}}^D = \frac{C_{\text{dot2}}}{C_{\Sigma}} \frac{e}{C_{\Sigma 2}}$$  \hspace{1cm} (B.21)

which are equal in magnitude only if

$$\frac{C_g}{C_{\text{dot2}}} = \frac{C_{\text{dot2}}}{C_{\Sigma 2}}$$  \hspace{1cm} (B.22)

If the second dot is well-isolated from everything but the gate, then $C_{\Sigma 2} \approx C_{\text{dot2}}$, and this condition amounts to $C_g = C_{\text{dot2}}$. That is, for a well-isolated second dot, the potential steps are equal if the gate couples to the second dot as well as the two dots couple to each other. In the terms of section 5.1, this is $C_{g-SET} = C_{NT-SET}$. However, in general $C_{\Sigma 2} > C_{\text{dot2}}$, and thus the steps are usually smaller, leading to the jagged line evolution illustrated in Figure 5.2, and measured in experiments, as in Figure 5.3.

### B.3 Distinguishing Double Dots

As we saw in section 2.1, disorder can break up the 1D electron liquid into a series of puddles, forming a series of quantum dots at low temperature. How do we distinguish a single, clean quantum dot from a series of dots?

To see how this is possible, consider the simplest quantum dot chain: a pair of dots in series, called a double quantum dot. The electrostatic analysis proceeds similarly to that of section B.2.

We spare the details here, which can be found in \textsuperscript{35}. We include the cross-capacitances $C_{g12}$ and $C_{g21}$, and the capacitance between the two dots $C_m$, as illustrated in Figure B.4a. The total capacitance of dots 1 and 2 are $C_1 = C_L + C_{g1} + C_{g12} + C_m$ and $C_2 = C_R + C_{g2} + C_{g21} + C_m$. The charging energies

167
Figure B.4: Double quantum dot electrostatics. a) Circuit diagram of the double quantum dot. b) Charge stability diagram in the limit of zero inter-dot coupling (and zero cross-capacitance). The two dots act as two independently-gated quantum dots. b) Limit of large inter-dot coupling. The two dots act electrostatically identically to a single dot. d) Intermediate regime. The charge stability diagram is a honeycomb lattice of charge states. e) Zoom-in on one of the honeycomb vertices highlighted in d). Transport occurs at the vertices where three charge states are degenerate. The pair of points are analogous to electron-hole conjugates. f) Dimensions of the honeycomb. See main text. (b)-(f) adapted from 155.

of the dots are:

$$U_{i(2)} = \frac{e^2}{C_{i(2)}} \left( \frac{1}{1 - \frac{C_m}{C_{i(2)}}} \right) = \frac{e^2}{C_{i(2)}} - \frac{C_m}{C_{i(2)}} \frac{C_{i(2)}}{C_{i(1)}}$$  \hspace{1cm} (B.23)

where $U_i$ is the charging energy of the $i^{th}$ dot.

We first proceed to describe the ground state charge configurations of the double dot. We denote the number of electrons in each dot by $N_1$ and $N_2$, and the gate voltages applied to each dot by $V_1$ and $V_2$. For each individual dot, the electrostatic energy defines the Coulomb valleys where $N_i$ is fixed in each dot. If we assume $C_m = 0$, then the two dots are entirely decoupled and $N_i$ depends
only on $V_i$, as shown in Figure B.4b. In this limit, we have two independent quantum dots. In the opposite limit, for $C_m \approx C_1, C_2$, the gate voltages couple equivalently to the two dots, charging either one equivalently, leading to a charge stability diagram equivalent to that of a single, two-gated quantum dot (Figure B.4c).

The intermediate regime, where $0 < C_m < C_1, C_2$, is shown in Figure B.4d. Now, each gate has an effective cross-capacitance to the neighboring dot, so that $N_{i(2)}$ depends also on $V_{2(1)}$, and the Coulomb peaks of dot 1(2) shift as a function of $V_{2(1)}$. In addition, electrons on the two dots experience a mutual repulsion. Thus for every electron added to dot 1(2), the Coulomb peaks of dot 2(1) shift by a discrete value. Denoting the charge state of the double dot $(N_1, N_2)$, the electrostatic energy of the state $(N_1 + 1, N_2 + 1)$ is not equal to the sum of the energies of the states $(N_1 + 1, N_2)$ and $(N_1, N_2 + 1)$, and contains an additional energy scale related to the interdot coupling $C_m$.

These considerations describe a honeycomb lattice of hexagons, where each hexagon corresponds to an electronic ground state $(N_1, N_2)$. Taking into account the cross-capacitances $C_{g12}$ and $C_{g21}$ and a level spacing $\Delta \epsilon$, the dimensions of the two-dimensional Coulomb valleys are given by

$$\Delta V_{g(2)} = \frac{eC_{2(1)}}{C_{g(2)} C_{2(1)} + C_mC_{g21(12)}} \left(1 + \frac{\Delta \epsilon}{U_{g(2)}}\right)$$  \hspace{1cm} (B.24)

and the shift of a Coulomb peak due to occupation of the neighboring dot is given by

$$\Delta V_{m(2)} = \frac{eC_m}{C_{g(2)} C_{2(1)} + C_mC_{g21(12)}} \left(1 + \frac{\Delta \epsilon}{U_m}\right)$$  \hspace{1cm} (B.25)

The quantity $U_m$ is given by

$$U_m = e^2 \frac{C_m}{C_1 C_2} \left(\frac{1}{1 - \frac{C_2}{C_1 C_2}}\right) = \frac{e^2}{\frac{C_1 C_2}{C_m} - C_m}$$  \hspace{1cm} (B.26)

and reflects the additional energy scale arising from the interdot coupling. The line corresponding to
the addition of an electron to dot 1 is given by

$$V_{g1} \propto -\left( \frac{C_2 C_{gl} + C_{g1} C_m}{C_2 C_{g12} + C_{g2} C_m} \right) V_{gl}$$

(B.27)

while the line corresponding to the addition of an electron to dot 2 is

$$V_{g2} \propto -\left( \frac{C_1 C_{g21} + C_{gl} C_m}{C_1 C_{g22} + C_{g12} C_m} \right) V_{gl}$$

(B.28)

Ignoring the cross-capacitances $C_{g12(21)}$, this is

$$V_{g2} \propto -\left( \frac{C_{g2(21)}}{C_{m(1)}} \right) \left( \frac{C_{gl}}{C_{g2}} \right) V_{gl}$$

(B.29)

with the subscript parentheses corresponding to the charging line of dot 2. The quantity $C_m/C_{1(2)}$ is the lever arm (denoted $\alpha$ in section B.2) of one dot to its neighbor. We find that the charging lines of the two individual dots have distinct slopes.

The two different slopes mean the complete charging line corresponding to $N_1 + N_2 \rightarrow N_1 + N_2 + 1$ has a kink, as visible in Figure B.4d. This feature is distinct from a single dot, and distinguishes the behavior of a double dot and a single, two-gated dot. In contrast, a single dot with two gates can never experience such a kink, since electrons are added uniformly to the entire dot (see Figure B.3). In the case of gate-defined dots where the capacitances are a function of the voltages, such as those shown in Figure 4.1c and d, the capacitances change smoothly with gate voltage, and thus no kinks are observed. In some cases, kinks may appear for other reasons, such as due to the presence of disorder potentials leading to multi-dot behavior or electron correlation effects (see section 4.2).

When is current allowed to flow through the double dot? For a single dot, current is allowed when the charge states $N$ and $N+1$ are degenerate, and the chemical potential of the dot is identical to
that of the leads. For a double dot, the chemical potential of each dot must be identical to that of the leads and in addition they must be identical to each other. Thus transport requires the degeneracy of three states. Denoting the charge state of the double dot \((N_1, N_2)\), we must have \((N_1, N_2), (N_1 + 1, N_2), \) and \((N_1, N_2 + 1)\) degenerate for current to flow. If \(C_m = 0\), then the state \((N_1 + 1, N_2 + 1)\) is also degenerate with this triplet. But for \(C_m > 0\), this state costs extra energy because of the electrostatic repulsion between the two dots. Thus it has an independent degeneracy condition, for the states \((N_1 + 1, N_2 + 1), (N_1 + 1, N_2), \) and \((N_1, N_2 + 1)\). Thus for each state \((N_1, N_2)\), there are two points in gate voltage space that allow current flow, called 'triple points', illustrated in Figure B.4e.

![Figure B.5: Double dot stability diagram with tunneling. Charge state boundaries become smooth and hyperbolic. The triple points are shifted by \(t_{12}\), the interdot tunnel coupling, to the maxima of the hyperbolae. Adapted from 156.](image)

In the presence of tunneling, the triple points become broadened by the coupling of the electronic states between the two dots and to the leads155,156. Hybridization of charge states rounds the two kinked slopes, as illustrated in Figure B.5. The electrostatic triple points are shifted to the peaks of the hyperbolic tunneling lines by an amount proportional to \(t_{12}\), the interdot tunnel coupling. For very strong tunneling much larger than \(e^2/C_{g1}\) and \(e^2/C_{g2}\), the two slopes will be washed out by the tunneling, resulting in one smooth charging line as in Figure B.4c. In this case, the ability to resolve the change in slope resulting from the double-dot nature depends on the width of the tunneling resonance, given by the larger of \(k_B T\) or \(h\Gamma\). Measurements as a function of two gate voltages
which show smooth, continuous charging lines, as in Figure 4.1c and d, therefore place an upper bound on the interdot tunneling given by the charging line widths.
In section 3.3, we discussed the fabrication of circuit chips for nano-assembly. The technique relied on lithographic patterning of contacts, gates, and capacitance pads, and then dry plasma etching to define the pillar for insertion into trenches. An alternative possibility to fabricate circuit chips that we developed is to first etch the silicon wafer, and afterwards perform lithography on the etched surface. This technique is trickier, but the tricks may be useful and thus we discuss that process here.

This alternative process requires a change in the design. It is nigh impossible to do lithography...
on a silicon pillar ~ $100\mu m$ tall and ~ $10\mu m$ wide. Instead, the pillar is made in two steps. In the first step, a deep, $100\mu m$ etch is performed with a design consisting of a large, rectangular region to hold the bonding pads and optical-lithographic lines, and a small 'stub' protruding from this rectangle, several hundred microns in length and width. Lithography is performed on this stub to create contacts and gates. A second dry etch is then performed to define the pillar on top of the stub.

The necessity of lithography on an etched wafer imposes a second design change. Typically wafers are designed with all chips pointing in one common direction. This isn’t a problem since resist spinning leads to uniform coverage. However, with a wafer etched into chips, spinning causes resist to bead on the edge of the etched chips facing out from the center of the wafer. The smallest features of the lithographic mask should therefore be on the inner side of the etched chips, closest to the wafer center. For our style of devices, this means all the 'stubs' are on the side of the rectangle closest to the wafer center.

The process flow is shown in Figure C.1, with panels (a)-(d) and (e)-(h) corresponding to two different etch solutions. The initial deep etching is done in wet solution. This creates a sidewall angle that is necessary for the resist spinning that comes later. We tried two different wet etchants, KOH (Figure C.1a-d) and TMAH (Figure C.1e-h).

KOH is a common silicon wet etchant that we successfully used, but which has key drawbacks. Etching with KOH is violent, leading to bubbles that can mask etching and lead to defects in geometry. Isopropyl alcohol (IPA) can be added to improve reduce bubble formation, but the addition of IPA doubles the lateral etch rate of KOH. This means small features like the stub that contains the device design is rapidly etched away, and compensating for this via design is difficult. It is thus best to avoid IPA and make up for the violent etching with good stirring of the solution. Another inconvenience is the doping-dependence of the KOH etch rate. For wafers of highly-doped silicon with resistivities of $0.005 - 0.020\Omega cm$, often used for backgates, this can mean an etch rate up to four times slower than a wafer with $1 - 20\Omega cm$. When bubble-masking and etch inhomogeneity is
an issue, many hours of etching can give poor results. A device after KOH deep etch without IPA is shown in Figure C.1a. The mask was defined with photolithography and BOE etching of the silicon, afterwards cleaning the wafer in solvents before the etch. The oxide here must be several microns thick to survive the wet etching.

In contrast, TMAH gives a much cleaner etch, without bubble masking (see also subsection 3.2.2). It gives a shallower sidewall angle, making resist spinning easier. In addition, it does not have an etch rate that depends on the silicon doping level. The difference between KOH and TMAH is apparent in Figure C.1a-d and e-h.

The following steps after deep wet etch are the same for the two etchants.

After the wet etching, a thick oxide is left on the silicon, which hangs over the underlying silicon. The second step is to thin this oxide in buffered oxide etch (BOE), shown in Figure C.1b and f. This eliminates the overhang, which is important for the resist spinning for lithography. It also allows one some freedom to choose the final thickness of oxide, which determines the capacitive coupling to the doped silicon backgate.

Exposing the surface to BOE makes resist-spinning difficult, with poor wetting properties. Therefore, after the BOE thinning, we dip the wafer in KOH at a temperature low enough to prevent etching. The result is shown in Figure C.1c and g, where noticeable stains have been removed from the surface.

To pattern the chips, we use a combination of photolithography and electron beam lithography. The resist spinning must be done carefully. Start the spinning from 0 RPM and slowly ramp up to 500 RPM, giving time for the resist to settle, then if necessary speed up to 1000 RPM or the appropriate speed. This should be carefully investigated to make sure the resist is properly covering the etched chips while thinning properly for exposure. The result, after evaporation and liftoff, is shown in Figure C.1d and h.

The final step is the dry plasma etch. The mask is made with photolithography to protect the
metal patterning. BOE is used to remove the oxide. After plasma etch and cleaning, the final result is shown in Figure C.1g (for KOH deep etching). Here, the plasma etch is $\sim 10\mu m$ deep, and can be made deeper to accommodate the nano-assembly process.

The critical point for the plasma etch is stub cleanliness. Because of the deep etch, spun resist often opens at the corner of the etched mesa. This means that after metal evaporation, there will be metal stuck to the silicon oxide at the edges of the stub. This metal is particularly difficult to remove in liftoff. It is necessary to do so to avoid the creation of a surrounding wall in the plasma etching process. This is shown in Figure C.2, where two views of a plasma-etched device show how a thin wall is created around the etched pillar due to the dirty stub edge. To keep the stub edge clean, proper cleaning should be performed after every step of the process, so that resist, oxide, and metal are all properly removed.
Figure C.1: SEM images of process flow for etch-first circuit chips (not including final plasma etch; see Figure C.2).

a-d: KOH etching. a) Deep etch in KOH solution. b) Oxide thinning in BOE. c) KOH surface treatment. d) Optical and e-beam patterning.

e-h: TMAH etching. e) Deep etch in TMAH solution. f) Oxide thinning in BOE. g) KOH surface treatment. h) Optical and e-beam patterning. All scale bars 100μm.
Figure C.2: Plasma etching of etch-first circuit chips and artifacts. a) Device after plasma etching. b) A second device after plasma etching. c) A different view of the device in panel (a). The wall surrounding the pillar at the stub edge originates in insufficient cleaning of the wafer, leading to resist residue and metal leftovers at the stub edge. All scale bars 20 \( \mu m \).
References


462(7269):74–7, November 2009. ISSN 1476-4687. URL http://dx.doi.org/10.1038/nature08482.


Anindya Das, Yuval Ronen, Yonatan Most, Yuval Oreg, Motty Heiblum, and Hadas Shtrikman. Zero-bias peaks and splitting in an AlInAs nanowire topological superconductor as a signature of Majorana fermions. *Nature Physics*, advance on, November 2012. ISSN 1745-2473. doi: 10.1038/nphys2479. URL http://dx.doi.org/10.1038/nphys2479.


This thesis was typeset using \LaTeX, originally developed by Leslie Lamport and based on Donald Knuth’s \TeX. The body text is set in 11 point Egenolff-Berner Garamond, a revival of Claude Garamont’s humanist typeface. A template that can be used to format a PhD thesis with this look and feel has been released under the permissive MIT (X11) license, and can be found online at github.com/suchow/Dissertate or from its author, Jordan Suchow, at suchow@post.harvard.edu.