STUDY OF BONDING AND DOPING PROPERTIES OF SP² CARBON NANOSTRUCTURES: NUMERICAL SIMULATIONS AND DEVELOPMENT OF EMPIRICAL INTERACTION POTENTIALS.

A Thesis in
Physics
by
Aleksey N. Kolmogorov

© 2004 Aleksey N. Kolmogorov

Submitted in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

August 2004
The Thesis of Aleksey N. Kolmogorov was reviewed and approved* by the following:

Date of signature

__________________________________________________________
Thesis Advisor, Chair of Committee
Vincent H. Crespi
Professor of Physics
Professor of Materials Science and Engineering
Downsbrough Professor

__________________________________________________________
Milton W. Cole
Distinguished Professor of Physics
Distinguished Professor of Materials Science and Engineering

__________________________________________________________
Ari M. Mizel
Assistant Professor of Physics

__________________________________________________________
Rajeev Sharma
Associate Professor of Computer Science and Engineering

__________________________________________________________
Richard Robinett
Professor of Physics
Director of Graduate Studies
Associate Department Head

*Signatures are on file in the Graduate School.
Abstract

The main topics of this dissertation are the development of empirical interaction potentials and the study of electronic and mechanical properties of $sp^2$-bonded carbon nanostructures.

First, we investigate the weak interlayer binding in graphitic structures. We argue that although the absolute cohesion is not properly described in the local DFT approximations, the variation in the binding energy under interlayer shifts appears to be much more sound than previously suspected. We combine this result with experimental data to introduce a new empirical potential, which describes the variation in the interlayer binding with the relative alignment (registry) of the layers. Lacking a registry dependence, the commonly used Lennard-Jones potential significantly underestimates the variation in energy. We use our potential to study interwall sliding in nested nanotubes. We find that the well-defined geometry and extreme structural anisotropy of a multiwalled carbon nanotube can bring qualitatively new features to its nanometer-scale tribology. Efficient cancellation of registration-dependent interactions in incommensurate tubes (and also, surprisingly, certain axial commensurate tubes) can induce extremely small and nonextensive shear strengths. This result suggests the use of multiwalled carbon nanotubes as nanoscale bearings. We also apply our potential to look at the alignment of nanotubes on a graphitic substrate. We discover that the interaction of a one-dimensional tube with a two-dimensional substrate then leads to an unusual registry phenomenon not visible in standard layer-on-layer growth: the system develops favorable orientations which clearly are incommensurate. This effect could be used for nanotube separation by their helical angle.

Next, we study the effect of doping in carbon nanostructures. Using a self-consistent tight-binding model we examine the radial distribution of charge in a bromine-doped double-walled carbon nanotube system. Our results confirm recent Raman measurements that most of the charge resides on the outer wall, even when the outer nanotube is semiconducting and the inner nanotube is metallic. According to experimental data nanoporous carbon also exhibits interesting behavior under doping with alkali atoms: it undergoes graphitization at relatively low temperatures. We propose a representative model of a fully $sp^2$-connected defect (a wormhole) in a carbon network, which could be used to study the graphitization phenomenon. We investigate structural properties of
the wormhole and consider possible mechanisms of its annihilation.

Finally, we explore the possibility to describe interatomic interaction in solids with neural networks methods. We find that with a proper choice of the network and input variables the forces on atoms and the total energy of the system around an equilibrium can be described with \textit{ab initio} accuracy. We formulate an algorithm how to build and train the network for modeling solid states systems.
# Table of Contents

List of Figures vii  
List of Symbols xiii  
Acknowledgments xiv  

1 Introduction 1  
1.1 Carbon structures ........................................... 1  
1.1.1 Binding in carbon structures ............................. 1  
1.1.2 $sp^2$ forms of carbon ..................................... 4  
1.2 Density Functional Theory ................................. 7  
1.2.1 Density Functional formalism: Hohenberg-Kohn theorems .... 7  
1.2.2 Kohn-Sham equations ..................................... 9  
1.2.3 Local Density and Generalized Gradient Approximations .... 12  
1.2.4 Nonlocal correlations .................................... 14  
1.3 Tight Binding method ....................................... 15  
1.3.1 Periodic boundary conditions in crystals ................. 16  
1.3.2 Secular equation ........................................ 17  

2 Interlayer interaction in graphitic structures I: Theoretical models 19  
2.1 Review .......................................................... 19  
2.2 Experimental data ............................................. 20  
2.3 $ab$ initio results .............................................. 21  
2.4 Registry Dependent Potential ................................. 24  
2.4.1 Original registry dependent potential ..................... 24  
2.4.2 Improved registry dependent model ....................... 26  
2.4.3 Tests ..................................................... 28  

3 Interlayer interaction in graphitic structures II: Applications 33  
3.1 Nanotube-nanotube interactions ............................... 33  
3.1.1 Manipulation of multiwalled carbon nanotubes .............. 33  
3.1.2 Simulation of nanotube-nanotube sliding .................. 34  
3.1.3 Intertube motion: recent developments and potential applications 40
List of Figures

1.1 Radial distribution of the 2s and 2p orbitals in carbon ............... 2
1.2 Most stable two- and three-dimensional forms of carbon: graphene and diamond. .................................................. 3
1.3 Band structure of graphene calculated in the LDA (see Section 1.2.3). The results agree with those obtained in [2]. ....................... 4
1.4 \(sp^2\) forms of carbon. ........................................... 5
1.5 Construction of an (8,2) nanotube out of a patch of graphene. The circumference is formed by a linear combination of primitive lattice vectors of graphene; in this case it is \(8a_1 + 2a_2\). \(\Theta_0\) is the helical angle of the tube, and \(L_x\) is the resulting period of the tube along its axis. ............ 6
1.6 Schematic representation of binding mechanism in graphite as in [17]. .... 15

2.1 Different stackings of graphite. ..................................... 20
2.2 a Interlayer energy between two graphitic sheets. b The energy difference between AA and AB stackings of graphite as a function of the interlayer distance. .................................................. 22
2.3 Contributions to corrugation to sliding for two graphene sheets 3.34 Å apart calculated in the LDA. The GGA gives very close results. From top to bottom: Hartree, exchange-correlation, total, Ewald and band structure energies. ............................................................. 23
2.4 The interlayer binding energy versus parallel shift for two rigid graphitic layers 3.44 Å apart. The dots are the local density approximation data; the solid curve is the fitted delocalization term. The insert gives \(E_{AA} - E_{AB}\) (normalized to 1 at 3.44 Å) as a function of the interplanar distance \(z\) with a fit to an exponential. ................................. 26
2.5 Definitions of the normal vector on an \(sp^2\) carbon site and the transverse distances between two atoms. ......................... 27
2.6 Corrugation to sliding for the nanotube/substrate interaction per atom of nanotube. The nanotubes are positioned 3.2 Å above a graphitic substrate; the energy difference is evaluated for structures when the lowest row of atoms on the nanotube is in AA or AB registration with the graphene layer. Armchair nanotubes, from left to right: (5,5), (6,6), (8,8), (10,10), (12,12), (15,15), (20,20); zigzag nanotubes, from left to right: (10,0), (16,0), (20,0), (24,0), (28,0).

2.7 An example of accidental commensuration in graphite. The two layers become commensurate when rotated by acos(11/14) with respect to each other. The resulting skewed unit cell has 14 atoms in each layer.

3.1 Three classes of DWCN’s with short commensurate, long commensurate, and incommensurate periods.

3.2 The corrugation as a function of number of atoms in the outer tube in a double-walled system without structural relaxation. In the commensurate (7,7)/(12,12) tube the corrugation grows linearly with a slope of 20 meV/Å. The corrugation in the incommensurate (14,0)/(16,10) tube does not grow with the system size. The commensurate (11,2)/(12,12) system has extremely small corrugation per period (i.e. 336 atoms of the outer tube). The curve is a guide for the eye.

3.3 The position of the nanotube above the graphitic substrate is defined by two angles $\phi$ and $\Theta$ and three coordinates of the nanotube center of mass $x$, $y$, and $z$.

3.4 The nanotube-graphene interaction energy, measured in meV per atom in the tube, as a function of angle about an axis perpendicular to the substrate. The zero angle is shifted to match the helical angle $\theta_0$. Here we assume matching bond lengths in the nanotube and graphene substrate. The distributions of interlayer registries for the minima marked “[2a]” and “[2b]” (at 30.0 and 38.2 degrees respectively) are shown in Fig. 3.5.

3.5 Solid points and lines represent the distribution of nanotube atoms in the area of contact with the substrate, mapped back into a single substrate unit cell, which is depicted by open circles and dashed lines. (a) and (b) are for equal bond lengths in tube and substrate. In (c) and (d) the bonds on the tubes are extended by 0.35% along the axis and 0.45% perpendicular to it.

3.6 Same as Fig. 3.4, but the C-C bonds in the nanotube are extended by 0.35% along the axis and 0.45% perpendicular to it, as described in the text. The samplings for c and d are shown in Figs. 3.5c,d.
3.7 **a:** Average (circles) and maximum (squares) interaction energy as a function of nanotube radius for nanotubes of different wrapping indices (from left to right: (5, 5), (6, 4), (7, 3), (8, 2), (9, 1), (10, 0), (6, 6), (7, 5), (8, 4), (9, 3), (10, 2), (11, 1), (12, 0)). The hollow (solid) points are for matching (mismatched) bond lengths, as described in the text. **b:** Tube-substrate interaction energy for nanotubes of various helicities and nearly constant radii (4.1Å to 4.3Å); the nanotube bonds are mismatched. From left to right: (11, 0), (10, 1), (9, 3), (8, 4), (7, 5), (6, 6), (5, 7), (4, 8), (3, 9), (1, 10).
All tubes make a 19.7° angle with the substrate, which was chosen to favor the (8, 4) tube. The (4, 8) tube is not favored, since the combined system lacks inversion symmetry.

4.1 Directions of the continuous wave vectors in nanotubes slicing through the Brillouin zone of graphene. Nanotubes are metallic if the allowed lines go through the K point of graphene, as shown on the left, and semiconducting otherwise, as shown on the right.

4.2 One-dimensional energy dispersion relations for (a) armchair (5, 5) nanotube, (b) (9, 0) zigzag nanotube, and (c) (10, 0) zigzag nanotube.

4.3 Unpolarized Raman scattering spectra (T=300K) of DWNTs using 1064 nm laser excitation [88]: (a) low frequency radial modes and (b) high frequency tangential modes. The thin solid line is the fit to the spectrum, with Lorentzian components represented by dashed lines.

4.4 Unpolarized Raman scattering spectra (T=300K) of DWNTs and SWNTs using 514.5 nm laser excitation [88]. The upper (lower) pair of spectra are for the SWNT control (DWNT) sample. The thin solid line is the fit to the spectrum. Lorentzian components for the outer (inner) tubes are represented by dashed thin (dotted thick) lines.

4.5 Schematic representation of the double-walled nanotube as a cylindrical capacitor. The potential difference between the tubes depends only on the charge on the inner tube.

4.6 Shift of bands ∆V in the (5,5)/(17,0) system caused by the electrostatic potential.

4.7 Calculated number of holes per Å on the inner tube versus the total number of holes on the double-walled pair. Metallic tubes are in boldface. Alternation of grey and black curves is solely for clarity in distinguishing adjacent curves.

5.1 HREM images of (a) NPC prior to Cs treatment, but after heating in vacuum for 18 h at T=450°C. (b) NPC after treatment with Cs and post-treatment at T=50°C [94]. (c) Post-treatment at T=350°C. (d) Post-treatment at T=500°C.

5.2 A neck (132 atoms), connecting two graphitic layers, consists of 12 heptagons and 12 additional carbons inserted in the middle.
5.3 Topological opposites: the wormhole (genus two) and the fullerene (genus zero).

5.4 A full square unit cell (224 atoms) containing a wormhole. The lattice vectors are 13.1 Å in each direction. The distance between the layers is 6 Å. The flat sheets can be reconstructed from the wormhole if the two sets of carbons (six green and six blue) retract back into the respective layers.

5.5 A series of transformation leading to the destruction of the wormhole and restoration of the flat $sp^2$ network. The energetics of these transformations are shown in Fig. 5.6

5.6 The energies of the wormhole annihilation steps calculated in LDA and tight-binding. The corresponding configurations before relaxation are shown in Fig. 5.5 (step zero is the wormhole, step six is graphite). The reference point is the energy of the wormhole. For the first three configurations the relaxed states are close to the starting structure. For step three the two methods find different metastable configurations: the one produced by the LDA does not have dangling bonds because some atoms are in the $sp^3$ configuration (see Fig. 5.7). For some configurations after step three the relaxation leaves the system in the graphitic state.

5.7 Step three relaxed with tight-binding (a) and LDA (b) methods. The structure on the right does not have dangling bonds; instead some atoms are in the $sp^3$ configuration.

5.8 The Stone-Wales transformations in the wormhole, graphite, and the bell (see Fig. 5.10). The initial, saddle point, and final states for graphite are shown in Fig. 5.9

5.9 The Stone-Wales transformation in graphite. a) the initial state; b) saddle point; c) final state.

5.10 Two finite-size structures used, in particular, for studying the effects of the periodic boundary constraints. a) The bell (from [105]) consists of 34 carbons and 12 hydrogens; in one of the tests we rotate the bond, shown in red. b) A wormhole after the SW transformation and the first bond-breaking (the starting configuration is as in Fig. 5.3, only with hydrogen-terminated ends).

5.11 Reaction paths found with the NEB for the first two transformations in the wormhole: the SW (from 0 to 1) and the bond breaking (from 1 to 2). These calculations are done for the wormhole in the square periodic unit cell; the pictures of the steps 0,1, and 2 show only the central part of the wormhole for convenience.

6.1 A two-layer perceptron. The first and second layers have three and two neurons respectively. Each layers has an additional bias input.
6.2 An example of a three-layer perceptron with two inputs and one output, which creates a localized peak at (0, 0). The first and second layers have the hyperbolic tangent as an activation functions; the last layer performs a linear transformation. The resulting weights are shown below (the units in each layer are ordered from left to right except that biases are labeled as zero as in Fig. 6.1; first index denotes the unit in a lower layer). First layer: \(w_{01} = 1, w_{02} = 1, w_{03} = 1, w_{04} = 1, w_{11} = 1, w_{12} = -1, w_{13} = 0, w_{14} = 0, w_{21} = 0, w_{22} = 0, w_{23} = 1, w_{24} = -1\); second layer: \(w_{01} = 0, w_{03} = 0, w_{11} = 0, 0.5, w_{12} = 0, 0.5, w_{13} = 0, w_{21} = 0, w_{22} = 0, 0.5, w_{23} = 0, 0.5, w_{31} = 0, 0.5, w_{32} = 0, 0.5, w_{41} = 0, w_{42} = 0, w_{43} = 0.5\); third layer: \(w_{01} = 0, w_{11} = -10, w_{21} = 10, w_{31} = 10\). The intermediate steps and the resulting peak are shown in Fig. 6.3.

6.3 An example of a three-layer perceptron with two inputs and one output, which creates a localized peak at (0, 0). First, two ridges are created along \(x_1\) (a) and \(x_2\) (b). When combined, they form a peak in the middle (c), while the secondary ridges are still present. The last two layers of the MLP suppress the ridges leaving the localized peak only.

6.4 Approximation of a two-dimensional surface with 40 RBFs, randomly scattered in the shown region. a) A wave propagates in the \(x_1\) direction; b) the width of the Gaussians is smaller than the average distance between them; b) the width is much greater than the average distance; c) the width and the average distance are comparable.

6.5 a) Local environment of an \(sp^2\)-bonded carbon atom consisting of nine nearest neighbors. b) A branch includes only five atoms: 1,2,3,4, and 5; two similar branches have atoms 1,2,3,6,7 and 1,2,3,8,9 respectively. c) The position of atoms in a branch can be described by 12 parameters \((x_1, x_2, y_2, x_3, y_3, z_3, x_4, y_4, z_4, x_5, y_5, z_5)\), if the first bond is put along the \(x\) direction and the second bond lies in the \((x, y)\) plane.

6.6 Performance test of the RBF network trained on 200 data structures. The black and red curves correspond to the E-type and EF-type training (see text) respectively. The top panel shows the error in the total energy calculated as \((\langle \Delta E^2 \rangle / \langle E^2 \rangle)^{1/2}\). The error in the force on the bottom panel is \((\langle \Delta F^2 \rangle / \langle F^2 \rangle)^{1/2}\). The tests are performed on 200 configurations, not included in the training set.

6.7 Same as in Fig. 6.6, only the training set has 600 structures.

6.8 Variation in energy caused by the displacement of one atom in a 128-atom graphitic unit cell in the LDA, tight-binding, Tersoff-Brenner and our RBF treatments.

6.9 The same setup as in Fig. 6.8. This plot shows the forces on the displaced atom label ’1’ (top panel) and its third neighbor labeled ’2’ (bottom panel).
6.10 The same setup as in Fig. 6.8. This plot shows the numerical derivatives of the forces on the atom being displaced. The derivatives are normalized to be one at $x = 0$. For the harmonic dependence the forces would be linear and the plotted value would remain at one. 

6.11 Reaction path for the Stone-Wales transformation found with the NEB. The calculations are done in LDA with VASP at PREC=HIGH (black dots). The unit cell is kept constant for all steps. Different curves show the energy of the system calculated with the RBF (in red), Tersoff-Brenner (in blue), and the tight-binding methods.
List of Acronyms

MWCN  Multiwalled Carbon Nanotubes
DWCN  Double-walled Carbon Nanotubes
HOPG  Highly Ordered Pyrolytic Graphite
DFT   Density-Functional Theory
LDA   Local Density Approximation
GGA   Generalized Gradient Approximation
NPC   Nanoporous Carbon
NN    Neural Networks
RBF   Radial Basis Functions
VASP  Vienna ab initio Simulation Package
EDIP  Environment-Dependent Interatomic Potential
RDP   Registry-dependent Potential
LJ    Lennard-Jones
SVD   Singular Value Decomposition
NEB   Nudged Elastic Band
MEP   Minimum Energy Path
SW    Stone-Wales
VdW   Van der Waals
Acknowledgments

First of all, I would like to thank my advisor, Prof. Vincent Crespi for the guidance, support, and encouragement he has provided throughout the course of my study at Penn State. His incredible physical intuition, and general knowledge on everything made my research an interesting and enjoyable learning experience. He has also shown a great deal of patience dealing with my natural stubbornness. I consider it a very lucky turn in my life that Vin gave me an opportunity to work and learn Physics in his research group.

I am also deeply grateful to Paul Lammert, who has always been available and willing to discuss any problem in my research. Without his always friendly and insightful help my thesis would have not been completed. Thanks, Paul.

I greatly appreciate the useful feedback on my thesis from my committee members, Milton Cole and Ari Mizel. I also would like to acknowledge Rajeev Sharma and Jesse Barlow for introducing me to the field of Computer Science and helping me with numerical algorithms.

I am indebted to Peihong Zhang for sharing his experience with me, especially at the beginning of my study, and to Jorge Sofo for valuable and fruitful discussions. Their vast knowledge of the Density Functional Theory helped me navigate through the complicated field of Computational Condensed Matter Physics.

I would like to thank our collaborator, James Ellenbogen for bringing the interesting idea of possible nanotube separation to us, which became a part of my thesis.

Many thanks go to my colleagues, Dragan Stojkovic, Kito Holliday, and David Stucke for the very friendly working environment and interesting discussions about physics and beyond it.

Finally, I would like to thank my family and Roxana Margine for their care and support. Without them I could not have come this far.
Chapter 1

Introduction

1.1 Carbon structures

Carbon ($1s^22s^22p^2$) is best known as an element of life: it is an essential component of millions of organic compounds. No less remarkable is its capability to form very stable pure structures, which vary in symmetry, dimensionality, and topology. Their physical properties are truly diverse: carbon materials can be dielectric, semiconducting, or metallic; they can be extremely hard, or soft to the point of being considered a lubricant. The key to understanding the flexibility of binding properties is in the interesting electronic structure of the valence states.

1.1.1 Binding in carbon structures

Carbon has four valence electrons which, in an isolated carbon atom, occupy $2s$ and $2p$ orbitals. An approximate expression for the orbitals can be derived by solving the Schrödinger equation in a simplified hydrogen-like model [1]:

\[
\begin{align*}
\phi_{2s}(r) & = \left( \frac{Z}{2a_0} \right)^{3/2} \left( 2 \frac{Zr}{a_0} \right) e^{-\frac{2r}{2a_0}} \\
\phi_{2p_x}(r) & = \left( \frac{Z}{2a_0} \right)^{3/2} \frac{1}{\sqrt{3}} \frac{Zx}{a_0} e^{-\frac{2r}{2a_0}} \\
\phi_{2p_y}(r) & = \left( \frac{Z}{2a_0} \right)^{3/2} \frac{1}{\sqrt{3}} \frac{Zy}{a_0} e^{-\frac{2r}{2a_0}} \\
\phi_{2p_z}(r) & = \left( \frac{Z}{2a_0} \right)^{3/2} \frac{1}{\sqrt{3}} \frac{Zz}{a_0} e^{-\frac{2r}{2a_0}},
\end{align*}
\]

where $Z$ is the effective charge of the nucleus.
An all electron Local Density Approximation (see Chapter 1.2.3) calculation gives more realistic wavefunctions shown in Fig. 1.1. The $2s$ and $2p$ orbitals peak at about the same distance from a carbon nucleus (0.7 Å, see Fig. 1.1), which allows them to mix together and form different sets of orbitals. The energy of the new $sp^x$ states is somewhere between the $2s$ and $2p$ energy levels. In an isolated carbon atom the four valence electrons prefer to be on the $2s$ and $2p$ levels but if there are other atoms around the mixed states are more energetically favorable: the hybridized orbitals extend more towards neighboring atoms increasing the wavefunction overlap, and therefore, the cohesion.

Depending on the number of orbitals involved in the hybridization the bonds can be linear ($sp$), planar ($sp^2$) or three-dimensional ($sp^3$). Fig. 1.2 shows two of the most stable planar and bulk crystal structures: graphene and diamond. The highly symmetric geometry and large overlap of $sp^3$ orbitals make diamond the hardest material known. All four valence electrons participate in the strong covalent bonding, filling deep energy levels ($\sigma$ states). Because of the pronounced localization of the valence electrons diamond is a very good dielectric: the bandgap of 5.5 eV is one of the largest among covalent materials.

In the case of graphene only three orbitals hybridize; they can be easily expressed in
The new orbitals are oriented at 120° with respect to one another and the resulting crystal structure is a planar hexagonal or honeycomb lattice. The equilibrium distance between atoms in graphene is smaller than in diamond, so the relative binding per bond in graphite is larger. In addition, the unhybridized $p_z$ orbitals overlap sideways forming extra $\pi$-bonds (see Fig. 1.3). They are not as strong as $\sigma$ bonds but they define the most interesting properties of graphene. First, the $\pi$ bands cross exactly at the Fermi level, which can be shown in a simple tight-binding model (Section 4.1, Eq. 4.1), or with ab initio calculations as shown in Fig. 1.3. It leads to a unique band structure for this two-dimensional system: at K point there is no gap between the valence and conduction bands but there are very few available electronic states around the point of crossing. For
Figure 1.3. Band structure of graphene calculated in the LDA (see Section 1.2.3). The results agree with those obtained in [2].

this reason graphene is called a zero-gap semiconductor or a semi-metal. Second, the $\pi$ states are delocalized above and below the sheet helping stabilize the planar geometry of graphene and making the total cohesion for this structure comparable to that of diamond (see Fig. 1.2). (In contrast, the larger silicon atoms are too far apart from one another to form $\pi$-bonds and planar structure has not been observed in nature).

1.1.2 $sp^2$ forms of carbon

As a result of this bonding mechanism the graphene layers are extremely stable. Hexagonal layers of carbon can be stacked on top of one another and it is in this form, called graphite, that the $sp^2$ form of carbon is mostly found in nature. The bonding of graphitic sheets retains its strength even when the layers bend, which allows new classes of purely $sp^2$ connected carbon structures.

Fullerenes are a family of carbon clusters discovered two decades ago. The most famous and most studied fullerene, $C_{60}$, has the shape of a soccerball. The carbon network in this structure is highly curved, causing the atoms to be in a configuration intermediate between the $sp^2$ and $sp^3$. Nevertheless, the molecules are very stable; they are considered as a promising element for building nanoscale devices. One of the most interesting properties of fullerenes is an unusually high temperature superconductivity of solid $C_{60}$ doped with alkali atoms (up to 40 K [3]).

Another form of pure carbon is an $sp^2$ network of atoms connected in three dimen-
sions. It has holes, or pores, ranging from nano- to macrometers in diameter. This structure, called nanoporous carbon, has interesting topology, which is discussed in more detail in Chapter 5 (Fig. 1.4 shows an example of a pore of just 3 Å in diameter). The pore size can be controlled, which suggests practical applications of the material as membranes, gas storage cells, etc. There has also been some experimental evidence of nanoporous carbon graphitization in the presence of alkali atoms at temperatures much lower than the naively expected barriers would suggest. A possible mechanism for this phenomenon is presented later in Chapter 5.

Finally, one of the exciting topics in condensed matter physics and materials science has been quasi-one-dimensional carbon structures — carbon nanotubes — discovered in 1991 by Iijima [4]. He first observed multi-walled carbon nanotubes: a set of coaxial cylinders made out of hexagonal lattice of carbon atoms. Since then single-walled carbon nanotubes have also been successfully produced. Very small in diameter (as thin as a few Angstroms across) and very long (up to a few microns), these geometrically perfect objects have fascinating physical and mechanical properties. Fig. 1.5 shows how a nanotube can be constructed by wrapping a patch of graphene into a cylinder, where the circumference is formed by a linear combination of the primitive lattice vectors of

![Figure 1.4. $sp^2$ forms of carbon.](image)
Figure 1.5. Construction of an (8, 2) nanotube out of a patch of graphene. The circumference is formed by a linear combination of primitive lattice vectors of graphene; in this case it is $8 \mathbf{a}_1 + 2 \mathbf{a}_2$. $\Theta_0$ is the helical angle of the tube, and $L_x$ is the resulting period of the tube along its axis.

The electronic properties of a single-walled nanotube are very sensitive to its structural specification, that is the diameter and helical pitch, since electrons are confined in the radial direction and along the circumference. Due to the additional circumferential constraint electrons can freely propagate only along the axis of the nanotube [5], which dramatically changes the structure of electronic states for this quasi-one-dimensional system. While graphene is semi-metallic, nanotubes can be either metallic or semiconducting depending on the wrapping indices. A simple relation between $n$ and $m$ allows one to distinguish the two cases: nanotubes are metallic when $2n + m$ is divisible by three and semiconducting otherwise (more details are in Chapter 4). Other physical properties, such as optical activity, mechanical strength, and charge transfer effects also
depend critically on the nanotube diameter and helicity. We present our study of alkali doping of multiwalled nanotube doping in Chapter 4.

So far only the strong in-plane binding has been discussed. The interlayer interaction between graphitic interfaces is much weaker: at equilibrium spacing, it is from two to three orders of magnitude smaller than that within the layer. This unusually strong anisotropy in binding makes it possible for pure carbon nanostructures to maintain their structural integrity when they interact with neighboring interfaces. For example, graphite is very soft perpendicular to the basal plane and graphene layers can be easily peeled off, but the individual sheets remain practically intact. In MWCN’s individual tubes can slide with respect to one another but their structure remains nearly unchanged (see Chapter 3).

The subtle interlayer interaction is a complicated many-body effect and deserves a close look. The binding is considered to come from a combination of long-ranged Van der Waals dispersion forces and interlayer $\pi - \pi$ orbital overlap. There have been numerous attempts to model the interaction with ab initio and empirical methods, but so far no approach has fully agreed with available experimental data. In Chapter 2 we discuss the difficulties associated with a full description of the interlayer interaction and propose an empirical registry dependent graphitic potential, which bridges the gap between the current classical and quantum-mechanical approaches.

Before presenting results of our numerical simulations I would like to review some main empirical and first-principle methods, which are presently used for calculation of solid state properties in Condensed Matter Physics.

1.2 Density Functional Theory

1.2.1 Density Functional formalism: Hohenberg-Kohn theorems

The conventional approach to solving quantum mechanics problems uses the wave function $\Psi$ as the central quantity. The reason is that once we know $\Psi$, we have access to all information that can be known about the state of the target system. Although justified in principle, this approach is not very practical. Even if we separate the electronic and nuclear degrees of freedom (this so-called Born-Oppenheimer approximation works well for most systems since nuclei are much heavier than electrons) the wavefunction still depends on $4N$ variables: three spatial and one spin variable for each of the $N$ electron
in the system. Solving the Schrödinger equation
\[ \hat{H} \Psi = \left[ \hat{T} + \hat{V}_{Ne} + \hat{V}_{ee} \right] \Psi = E \Psi, \quad (1.4) \]
is a formidable task even for a few electrons. Operators \( \hat{T}, \hat{V}_{ee} \) in Eq. (1.4) describe the kinetic energy and the electron-electron interaction respectively; operator \( \hat{V}_{Ne} \) accounts for the interaction of electron with nuclei, in case of external fields it can be generalized as \( \hat{V}^{ext} \).

A key step to make the problem treatable is to realize that the Hamiltonian \( \hat{H} \) contains operators that act only on one (\( \hat{T} \) and \( \hat{V}_{Ne} \)) or at most two (\( \hat{V}_{ee} \)) particles at a time. A few important questions arise: is the full many-body wavefunction really needed to obtain the total energy and other properties of interest? Is some information in it irrelevant so that a simpler quantity as the central variable could still allow extraction of all desired information about the quantum state?

The answer to these questions is a tenet of the Density Functional Theory: any property of a system of many interacting particles can be viewed as a functional of the ground state density \( \rho_0(r) \):
\[ \rho_0(r) \equiv \langle \psi_0 | \sum_i \delta(r - r_i) | \psi_0 \rangle. \quad (1.5) \]
Hohenberg and Kohn [6] physically justified it in a disarmingly simple way. In their first theorem that proved that, quoting:

the external potential \( V^{ext}(r) \rho(r) \) is (to within a constant) a unique functional of \( \rho(r) \); since, in turn \( V^{ext}(r) \rho(r) \) fixes \( \hat{H} \) we see that the full many particle ground state is a unique functional of \( \rho(r) \).

Diagrammatically, this can be expressed as:
\[ \rho_0(r) \Rightarrow \hat{V}^{ext} \Rightarrow \hat{H} \Rightarrow \Psi_0 \Rightarrow E_0 \quad (and \ all \ other \ properties) \quad (1.6) \]

If no external fields are present the external potential is fully determined by the attraction due to the nuclei; we will use \( \hat{V}_{Ne} \) instead of \( \hat{V}^{ext} \) from now on.

It is convenient to separate the total energy expression
\[ E_0[\rho] = T[\rho_0] + E_{ee}[\rho_0] + E_{Ne}[\rho_0] \quad (1.7) \]
into parts that depend on the actual system (the last term) and those that are universal
in the sense that their forms are independent of the external parameters \(N, R_A, Z_A\) (the positions and charges of \(N\) nuclei):

\[
E_0[\rho(r)] = F_{HK}[\rho(r)] + \int V_{Ne}(r)\rho(r)\,dr.
\]  

The Hohenberg-Kohn functional \(F_{HK}[\rho(r)]\) contains the electronic kinetic energy and the electronic Coulomb interaction

\[
F_{HK}[\rho(r)] \equiv \langle \psi|\hat{T} + \hat{V}_{ee}|\psi\rangle,
\]

and is completely independent of the system at hand.

The second, fundamental Hohenberg-Kohn theorem of density functional theory states that the ground state energy can be obtained by minimization of the functional Eq. (1.8), constrained with the conservation of the total charge [6, 7]:

\[
N = \int \rho(r)\,dr,
\]

The minimization proceeds over the set of all densities \(\rho\) that can be obtained with antisymmetric wave functions [7].

### 1.2.2 Kohn-Sham equations

Unfortunately the universal functional \(F_{HK}[\rho(r)]\) is not known. The Hohenberg-Kohn theorems are just proofs of existence, but they do not provide any guidance at all on how the functional should be constructed.

A groundbreaking approach to solving this problem was proposed by Kohn and Sham [8] in 1965. They introduced the concept of a non-interacting reference system built from a set of orbitals (i.e., one electron functions) such that the major part of the kinetic energy can be computed to good accuracy. The remainder is merged with the non-classical contributions to the electron-electron repulsion - which are also unknown, but usually fairly small. By this method, as much information as possible is computed exactly, leaving only a small part of the total energy determined by an approximate functional.

First, let us see what advantages we get from considering a system of non-interacting
particles with an effective, local potential $V_S(r)$:

$$\hat{H}_S = -\frac{1}{2} \sum_i^N \nabla_i^2 + \sum_i^N V_S(r_i)$$  \hspace{1cm} (1.11)

It can easily be shown [1] that the exact wave functions of non-interaction fermions are Slater determinants

$$\Theta_S = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(r_1) & \phi_2(r_1) & \ldots & \phi_N(r_1) \\ \phi_1(r_2) & \phi_2(r_2) & \ldots & \phi_N(r_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(r_N) & \phi_2(r_N) & \ldots & \phi_N(r_N) \end{vmatrix}$$  \hspace{1cm} (1.12)

where the spin orbitals are determined by solving a system of one-electron equations

$$\hat{H}^{KS} \phi_i = \epsilon_i \phi_i$$  \hspace{1cm} (1.13)

with the one-electron Kohn-Sham operator $\hat{H}^{KS}$ defined as

$$\hat{H}^{KS} = -\frac{1}{2} \nabla^2 + V_S(r).$$  \hspace{1cm} (1.14)

All we need to do now to connect this artificial system to the one in which we are really interested is to choose the effective potential $V_S$ such that the total density from occupied Kohn-Sham orbitals $\phi_i$ exactly equals the ground state density of our real target system of interacting electrons,

$$\rho_S(r) = \sum_i^N \sum_s |\phi_i(r, s)|^2 = \rho_0(r)$$  \hspace{1cm} (1.15)

In order to do that, Kohn and Sham suggested separation of the functional $F[\rho]$

$$F[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho],$$  \hspace{1cm} (1.16)

into the kinetic energy of a non-interacting electron gas

$$T_s[\rho] = -\frac{1}{2} \sum_i^N \langle \phi_i | \nabla^2 | \phi_i \rangle$$  \hspace{1cm} (1.17)
the classical Coulomb energy (often called Hartree term)

\[ J[\rho] = \frac{1}{2} \int \frac{\rho(r)\rho(r')}{|r - r'|} dr dr', \]  

(1.18)

and the exchange-correlation energy:

\[ E_{xc}[\rho] \equiv \{ T[\rho] + V_{ee}[\rho] \} - \{ T_s[\rho] + J[\rho] \}. \]  

(1.19)

\( E_{xc} \) contains not only the non-classical effects of self-interaction correction, exchange and correlation, which are contributions to the potential energy of the system, but also a portion belonging to the kinetic energy (the non-interacting kinetic energy \( T_s \) is not equal to the true kinetic energy of the interacting system \( T \), even if the systems share the same density). This is a big unknown; construction of exchange-correlation functionals and their functional derivatives, exchange-correlation potentials: \( V_{xc}(r) = \frac{\delta E_{xc}[\rho(r)]}{\delta \rho(r)} \), is one of the most important and complicated steps in solving the many-body problems. Some approximate forms of the functionals have been successfully used for various problems in physics and chemistry; they will be discussed in detail in the following sections.

If the exact form of the exchange-correlation functional was known, \( V_S \) could be determined in the following way. Let us write down the expression for the energy of the interacting system in terms of the separation described by Eq. (1.16), highlighting the dependence on the orbitals shown in Eqs. (1.17, 1.18):

\[
E_0[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] + E_{Ne}[\rho] \\
= T_s[\rho] + \frac{1}{2} \int \frac{\rho(r)\rho(r')}{|r - r'|} dr dr' + E_{xc}[\rho] + \int V_{Ne}(r)\rho(r)dr \\
= -\frac{1}{2} \sum_i^N \langle \phi_i | \nabla^2 | \phi_i \rangle + \frac{1}{2} \sum_i^N \sum_j^N \int \int |\phi_i(r)|^2 \frac{1}{|r - r'|} |\phi_j(r')|^2 dr dr' \\
+E_{xc}[\rho] - \sum_i^N \int \sum_A^M \frac{Z_A}{|r - r_A|} |\phi_i(r)|^2 dr.
\]  

(1.20)

Applying variational principle to minimize this functional under the usual constraint \( \langle \phi_i | \phi_j \rangle = \delta_{ij} \) as in [9] we get

\[
\left( -\frac{1}{2} \nabla^2 + \int \frac{\rho(r')}{|r - r'|} dr' + V_{xc}(r) - \sum_A^M \frac{Z_A}{|r - r_A|} \right) \phi_i = \left( -\frac{1}{2} \nabla^2 + V_{eff}(r) \right) \phi = \epsilon_i \phi_i
\]  

(1.21)
Comparing this result with the one-particle equation from the non-interacting reference system (1.11) we obtain the expression for the Kohn-Sham potential:

$$V_S(r) \equiv V_{\text{eff}}(r) = \int \frac{\rho(r')}{|r - r'|} dr' + V_{xc}(r) - \sum_A \frac{Z_A}{|r - r_A|}. \tag{1.22}$$

Thus, once we know the various contributions in (1.22) we have a grip on the potential $V_S$ which we need to insert into the one-particle equations, which in turn determine the orbitals and hence the ground state density and the ground state energy by employing the energy expression (1.20). It should be noted that $V_{\text{eff}}$ already depends on the density (and thus on the orbitals) through the Coulomb term. Therefore, the Kohn-Sham one-electron equations (1.21) have to be solved iteratively.

One needs to point out that the wave functions $\phi_i(r)$ appearing in the Kohn-Sham equations describe electronic orbitals for the non-interacting gas; they are the eigenstates of the effective single electron non-interacting problem and should not be considered as wave functions for the electrons of the real system. It is only the total energy and the electronic density $\rho(r)$ that have a physical meaning. However, in many cases the Kohn-Sham energy levels give adequate description of the band structure of the real crystal.

### 1.2.3 Local Density and Generalized Gradient Approximations

To apply the Kohn-Sham scheme one needs to have an explicit form of the exchange-correlation energy, $E_{xc}[\rho(r)]$. Since the exact functional form for it is not known, the quality of the density functional method depends solely on the accuracy of the chosen approximation to $E_{xc}[\rho(r)]$.

Local Density Approximation (LDA), proposed in the original work by Kohn and Sham, is among the simplest approximations. It centers on the idea of a hypothetical uniform electron gas, where electrons move on a positive background charge distribution such that the total ensemble is electrically neutral. The situation is far from any realistic situation in atoms and molecules, which are characterized by rapidly varying electron densities. The attraction of this approach is that the form of the exchange and correlation energy functionals are known to very high accuracy and systematic refinement is in principle possible.

$E_{xc}[\rho(r)]$ in this approximation is equal to the exchange correlation energy per elec-
tron of a homogeneous electron gas with the same density $\rho(r)$:

$$E_{xc}^{LDA}[\rho(r)] = \int \epsilon_{xc}^{hom}(\rho(r)) \rho(r) \, dr,$$

where $\epsilon_{xc}^{hom}(\rho(r))$ is the exchange-correlation energy per particle of a uniform electron gas of density $\rho(r)$. This energy is weighted with the probability $\rho(r)$ that there is in fact an electron at this position in space. The quantity $\epsilon_{xc}^{hom}(\rho(r))$ can be further split into exchange and correlation contributions as,

$$\epsilon_{xc}^{hom}(\rho(r)) = \epsilon_{x}^{hom}(\rho(r)) + \epsilon_{c}^{hom}(\rho(r)).$$

(1.23)

The exchange part $\epsilon_{x}^{hom}(\rho(r))$ was originally derived by Bloch and Dirac in the late 1920’s:

$$\epsilon_{x}^{hom}(\rho(r)) = -\frac{3}{4} \left( \frac{3\rho(r)}{\pi} \right)^{\frac{1}{3}}$$

(1.24)

No such explicit expression is known for the correlation part $\epsilon_{c}^{hom}(\rho(r))$. However, highly accurate numerical Monte-Carlo simulations of the homogeneous electron gas are available from the work of Ceperly and Alder, 1980 [10]. On the basis of these results various authors presented analytical expressions of $\epsilon_{c}^{hom}(\rho(r))$ based on sophisticated interpolation schemes. One of the most accurate and popular is given by Perdew and Zunger, 1992 [11].

One would expect the LDA to perform well only for systems with slowly varying density. However, it also produces surprisingly good results for non-homogeneous systems, like covalently bonded materials and some transition metals. The agreement between theory and experiments is good for structural and vibrational properties. Among the drawbacks is a slight underestimation of the bonding energies and bond lengths. Furthermore, LDA underestimates the band gap for semiconductors, but implementation of the GW method or the Self-Interaction Correction, proposed by Perdew and Zunger, could alleviate this problem [11, 12, 13].

To extend the Local Density Approximation to systems with more significant non-homogeneous densities, several techniques have been proposed. The logical step in that direction was the suggestion to use the information not only about the density $\rho(r)$ at a particular point $r$, but also the gradient of the charge density, $\nabla \rho(r)$:

$$E_{xc}^{GGA}[\rho(r)] = \int \epsilon_{xc}^{GGA}(\rho(r), |\nabla \rho(r)|) \rho(r) \, dr,$$

(1.25)
This is the so-called Generalized Gradient Approximation (GGA). The energy per particle \( \epsilon_{xc}^{GGA} \) again needs to be approximated. The most commonly used parameterization for \( \epsilon_{xc}^{GGA} \) are PW91, proposed by Perdew and Wang [14, 15], and PBE [16], which is a simplified version of PW91. The GGA formalism usually outperforms the LDA method, providing better descriptions of inhomogeneous systems, like transition metals, and significantly improving the binding energy of molecules and solids. One needs to remember that even though the GGA exchange-correlation potential is often called semi-local, it still depends on only one point in space and does not give any information about nonlocal effects.

1.2.4 Nonlocal correlations

Further improvement of the exchange-correlation functionals beyond the LDA and the GGA is difficult. One of the earliest attempts, the Self-Interaction Correction to the LDA proposed by Perdew and Zunger in 1981, offers a number of advantages, such as more accurate total, binding, and removal energies and correct long-range behavior of the one-electron potential and density. But this becomes possible at the cost of the nonlocality of the orbitals (the value of the orbital at a particular point depends on all orbitals at each point in space), which makes this method impossible to realize in practice.

Nevertheless, for some important situations corrections to the exchange-correlation functional have been worked out and successfully implemented. For example, binding in layered structures such as graphite, BN, and MoS\(_2\) is attributed to weak orbital overlap and nonlocal Van der Waals (VdW) forces. As was pointed out by Rydberg et al. [17], standard LDA and GGA methods cannot describe this interaction properly because neither of them accounts for the nonlocal electrodynamic coupling between layers. Fig. 1.6 illustrates that in the LDA or GGA methods the interlayer binding for this system comes from the small region between the layers, where exponential tails from orbitals in different layers weakly overlap. As a result the interlayer binding is underestimated and has the wrong asymptotic behavior: exponential decay, characteristic of local DFT approximations, instead of the expected \( r^{-4} \) dependence from the VdW attraction.

Rydberg et al. [17] proposed inclusion of nonlocal correlations to the density functional, based on the calculation of polarizability of individual layers. The correlation energy functional is divided into two parts,

\[
E_c[\rho(r)] = E_c^{LDA}[\rho(r)] + E_c^{nl}[\rho(r)]
\]

(1.26)
where the $E_c^{\text{LDA}}[\rho(r)]$ term is nonlocal but approaches the LDA in the limit of a slowly varying density, and $E_c^{\text{nl}}[\rho(r)]$ is defined to include the longest ranged or most nonlocal terms that give the VdW interaction, and to approach zero in this limit. They argued that long range terms are less sensitive to details of the system’s dielectric response. Thus, very simple approximations for the dielectric function are made for $E_c^{\text{nl}}[\rho(r)]$. In particular, the charge density, calculated in GGA, is averaged in the lateral direction to yield $n(z)$. Included in this natural way, the nonlocal correlations give the right qualitative character of soft bonds along the $z$ direction, including saturation of the VdW potential at small separations, and better agreement of the interlayer spacing, binding energy, and compressibility with experimental values.

In the following chapter we further investigate the capabilities of local DFT in describing the weak interlayer binding.

### 1.3 Tight Binding method

The first principle methods are powerful tools for studying many-body problems but unfortunately, very demanding computationally. For a system of $N$ atoms the computation time usually scales as $N^3$, so these methods can presently handle systems no more than a few hundred atoms in size [18]. A tight binding approach is much simpler, yet provides important insights for understanding electronic properties of quantum mechanical systems. Parameterized with accurate *ab initio* calculations, the tight binding methods...
perform very well, when applied to structures around an equilibrium state, and also offer adequate description of transition states. In my research I have used tight binding calculations for the first stage of a reaction path search (Chapter 5) and for benchmarking the performances of empirical methods (Chapter 6). In this section I will give a brief description of the method.

1.3.1 Periodic boundary conditions in crystals

Because of the translational symmetry of the unit cells in the direction of the lattice vectors \( \mathbf{a}_i \) \((i = 1, 2, 3)\), any wave function of the lattice, \( \Psi \), should satisfy Bloch’s theorem:

\[
T_{\mathbf{a}_i} \Psi = e^{i \mathbf{k} \cdot \mathbf{a}_i} \Psi,
\]

where \( T_{\mathbf{a}_i} \) is a translational operator along the lattice vector \( \mathbf{a}_i \) and \( \mathbf{k} \) is the wave vector. The Bloch states are usually represented as a linear combination of plane waves. One of the drawbacks of this approach is the difficulty to relate the plane wave wavefunction to the atomic orbitals in the solid. A more natural form of the Bloch function \( \Phi_j(\mathbf{k}, \mathbf{r}) \) is given by

\[
\Phi_j(\mathbf{k}, \mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}} e^{i \mathbf{k} \cdot \mathbf{R}} \phi_j(\mathbf{r} - \mathbf{R}) \quad (j = 1, \cdots, n).
\]

Here \( \mathbf{R} \) is the position of the atom and \( \phi_j \) denotes the atomic wavefunction in state \( j \). The number of atomic wavefunctions in the unit cell is denoted by \( n \).

Periodic boundary conditions impose a restriction on possible values of the wave vector \( \mathbf{k} \). The wavefunction of a crystal consisting of \( N_i \) unit cells in each direction \((i = 1, 2, 3)\), should be seamless on the boundaries of the crystal:

\[
\Phi_j(\mathbf{k}, \mathbf{r} + N_i \mathbf{a}_i) = e^{i N_i \mathbf{k} \cdot \mathbf{a}_i} \Phi_j(\mathbf{k}, \mathbf{r}),
\]

which is possible only when the phase factor is equal to one. Therefore the allowed values of \( \mathbf{k} \) should satisfy the following condition:

\[
k_i = \frac{2\pi p}{N_i \mathbf{a}_i}, \quad (p = 0, \cdots, N_i - 1), \quad (i = 1, 2, 3),
\]

Due to the large number of unit cells in a three-dimensional crystal \((N_1 N_2 N_3 = N \approx 10^{24})\) wave vectors \( k_x, k_y, \) and \( k_z \) are considered as continuum variables.
The eigenfunctions in the solid $\Psi_j(k, r)$ ($j = 1, \cdots, n$) are expressed as a linear combination of Bloch functions $\Phi_{j'}(k, r)$ as follows:

$$\Psi_j(k, r + a) = \sum_{j'=1}^n C_{jj'}(k) \Phi_{j'}(k, r), \quad (1.31)$$

where $C_{jj'}(k)$ are coefficients to be determined. Since the functions $\Psi_j(k, r)$ should also satisfy Bloch’s theorem, the summation in Eq. (1.31) is taken only for the Bloch orbitals $\Phi_{j'}(k, r)$ with the same value of $k$.

### 1.3.2 Secular equation

If $\hat{H}$ is the Hamiltonian of the system the $j$-th eigenvalue $E_j(k)$ ($j = 1, \cdots, n$) as a function of $k$ is given by

$$E_j(k) = \frac{\langle \Psi_j | \hat{H} | \Psi_j \rangle}{\langle \Psi_j | \Psi_j \rangle} = \int \frac{\Psi_j^* \hat{H} \Psi_j}{\Psi_j^* \Psi_j} dr \quad (1.32)$$

Substituting Eq. (1.31) into Eq. (1.32) and making a change of subscripts, we obtain the following equation,

$$E_i(k) = \frac{\sum_{j,j'=1}^n C_{ij}^* C_{ij'} \langle \Phi_j | \hat{H} | \Phi_{j'} \rangle}{\sum_{j,j'=1}^n C_{ij}^* C_{ij'} \langle \Phi_j | \Phi_{j'} \rangle} = \frac{\sum_{j,j'=1}^n H_{jj'} C_{ij}^* C_{ij'}}{\sum_{j,j'=1}^n S_{jj'} C_{ij}^* C_{ij'}} \quad (1.33)$$

where the integrals over the Bloch orbitals $H_{jj'}(k)$ and $S_{jj'}(k)$ are called transfer integral and overlap integral matrices respectively, which are defined by

$$H_{jj'} \equiv \langle \Phi_j | \hat{H} | \Phi_{j'} \rangle, \quad S_{jj'} \equiv \langle \Phi_j | \Phi_{j'} \rangle \quad (j,j' = 1, \cdots, n). \quad (1.34)$$

In order to find the eigenvalues $E_i(k)$ the values of the $n \times n$ matrices $H_{jj'}(k)$ and $S_{jj'}(k)$ are fixed for a given value of $k$ and the coefficients $C_{ij}$ are optimized so that the partial derivatives $\frac{\partial E_i(k)}{\partial C_{ij}}$ are zero. When we work out the system of linear equations the eigenvalues will be given as a solution of the secular equation:

$$\text{det} [H - ES] = 0 \quad (1.35)$$

The secular equation (1.35) is solved for each value of $k$. The total band structure energy is an integral over all occupied states in the Brillouin zone. For complicated multidimensional potential surfaces the integration is replaced by a summation over
special high symmetry k-points [19]. In order to obtain the energy dispersion relations the calculations are done for a whole range of $k$ values. In the case of a two or three dimensional crystal it is done along high symmetry directions in the Brillouin zone.

One needs to stress that standard tight binding calculations are not self-consistent. The transfer and overlap integrals are kept constant throughout the energy minimization steps. Since potential is not adjusted after the orbitals are found the screening is usually underestimated in tight binding calculations. One should also be careful when the method is applied to systems with considerable charge transfer. The transfer and overlap integrals are usually tabulated for neutral systems and might not be accurate in this case.
Chapter 2

Interlayer interaction in graphitic structures I: Theoretical models

2.1 Review

The discovery of carbon nanotubes [4] has opened up the possibility of building electronic and mechanical carbon-based devices at a nanoscale. Made from strongly bound \( sp^2 \) carbon networks, nanotubes interact only weakly with one another. But it is this interaction that will define the function of the devices as nanoscale bearings, motors, etc [20, 21]. The nanotube-graphite interaction could also be important for distinguishing nanotubes by helical angle [22]. Large-scale simulations of nanomechanical systems are beyond capabilities of modern quantum mechanical methods, so empirical potentials must be applied. A pairwise Lennard-Jones (LJ) potential can successfully describe the total cohesion between graphene layers, but is much too smooth under variation in the relative alignment of the layers [23]. To account for this registry dependence and realistically simulate the interaction of graphitic interfaces containing thousands of atoms we have developed a new empirical interlayer potential making use of the latest experimental and theoretical knowledge about the interaction.

The cohesion in layered graphitic structures is a combination of long-range Van der Waals (VdW) and short-ranged \( \pi \) overlap contributions. The VdW coupling is missing in the local density (LDA) and the generalized gradient approximations (GGA) to DFT, so these methods do not adequately describe the interlayer cohesion in graphite [23, 17]. An important step towards a more complete model has been taken by Rydberg et al. [17], who included the long-ranged VdW coupling as a nonlocal functional of charge density incorporating the polarizability of the graphitic layers. For calculation
of the dielectric function in [17] the charge density was smoothed out in the in-plane directions. This is a good starting point for improving the description of the total interlayer binding; for revealing the registry dependence of the VdW term one would need to go beyond the jellium model. A. Palser, who proposed a tight-binding-plus-dispersion model [24], suggested that the differences in the binding energy for different forms of graphite calculated in LDA gives a qualitative measure for the registry dependence of the interlayer graphitic interaction.

In the present work we argue that although the absolute cohesion is not properly described in the local DFT approximations, the variation in the binding energy under interlayer shifts appears to be much more sound than previously suspected. We combine this result with experimental data to introduce a new empirical potential, which describes the variation in the interlayer binding with the relative alignment (registry) of the layers. Henceforth, we refer to this variation as corrugation for brevity.

### 2.2 Experimental data

Experimental measurements are usually done on natural graphite, which consists of 80% AB and 14% ABC stackings and 6% turbostratic (Fig. 2.1).

The equilibrium interlayer distance (3.34 Å [38]) and compressibility (2.7 cm²/dyn [39]) along the c-axis of graphite can be measured with good accuracy. The absolute interlayer binding energy is harder to extract; only a handful of results exist in the
literature. A heat of wetting experiment by Girifalco gives an exfoliation energy of 43 meV/atom [25]. A microscopic determination of the interlayer binding through study of collapsed nanotubes yields $35 \pm 10$ meV/atom [26]. Later experiments on thermal desorption of polyaromatic hydrocarbons give a larger binding energy of $61\pm5$ meV/atom [27]. The discrepancies in the experimental values might arise from different structural constraints in the studied systems. Conjugated molecules on a graphitic substrate are the least constrained system: they have rotational and translational degrees of freedom. Thus, the molecules can find an optimal finite-size registration with the substrate and achieve higher cohesion. In contrast, collapsed nanotubes are highly constrained due to circumferential continuity; the opposite walls of a nanotube of arbitrary helicity are skewed and therefore approximate turbostratic graphite (material with randomly rotated layers, see Fig. 2.1), which has weaker binding. The increase in the interlayer spacing of turbostratic graphite due to the random orientation of graphene layers can be used as an independent measure of the registry dependent interlayer interaction. Experimental data on AA stacking (Fig. 2.1a) would help reveal the registry dependence but this allotropic form has not been observed in nature.

2.3 \textit{ab initio} results

Because there are no direct experimental data on registry dependence in graphite we investigate what information about the interaction can be obtained theoretically. We perform first-principle calculations using VASP [28, 29, 30, 31] with ultrasoft pseudopotentials [32]. The scale of most of the effects we study is a few meV/atom, so high accuracy is required. We use an energy cutoff of 358 eV with a sufficient number of special k-points [19] to ensure convergence of our results. More details are discussed in Section 2.4.3.

Fig. 2.2 illustrates how LDA and GGA functionals describe the interplanar binding of graphite. The LDA finds the right interlayer spacing for AB stacking, but it underestimates the exfoliation energy and overestimates the compressibility by a factor of roughly two. The GGA shows no binding at all at physically meaningful spacings. The exfoliation energies are strongly affected by the locality of the exchange-correlation functionals used in the two approximations. First, the electrodynamic coupling between the layers is missing resulting in the weaker binding. Second, the two methods employ different approximate forms for the exchange-correlation functional [11]; by construction they differ the most in regions of rapidly varying charge density. Due to the large interlayer
spacing in graphite the orbital overlap between neighboring layers occurs in the region of exponentially decaying charge density. This must be the reason for the quite different exfoliation energies produced by the LDA and GGA functionals. For these reasons, standard DFT calculations for these structures have been looked at with skepticism and the need for nonlocal correlations is evident [17].

However, Fig. 2.2b shows that even without the nonlocal corrections the difference between the AA and AB stackings of graphite at constant interlayer spacing is essentially the same in the LDA and GGA approximations (15 meV/atom at 3.34 Å). The result is also independent of the details of the calculations: the hard pseudopotential [33], ultrasoft pseudopotential [32] and the projector augmented waves [34] approaches give very close values for the corrugation if sufficient number of k-points is used for convergence.
The accuracy of our relative energy calculations can be analyzed by breaking down the total corrugation into individual contributions from different terms (plotted in Fig. 2.3). The exchange-correlation term should be the least accurate, since the explicit form for nonlocal correlation effects is not known. However, the magnitude of its variation under sliding is the smallest of the four. The local description of the correlation also affects the accuracy of the other terms, but weakly and indirectly through the charge density.

While the absence of the VdW attraction leads to the reduction of the absolute binding, it should have little effect on the magnitude of corrugation. This term is long-ranged, therefore the interaction energy is an average over large area and depends more on the distance between the layers than on their relative alignment. A simple estimate of the VdW contribution, taken in the form of an \( r^{-6} \) dependence (as in the LJ potential [35], Eq. 2.1), gives a corrugation just 2% of the LDA value. The long-ranged correlations could play a somewhat bigger role in the corrugation to sliding for structures of lower dimensionality (nanotubes, fullerenes) due to less efficient averaging. Another possible increase in the VdW contribution to corrugation could be caused by local field effects as a careful study of the out-of-plane polarizability suggests [36]. It still looks likely that this is only a minor correction to the corrugation. To sum up, our tests indicate that corrugation comes primarily from short-ranged, exponentially decaying orbital overlaps. The corrugation coming from this contribution is adequately described in the local DFT.
approximations, and by shifting/rotating one layer with respect to the other one can extract physically meaningful information about the interaction.

2.4 Registry Dependent Potential

Now we can check how empirical potentials compare against the registry dependence calculated in the LDA. We consider the widely used LJ pair potential in the form given in [35]:

$$V_{LJ}(r) = 4\epsilon \left[ -\left(\frac{\sigma}{r}\right)^6 + \left(\frac{\sigma}{r}\right)^{12} \right], \quad (2.1)$$

where for graphene/graphene $\epsilon = 2.39$ meV and $\sigma = 3.41$ Å. The LJ curve shown in Fig. 2.2a is in good agreement with experiment: a two-parameter potential reproduces the equilibrium spacing, cohesion and the compressibility along the c-axis. This justifies its popularity for the description of averaged interlayer interactions. But Fig. 2.2b reveals that the potential is much too smooth under interlayer shift: at the equilibrium spacing the corrugation is less than 10% of that found in the LDA calculation. The corrugation in the LJ potential comes predominantly from the more singular $r^{-12}$ repulsive term, but it is still insufficient to reproduce the LDA results. The real $\pi$ overlap is anisotropic, defining a specific direction along the c-axis and requires a more detailed description. If the potential depends only on the radial distance between two atoms, then the compressibility along the c-axis and the corrugation are closely related and cannot be fit simultaneously. To decouple these dependencies one has to differentiate between in-plane and out-of-plane directions.

2.4.1 Original registry dependent potential

To study interlayer sliding in MWCNs we proposed a new registry dependent interatomic potential [37]. This original potential has an $r^{-6}$ two-body Van der Waals attraction, an exponential atomic-core repulsion, and a short-ranged term describing the energy gain due to interlayer delocalization of $\pi$ orbitals; this term dominates corrugation against interlayer sliding. We parameterize this contribution as a short-ranged function of the transverse separation $\rho$ between atoms in different layers, times a factor which is exponential in the interlayer spacing $z$. 
\[ V(r_{ij}) = -\left(\frac{r_{ij}}{d}\right)^{-6} + e^{-\lambda_1 (r_{ij} - r_0)} + e^{-\lambda_2 (z_{ij} - z_0)} e^{-\left(\frac{\rho_{ij}}{\delta}\right)^2} \sum C_{2n} \left(\frac{\rho_{ij}}{\delta}\right)^{2n}. \]

(2.2)

The units of energy are meV; \(d, r_0, \lambda_1, \lambda_2, z_0, \delta,\) and \(C_{2n}\) are fitting parameters.

We fit these parameters to a careful combination of experimental and \textit{ab initio} results. From experiment, we take the surface energy near optimal registry (approximately 40 meV/atom[25, 26]), the equilibrium interplanar spacing (3.34 Å [38]), and the \(c\)-axis compressibility of graphite (2.7 \times 10^{-12} \text{ cm}^2/\text{dyn} [39]). Since the calculated energy difference between AB and ABC graphite is very small [40] we describe the corrugation with a two-body potential, which gives nearly identical energies for AB and ABC stackings. As discussed at the beginning of this chapter, the local density approximation does not accurately describe the Van der Waals interaction, so we use it only to calculate the \textit{variation} in energy under changes in interlayer registry, and take the other quantities from experiment. We focus on two-layer graphene (see Fig. 2.4), since we apply this potential to double-wall nanotubes. (Unlike bulk graphite, this two-layered rigid system has a minimum slightly off from AB stacking, whereas the paired nature of the potential induces a small flat region near AB stacking.) The results are consistent with those for bulk graphite[41].

The resulting coefficients are: \(C_0 = 11.964, C_2 = 6.728, C_4 = -18.418, C_6 = 9.836, C_8 = -1.8938, C_{10} = -0.6391, C_{12} = 0.08652, \delta = 0.568 \text{ Å},\) and \(\lambda_2 = 3.444 \text{ Å}^{-1}.\) For convenience, we introduce the parameter \(z_0 = 3.44 \text{ Å}.\) The interlayer \(\pi\) overlap term is very short-ranged, falling essentially to zero at two transverse interatomic distances. Different interplanar separations give very similar \(\rho\)-dependences which scale exponentially in the interlayer distance (see the insert in Fig. 2.4). Such exponential behavior is expected for wavefunction overlap and is also characteristic of the local density approximation. The aforementioned experimental results then fix the remaining parameters: \(r_0 = 4.00 \text{ Å}, \lambda_1 = 4.19 \text{ Å}^{-1},\) and \(d = 4.68 \text{ Å}.\) There is one ambiguity: the data cannot precisely separate the delocalization and Van der Waals contributions to the uncorrugated component of the attractive interaction. For simplicity, we assume that the Van der Waals term dominates the interlayer attractive interaction for AA stacking because of the lack of interlayer \(\pi-\pi^*\) dispersion at this registry. Curvature-induced rehybridization in nanotubes is second-order in the inverse radius of curvature, so we anticipate that the graphene bilayer data is sufficient to describe multilayered nanotubes above
Figure 2.4. The interlayer binding energy versus parallel shift for two rigid graphitic layers 3.44 Å apart. The dots are the local density approximation data; the solid curve is the fitted delocalization term. The insert gives $E_{AA} - E_{AB}$ (normalized to 1 at 3.44 Å) as a function of the interplanar distance $z$ with a fit to an exponential.

$\sim 1$ nm diameter. We estimate that the final results for corrugation are typically accurate to about 50%. Sources of error include many-body and angular-dependent Van der Waals contributions, ambiguities in determining the registry-independent portion of the electron delocalization term and an absence of fitting to rotationally skewed graphitic bilayers.

We used this potential (we will refer to it as registry-dependent potential zero, or RDP0) to study interwall sliding in MWCNs [37]. According to our latest LDA calculations the potential tends to overestimate the corrugation in nanotube/graphene and nanotube/nanotube systems, as discussed later in Section 2.4.3. In the following section we develop an improved empirical model of the interlayer interaction.

### 2.4.2 Improved registry dependent model

The original potential RDP0 [37] fits the set of experimental data for AB stacked graphite and is in a reasonable agreement with the LDA results on corrugation to sliding. In the present work we improve the potential by addressing a few important issues. First,
RDP0 was fitted to sliding of two graphene sheets. Our latest LDA calculations for nanotube/graphene and nanotube/nanotube interaction showed that RDP0 tends to overestimate the corrugation. Now we use a larger set of data to train our new potential. Second, the in-plane and out-of-plane distances were defined globally assuming knowledge about the symmetry of the system, which prohibited its application to structures of arbitrary geometry. This time we set up the normal direction using only a small patch of neighboring atoms to reflect the fact that the orientation of the $p_z$ orbital on an $sp^2$ carbon site is determined by the local environment.

A simple way to define a plane for an $sp^2$-bonded carbon atom is based on its three neighbors. We calculate cross products for each pair of displacement vectors to neighbors and take the normal as the average of the three normalized vector products. These local normals tend to tilt towards the center of a hexagon for nanotubes, more so for smaller-diameter tubes. A broader definition could include second neighbors as well: after the local normals are found the direction of a semi-local normal is calculated as

$$n_i \propto an_{i1}^{loc} + n_{i1}^{loc} + n_{i2}^{loc} + n_{i3}^{loc}$$ (2.4)

where $n_{i1}^{loc}, n_{i2}^{loc}$, and $n_{i3}^{loc}$ are local normals for the first neighbors of atom $i$. The parameter $a$ interpolates between the two cases ($a = \infty$ corresponds to the local normals).

Once a normal is found the transverse distance is evaluated as the projection of the radius vector onto the plane perpendicular to the normal (see Fig. 2.5).
Our potential has an $r^{-6}$ Van der Waals attraction and an exponentially decaying repulsion due to the interlayer wavefunction overlap. To reflect the directionality of the overlap we introduce a function $f$, which rapidly decays with the transverse distance $\rho$:

$$V(r_{ij}) = -A \left( \frac{r_{ij}}{z_0} \right)^{-6} + e^{-\lambda(r_{ij}-z_0)}(C + f(\rho_{ij}) + f(\rho_{ji})) \quad (2.5)$$

$$\rho_{ij}^2 = r_{ij}^2 - (n_i r_{ij})^2, \quad \rho_{ji}^2 = r_{ij}^2 - (n_j r_{ij})^2. \quad (2.6)$$

$$f(\rho) = e^{-(\rho/\delta)^2} \sum C_{2n}(\rho/\delta)^{2n}. \quad (2.7)$$

The energy is measured in meV, and the vector $n_k (k = i, j)$ is a normal vector for atom $k$ (local or semi-local).

We use a conjugate gradient technique to adjust the seven parameters simultaneously by minimizing the least square error function. The experimental and theoretical data are weighed according to their accuracy. The compressibility ($2.7$ cm$^2$/dyn) and the equilibrium interlayer spacing ($3.34$ Å) of the AB stacking count the most and the resulting values closely match the measured ones. The experimental value of the interlayer binding energy for natural graphite is less accurate; our potential gives $48$ meV/atom. From the LDA calculations we took sliding data for graphene/graphene (shown in Fig 2.2b) and nanotube/graphene (for $(5,5)$ and $(10,0)$ nanotubes as shown in Fig 2.6) interaction. The resulting parameters are: $C_0 = 15.71$, $C_2 = 12.29$, $C_4 = 4.933$, $C = 1.515$, $\delta = 0.578$ Å, $\lambda = 3.629$ Å$^{-1}$, $A = 10.238$; the scaling factor $z_0 = 3.34$ Å is added for convenience.

### 2.4.3 Tests

Fig. 2.6 shows corrugation to sliding for armchair and zigzag nanotubes. Our potential with locally defined normals (registry dependent potential one, or RDP1) performs well, producing errors no more than $50\%$ compared to the LDA data. The consistent overestimation is likely caused by the curvature-induced rehybridization of orbitals [42] on a nanotube. The accuracy may also be affected by the way the normals are defined. To investigate this, we use semi-local normals (RDP2) with $a = 0.25$ in equation (2.4). The value of the parameter $a$ is chosen so that the normals are close to the radial direction in small-diameter nanotubes (as was mentioned earlier, local normals tilt towards the center of a hexagon). The RDP2 results are in better agreement with the LDA data; the biggest difference occurs, as expected, for nanotubes with smallest radius of curvature. The advantage of the RDP1 is that the accuracy remains almost constant for a large class of tubes. We used the local definition of the normals in our study of the nanotube/graphite interaction [37] to make the comparison of alignment effects for tubes of
Figure 2.6. Corrugation to sliding for the nanotube/substrate interaction per atom of nanotube. The nanotubes are positioned 3.2 Å above a graphitic substrate; the energy difference is evaluated for structures when the lowest row of atoms on the nanotube is in AA or AB registration with the graphene layer. Armchair nanotubes, from left to right: (5,5), (6,6), (8,8), (10,10), (12,12), (15,15), (20,20); zigzag nanotubes, from left to right: (10,0), (16,0), (20,0), (24,0), (28,0).

The nanotube/nanotube tests are even more challenging, since the relative distribution of atoms on different walls is smeared out (as discussed, for example, in [24]) and the resulting corrugation to rotation and sliding are very small. For double-walled carbon nanotubes with incommensurate axial periods the uniform smearing over all possible registrations causes the total corrugation not to grow with the system size as was theoretically predicted by Damnjanovic et al. [43] and later observed experimentally [20, 44] and numerically [37]. Even when nanotubes are axially commensurate the lack of circumferential period less than $2\pi$ can lead to efficient averaging of the intertube interaction and allow smooth relative motion of nested nanotubes.

One of the special cases, widely discussed in the literature [24, 45], is a (5,5)/(10,10) system with a short angular ($\pi/10$) and axial ($\sqrt{3}/2a_0$) periods. The maximum corrugation is achieved when one of the nanotubes is shifted by a quarter of the axial period or half of the angular period. Our LDA corrugation, shown in table 2.1, is lower than the respective values calculated in [45]: 14 meV for sliding and 31 meV for rotation. We attribute this discrepancy to the small magnitude of the corrugation (less than 1
Table 2.1. Barriers to sliding and rotation in meV per unit cell of a double-wall carbon nanotube (60, 108, 76 atoms respectively). The last column shows barriers at lower cut-off energy of 287 eV.

<table>
<thead>
<tr>
<th>DWCN</th>
<th>Motion</th>
<th>LDA</th>
<th>RDP0</th>
<th>RDP1</th>
<th>RDP2</th>
<th>LJ</th>
<th>LDA*</th>
</tr>
</thead>
<tbody>
<tr>
<td>0505/1010</td>
<td>slide</td>
<td>&lt; 1</td>
<td>20</td>
<td>25</td>
<td>2</td>
<td>-0.3</td>
<td>-34</td>
</tr>
<tr>
<td></td>
<td>rotate</td>
<td>16</td>
<td>55</td>
<td>-91</td>
<td>18</td>
<td>-0.9</td>
<td>19</td>
</tr>
<tr>
<td>0900/1800</td>
<td>slide</td>
<td>-110</td>
<td>-280</td>
<td>-149</td>
<td>-154</td>
<td>-11</td>
<td>-120</td>
</tr>
<tr>
<td></td>
<td>rotate</td>
<td>&lt; 1</td>
<td>36</td>
<td>29</td>
<td>10</td>
<td>&lt; 0.1</td>
<td>-5</td>
</tr>
<tr>
<td>0707/1212</td>
<td>slide</td>
<td>&lt; 1</td>
<td>48</td>
<td>6</td>
<td>6</td>
<td>-0.4</td>
<td>-40</td>
</tr>
</tbody>
</table>

meV/atom), which makes the results more sensitive to details of calculations, such as a choice of pseudopotential and convergence criteria. If, for example, we use a lower energy cut-off of 287 eV in our soft-pseudopotential calculations, some corrugation values increase significantly, as shown in the last column of table 2.1 (9 special points along the axis provide good k-point convergence for both zigzag and armchair tubes). For comparison, at this cut-off the difference between AA and AB stacking of graphite is just 10% lower than the converged value of 15 meV/atom at cut-off energy of 358 eV.

It is therefore difficult to evaluate the performances of classical (the LJ and our) potentials in this case. Table 2.1 shows that the fine averaging of interlayer registries for small-diameter nanotubes is described better with semi-local normals, so RDP2 is more suitable for these systems. The LJ results for some nanotubes with small values of corrugation might seem to be in better agreement with LDA data. We think that this is related to the general trend of the LJ potential to underestimate the registry effects, which can be seen in a few cases with relatively large corrugation (sliding in (5,5)/(10,10) and (9,0)/(18,0) systems). Interwalled motion in chiral nanotubes would be an even more difficult test since the corrugation is expected to be even smaller and beyond the present accuracy of either classical or ab initio methods. We would like to stress that the LDA results for nanotube/nanotube interaction should be used with caution; further adjustment of classical models will require more reliable ab initio data.

The importance of the corrugation for flat surfaces can be tested against the equilibrium interlayer spacing in turbostratic graphite. In this form of graphite the layers are randomly rotated with respect to one another which makes atoms in different layers be in all possible registrations. Therefore one expects the equilibrium interlayer distance to be somewhere midway between the optimal AB and AA spacings. Various measurements produce close results for the equilibrium interlayer spacing: 3.44 Å [46], a noticeable
Table 2.2. The equilibrium interlayer spacing for different allotropic forms of graphite. The theoretical value for turbostratic graphite, calculated with the RDP1, is in much better agreement with experiment than the value produced by the LJ potential.

<table>
<thead>
<tr>
<th>Stacking</th>
<th>LJ</th>
<th>RDP1</th>
<th>Exp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AB</td>
<td>3.34</td>
<td>3.34</td>
<td>3.34</td>
</tr>
<tr>
<td>Turbostratic</td>
<td>3.34</td>
<td>3.43</td>
<td>3.44</td>
</tr>
<tr>
<td>AA</td>
<td>3.36</td>
<td>3.59</td>
<td></td>
</tr>
</tbody>
</table>

increase compared to natural graphite.

We simulated the three allotropic (AA, AB, and turbostratic) forms of graphite with the LJ [47] and our potentials (for flat surfaces local and semilocal normals are the same); the results are shown in Table 2.4.3.

The interlayer spacing given by the present potential, not fitted to this data directly, matches the experimental value surprisingly well. The perfect fit is probably accidental but it does indicate the right magnitude of the registry dependence necessary for description of turbostratic graphite. As in the other examples throughout the paper the LJ potential turns out to be too smooth and does not distinguish different stackings of graphite properly.

Our potential does not have an explicit dependence on the relative orientation of hexagons in different layers. To test this dependence with first-principle calculations we consider skewed configurations allowing periodicity. Orientations with accidental commensurations have been observed in graphite [48] and Fig. 2.7 shows how such a structure can be constructed. Two patches of hexagonal lattice containing 14 atoms each are rotated by acos(11/14) with respect to each other. Now the atoms are in a different registration but the resulting unit cell retains its original shape. This allows a direct comparison of the aligned and the skewed structures.

We find that the energy of the mismatched system at 3.34 Å is 11.6 meV/atom higher than that of the AB stacking and the corrugation to sliding is very small. A similar increase in energy with respect to the AB stacking (11.8 meV/atom) and practically no corrugation to sliding are observed for a larger structure with a 26 plus 26 atom unit cell and an acos(11/13) angular mismatch. Our potential gives 12.0 meV/atom for both cases. The results indicate that even for these small unit cell structures the interlayer interaction averages out over all possible registrations and we obtain the mean binding energy.
Figure 2.7. An example of accidental commensuration in graphite. The two layers become commensurate when rotated by $\acos(11/14)$ with respect to each other. The resulting skewed unit cell has 14 atoms in each layer.

In summary, our tests have shown that although the absolute cohesion is not properly modeled in the local DFT approximations, the corrugation is described more accurately than previously suspected. The corrugation for flat and curved interfaces in the LDA and GGA are nearly identical. Based on these results and experimental data we introduce a new empirical registry-dependent potential (RDP1). It matches the LDA values of the corrugation for flat graphitic layers and reproduces the experimentally observed increase in the interlayer spacing in turbostratic graphite. The potential is also in good agreement with the LDA results for the nanotube/graphene system; the small overestimation of the corrugation in this case is probably caused by the curvature-induced rehybridization of the orbitals. The effect is more pronounced in nanotube/nanotube systems. We propose a way to adjust our potential (RDP2) to this effect; the fine tuning though will require more reliable \textit{ab initio} data.

The results of this work will be submitted to Physical Review B shortly.
Chapter 3

Interlayer interaction in graphitic structures II: Applications

3.1 Nanotube-nanotube interactions

3.1.1 Manipulation of multiwalled carbon nanotubes

Modern experimental techniques allow researchers to handle individual nanotubes [20, 49] which offers an intriguing possibility of building devices on nanoscale. J. Cummins and A. Zettl [20] reported a controlled and reversible telescopic extension of multiwalled carbon nanotubes, thus realizing ultralow-friction nanoscale linear bearings and constant-force nanosprings. By using high resolution transmission electron microscope, they placed quantitative limits on the static and dynamic interwall frictional force between the nested tubes. Repeated extension and retraction of telescoping nanotube segments revealed no wear or fatigue on the atomic scale. On the basis of their experimental study the authors suggested that these nanotubes may constitute near-perfect, wear-free surfaces.

In another interesting experiment R. Ruoff et al. [44] studied sliding between nested shells of MWCNs under tensile load. They observed a stick-slip motion and a smooth pullout motion and modeled the intertube sliding by taking into account the shear interaction, the capillary effect, and the edge effect. By calibrating the force constants of the cantilevers used in the experiment they were able to estimate the shear strength between the walls and the surface energy of the MWCNs.
3.1.2 Simulation of nanotube-nanotube sliding

M. Damnjanovic et al. have done a thorough study of properties of single- and multi-walled carbon nanotubes based on the symmetry of these quasi-one dimensional objects [43, 50]. In one of their first papers [43] they showed that optical activity and interwall interaction of nested nanotubes depend critically on the symmetry of the combined system. In particular, they predicted that intertube sliding and rotation in infinite MWCN’s with incommensurate axial periods is frictionless.

In our work we discuss issues related to intertube motion in real nanotubes and simulate the intertube sliding for a few special cases. Using our preliminary interlayer potential (see Section 2.4.1) we make an estimate of the magnitude of barriers to intertube sliding and compare them with experimental values. More importantly, we demonstrate that the tightly constrained geometry of a multiwalled nanotube can produce an extremely smooth solid-solid interface not only in incommensurate tubes, but also, surprisingly, in certain axial commensurate tubes. The energetic barrier to interlayer sliding in defect-free nanotubes containing thousands of atoms can be comparable to that for a single unit cell of crystalline graphite. To our knowledge these systems represent the smoothest solid-solid interfaces consisting of the same element.

The tribology of graphitic lubrication is a very complex heterogeneous process. In contrast, concentric multiwalled carbon nanotubes present a very simple geometry, which restricts interlayer motion to a single (axial) direction with a fixed interlayer orientation of stiff, smooth layers. The stiffness suppresses shear-induced dislocation slip and the small system size promises the possibility of attaining nearly defect-free structures. Sliding within a many-walled nanotube should occur between the two layers with the smoothest interface, so we focus on two-walled systems here.

As was discussed in Chapter 1, each layer in a concentric double-walled carbon nanotube (DWCN) is indexed by two integers \((n, m)\) which give the circumference in graphitic lattice coordinates. The difference in radii between successive layers frustrates the circumferential interlayer registry. Two layers with \(n_1/m_1 \neq n_2/m_2\) are typically (but not always [43]) axially incommensurate also. In contrast, two layers with \(n_1/m_1 = n_2/m_2\) (and other special cases) are axially commensurate. We divide nanotubes into three classes, which are expected to exhibit different behavior for interwall sliding: axially commensurate tubes with a short and long periods and axially incommensurate tubes. The representative DWCN systems are: commensurate, short-period \((7,7)/(12,12)\); commensurate, long-period \((11,2)/(12,12)\); and incommensurate \((14,0)/(16,10)\) (see Fig. 3.1).
Figure 3.1. Three classes of DWCN’s with short commensurate, long commensurate, and incommensurate periods.

We calculate the barriers to sliding using the following setup. To isolate the effects of variations in contact area (where “contact” is broadly defined as stretching up to a cutoff distance in the interlayer interaction) and to treat both commensurate and incommensurate finite-sized systems on an equal footing, we extend the inner wall of the tube while keeping the outer layer as a relatively short collar; in this configuration the motion of the outer layer does not change the contact area. Forces induced by changes in contact area are easily incorporated later using the surface energy of graphite [25, 26].

We define the interlayer corrugation as the maximal variation in total energy as the outer layer slides through the full unit cell of the inner layer. This quantity is straightforward to extract theoretically; it is distinct from, yet bears a strong relation to, both shear modulus and frictional coefficient. We focus on the corrugation so defined because, as we shall see later, the interlayer shear modulus is actually ill-defined for a large class of tubes. First we analyze perfectly rigid tubes; next we introduce the intraplanar degrees of freedom.

As expected, the (7,7)/(12,12) system shows a relatively large barrier that increases linearly with system size (see Fig. 3.2). This result is consistent with the preference for commensurate tube/substrate orientations for tubes on graphitic substrates [51]. This linear size dependence allows definition of a shear modulus, at least as reflects variation...
Figure 3.2. The corrugation as a function of number of atoms in the outer tube in a double-walled system without structural relaxation. In the commensurate \((7,7)/(12,12)\) tube the corrugation grows linearly with a slope of 20 meV/Å. The corrugation in the incommensurate \((14,0)/(16,10)\) tube does not grow with the system size. The commensurate \((11,2)/(12,12)\) system has extremely small corrugation per period (i.e. 336 atoms of the outer tube). The curve is a guide for the eye.

in the axial length.

The other two systems show intriguing behaviors. The commensurate but mixed-wrapping-angle \((11,2)/(12,12)\) nanotube (Fig. 3.2) shows nearly zero corrugation for lengths of the outer layer which are integral and half-integral numbers of the overall lattice constant. The growth in corrugation with length is therefore linear, but extremely slow. The averaging in local interlayer registry across the unit cell is very efficient, much more so than would be a random distribution of local registries. From the figure it might appear that the minimal barrier is exactly zero; however, artificially sharpening the interlayer potential by reducing the interlayer \(\pi\) overlap decay length \(\delta\) (defined in Section 2.4.1) by a factor of two increases the corrugation significantly above zero, thereby revealing that the corrugation for the pure system at the physically relevant parameters is very small, but not zero.
The incommensurate \((14,0)/(16,10)\) nanotube has a corrugation from 0.0 – 0.3 eV which, excepting fluctuations, is independent of system size. The interlayer sliding mode (phason) of a perfectly rigid incommensurate system is gapless [52] (the case of not perfectly rigid nanotubes is discussed in the next paragraph). For finite (and therefore imperfectly) incommensurate systems, a barrier to sliding arises from fluctuations in the finite-sampled registry-averaged interlayer binding energy. The size of the fluctuations depends on the number and distribution of the sample points for the different local registries. The unusual nonextensive behavior observed here is in fact a generic behavior of incommensurate systems, such as e.g. the average value of \(\sin(x)\) when finite-sampled an incommensurate period. The minimal value of the reciprocal error in this averaging is linear in the number of sample points \(N\). (Such scaling occurs for integration sampled in low discrepancy sequences [53]; our result could then be relevant to the efficient numerical integration of periodic functions). Therefore the error per sample point (i.e. per atom) scales as \(1/N\) and the total deviation from zero is independent of \(N\), excepting fluctuations of constant characteristic size [54]. This is, to our knowledge, the first time such behavior has been observed for physically relevant parameters: most previous examples of incommensuration either comprise two qualitatively distinct subsystems (such as electronic charge density waves in atomic lattices or a rare-gas monolayers on rigid substrates [55]) or have a much stronger interlayer interaction which locks in commensurate domains with strongly hindered motion of the misfit dislocations under interlayer sliding. The 0.0 – 0.3 eV corrugation arising from this sample-point analysis also appears in periodic cases, so long as one looks within a single unit cell: i.e. planar graphite [41] and \((5,5)/(10,10)\) two-walled tubes [45] also exhibit corrugations of this magnitude per unit cell.

The calculations so far have assumed perfectly rigid layers. However, real incommensurate systems distort perpendicular to the interface so as to improve the registry within finite domains separated by domain walls. (Commensurate systems also relax, but are not prone to this domain formation). The energetic cost of the in-plane distortion is second order, while an in-plane coordinate \(Q\) describing domain formation gives a linear energy gain due to enhanced registry: \(E(Q) = -AQ + BQ^2\). \(B\) measures the in-plane elastic response and \(A\) is an interlayer shear energy divided by a characteristic length, such as the interatomic distance \(a\). \(A/Ba\) governs the propensity towards the incommensurate-commensurate transition [52]. Whereas for typical atomic interfaces \(A/Ba \sim 1\) (i.e. bonding is roughly isotropic), in graphite the in-plane elastic stiffness greatly exceeds the interplanar shear strength: \(A/Ba \sim 10^{-2} \) to \(10^{-3}\). Not only is domain formation
strongly suppressed, but the noncovalent nature of the interlayer interaction suggests that the kinetic barriers against domain-wall motion are small.

We introduce the in-plane degrees of freedom through a Tersoff-Brenner [56] potential. Since the in-plane coordinates can now relax, we must impose boundary conditions on the layers to control the shifts in registry. We adopt simple boundary conditions similar to those which can be imposed experimentally: we fix one end of each layer and allow the remaining atoms to relax. To eliminate edge relaxations, we lock the last few rings of atoms into a collar with a single axial mechanical degree of freedom. In this geometry, the nongeneric effects of the strain boundary conditions cannot be completely eliminated, so we extract the size-dependent corrugation by comparing tubes of different lengths.

For sizes studied (up to a few hundred Angstroms long), relaxation can increase the corrugation by roughly 50 to 100 percent. The barrier of the (7,7)/(12,12) system increases from 20 meV to 35 meV per Å of length. (Other short-cell commensurate systems might show somewhat different relaxation behavior due to a different local minimum in the corrugation across one unit cell). The (14,0)/(16,10) incommensurate tube shows a variable increase in corrugation (depending on system size) but not above a factor of two. The corrugation remains small in the (11,2)/(12,12) tube at integral unit cell lengths, excepting a length-independent offset of about 0.1 eV arising from the strain boundary conditions. These results suggest that relaxation induces only moderate changes in corrugation for defect-free systems in the size range studied. As discussed above, relaxation effects are suppressed by the extreme rigidity of the graphitic layers and the weakness of the interlayer interaction.

The corrugation studied here comes from the bulk interaction of defect-free nanotube walls at a constant interwall contact area. In all these systems, the circumferential direction is essentially incommensurate (even for multilayer tubes with a rotational symmetry, although there the total corrugation is likely somewhat larger). Therefore the shear strength does not scale with diameter, at least for defect-free tubes of circular cross-section. Although a shear modulus cannot be defined in terms of a total contact area, a shear strength proportional to axial length is well-defined for axially commensurate tubes. In such cases, our results imply a relatively large shear strength for defect-free uniform-wrapping-angle commensurate systems. Taking the maximum slope of all lines connecting successive maxima and minima in the registry-dependent interlayer binding energy as a crude measure of the threshold force required to shear a defect-free system, we obtain shear strengths of about a hundred MPa [57] for these commensurate car-
bon nanotubes, and extremely small axial shear strengths for the very low-corrugation nonuniform wrapping angle commensurate nanotubes. In incommensurate defect-free systems, even this “axial” shear strength is ill-defined, since the corrugation does not scale with length. Two additional contributions to the interlayer interaction scale linearly with tube diameter, but are essentially independent of length: namely, the capillary force due to any changes in contact area and edge effects due to the terminal groups on tube ends.

Structural defects can play a very important role in the response under interlayer sliding. The commensurate nanotubes with uniform wrapping angle have the expected behavior under shear: disruptions in translational symmetry due to e.g. on-wall defects should typically lower the corrugation below that expected for a perfect crystal. However, for incommensurate layers, the defects play the opposite role: the perfect system has an extremely low corrugation, and the introduction of a small concentration of defects increases the shear strength, through a contribution that scales linearly with length at constant defect density.

Now we can compare our results to experimental values found in the challenging experiments by A. Zettl and R. Ruoff [20, 44], which were described in Section 3.1.1. For two large-diameter multiwalled tubes (with diameters of 300 and 360 Å) R. Ruoff et al estimated the shear moduli to be 0.3 and 0.08 MPa respectively [44]. These values, intermediate between the ultra-smooth incommensurate systems and the highly corrugated uniform-wrapping commensurate systems studied here, indicate that defects and/or deviations from circularity likely dominate in these particular very large-diameter multiwalled tubes. This result raises an interesting question on the prospects for obtaining more ideal behavior in samples with lower defect concentrations. Nanoscale systems can often grow defect-free over significant distances. For example, the straight growth of single-walled nanotubes over long (∼μm) distances suggests a low density of isolated pentagonal or heptagonal ring defects. Hindrances due to scrolling of layers [58] could be avoided since the occasional two-walled tubes formed during single-walled tube synthesis likely lack scrolling. Radial distortions which polygonalize the cross-section constitute a circumferential domain formation (with a much softer coordinate than a purely in-plane distortion) which could also hinder interlayer sliding. However, such distortions are most favored for layers with equal wrapping angles, while electron diffraction reveals multiple wrapping angles within single multiwalled tubes [59]. In addition, such distortions impose a cost in curvature energy and therefore are favorable only for tubes with large radii. (Intermittent nongraphitic interwall spacings are more naturally attributed to scrolling
or nonconcentric layers than faceting [60]). Finally, hindrances due to end-of-tube effects are amenable to chemical modification. These systems could asymptote to nearly perfect smoothness as structural perfection improves.

The results of this work were published in 2000 [37]. Since then we have tested the potential used in these simulations (see Section 2.4.1) and concluded that it tends to overestimate corrugation to sliding in DWCNs, as discussed in Chapter 2. According to our latest LDA results the corrugation to sliding in the angular-incommensurate (7,7)/(12,12) DWCN could be much smaller (see Table 2.1). However, we expect the same qualitative growth of the corrugation with the system size for other systems with short axial periods, such as (5,5)/(10,10) and (9,0)/(18,0). These tubes also have short angular periods and according to the LDA results presented in Table 2.1, they have relatively large rotational (the (5,5)/(10,10) tube) and translational (the (9,0)/(18,0) tube) barriers per unit cell. It is likely that RDP0 also overestimates the corrugation to sliding in the long-period and incommensurate DWCN, therefore the intertube sliding in these systems could be even smoother than our original estimates suggest.

### 3.1.3 Intertube motion: recent developments and potential applications

The possibility of the very smooth interlayer motion in MWCNs theoretically predicted by M. Damnjanovic et al [43], experimentally observed by R. Ruoff et al [44] and A. Zettl et al [20], and numerically simulated by us [37] has brought a lot of interest to these systems in the materials science community. In a number of publications that followed, the intertube motion has been studied in more detail and possible applications have been proposed and, moreover, realized.

T. Vukovic et al. [50] published a systematic study of the intertube motion in MWCNs based on the symmetry of the system. They expanded the interaction potential in a set of harmonics invariant under symmetry operations of the combined system. Then the value of the corrugation depends on the lowest common harmonics allowed by the symmetry of the individual tubes. They conclude that there are only a small number of systems where the common harmonics are small enough to result in any significant corrugation to intertube motion. Most of these cases are zigzag-zigzag and armchair-armchair DWCNs. But even for these systems the symmetry-breaking factor is often too large to permit any noticeable barrier to intertube motion, which is the case in our (7,7)/(12,12) example.

Potential applications of MWCNs as nanoscale devices include a gigahertz oscillator,
proposed by Q. Zheng and Q. Jiang [61]. They speculated that if both ends of the outer shells of a MWCN are opened up, the core shells will not stop once they extrude and retract inside the outer shells but will perform oscillatory motion. They estimate that the frequency of such oscillations could be as high as a few gigahertz, which is much higher than frequencies achieved to date. In order to realize this device, though, a more detailed study of dissipation processes is required [62]. Another interesting nanomachine, proposed by Yu. Lozovik et al. [21], is based on a thread-like relative motion of nanotube walls. If two neighboring shells have a suitable helical symmetry the sliding motion can be converted to the rotational motion around the axis and vice versa. The function of a MWCN as a bolt-and-nut pair depends critically on the depth of the grooves in the potential surface of a chiral nanotube. As was discussed in Chapter 2, available theoretical methods are unable to make a realistic estimate of such subtle variations in interlayer energy at this point.

Finally, the original idea of using MWCNs as nanoscale bearing has been successfully realized [63]. Zettl et al. built a nanomotor with a rotatable metal plate attached to a MWCN, which serves as the key bearing-like element. They estimate the diameter of the MWCN to be in the range from 10 to 40 nm. For a 10-nm-diameter MWCN with an effective length of 2 \( \mu \)m the estimate of the shear modulus is 100 to 300 GPa. Unlike existing bio-motors this device is designed to operate over a wide range of frequencies, temperature, and environmental conditions. The authors operated the actuator for thousands of cycles, with no apparent wear or degradation in performance. This makes the device a promising element of electro-mechanical, optical, and bio-mechanical circuits.

3.2 Nanotube-substrate interactions

3.2.1 Observation of nanotube alignment on a graphitic substrate

There is another interesting system where the registry-dependence plays an important role in the interlayer interaction and leads to interesting alignment effects. It has been observed experimentally that nanotubes deposited on highly oriented pyrolytic graphite (HOPG) tend to align 60° to one another, presumably matching the underlying hexagonal lattice. Another set of experiments confirmed the correlation and the threefold symmetry of the preferred orientations of SWCN on a graphitic substrate [64].

The first numerical simulations of the tube/substrate alignment were done by Buldum and Lu ([65]). They considered long, perfectly rigid nanotubes in contact with a graphitic substrate. The interlayer interaction energy was calculated with a Lennard-Jones (LJ)
potential and minimized by adjustment of the nanotubes' position. In their model the bond length of a nanotube matched that of graphite, so for a number of orientations a commensuration between the nanotube and the graphitic lattice was expected. Sharp peaks in the binding energy repeating every 60° were indeed observed.

The angular selectivity, though, was rather small: the principal dips were just 0.5% of the average binding energy. One may wonder if such an energy gain could be big enough to force a nanotube into a particular commensurate orientation with the graphitic substrate and keep it there. Possible nanotube imperfections, edge effects, thermal fluctuations, and bond mismatch can hinder this alignment process. But as was discussed in Chapter 2 the LJ potential underestimates registration effects. In the present work we would like to refine these important preliminary theoretical results by addressing a number of important issues:

- simulate the interaction with a more accurate registry dependent potential.
- perform a more thorough minimization of the tube position.
- consider the realistic case, in which the bond lengths on the nanotube and the substrate are mismatched.

3.2.2 Numerical simulations of nanotube alignment

Our setup is shown in Fig. 3.3. We first consider a rigid nanotube on a flat graphitic surface. The tube center of mass has three translational degrees of freedom. The tube can also spin about its axis. We assume that the tube axis is parallel to the substrate, which eliminates one rotational degree of freedom. The remaining angular degree of freedom, namely the angle about an axis perpendicular to the substrate, is the main variable of interest. We hold this angle fixed at a succession of values (every 0.1°) while optimizing the other parameters $x$, $y$, $z$, and $\phi$.

For a rigid fully incommensurate system, shifts of the center of mass parallel to the plane of the substrate do not change the energy; otherwise a careful minimization is required. To reduce finite-size fluctuations, the tubes are very long (about 300 nm). We treat the substrate as a single rigid graphene layer. The deeper layers are too far away to substantially affect the registry dependence. However, by pulling the nanotube closer to the surface they could indirectly increase most of the corrugation effects described here by about 10%. Subtle long-ranged effects (e.g., due to scattering by the tube of graphene electrons in low energy states) also might affect the orientation dependence.
Figure 3.3. The position of the nanotube above the graphitic substrate is defined by two angles $\phi$ and $\Theta$ and three coordinates of the nanotube center of mass $x$, $y$, and $z$.

of the tube-substrate binding energy. Such effects, which are not treated in the short-ranged potential used here, might be suppressed by using a substrate such as hexagonal B-N.

We begin with the case where all nanotube C-C bonds have the same length as in bulk graphite; this provides a helpful starting point for the later inclusion of more subtle effects. Fig. 3.4 shows the angular dependence of the tube-substrate interaction energy for three nanotubes with similar diameters but different helical angles. As expected, all tubes have a sharp minimum precisely at $\theta_0$, repeating every 60° [65]. The gain in the binding energy at the preferred orientation is substantial, about 8% of the total binding energy and tens of electron volts for tubes of a few hundred nanometers in length.

The curve between the principal minima is not quite flat. The additional dips do not shrink as the tubes are extended, so they are not finite-size fluctuations. Some of the features are known accidental commensurations for graphite [48]. For example, two graphitic layers become commensurate at 38.2° ($\cos^{-1}(11/14)$) (see Fig. 2.7). The amplitude of these peaks depends not only on the period of the structure at these angles, but also on the chirality of the tube. Here, by “chirality” we mean not the specific $(n,m)$ values, as the term is sometimes used for nanotubes, but the true left/right distinction. For example, the 38.2° peak for the (8,2) tube in Fig. 3.4 is missing its chiral companion at 21.8°, due to the left/right asymmetry of the whole system. These accidental commensurations are familiar from thin film growth on mismatched substrates.

However, the highly symmetric (10,0) and (5,5) tubes have additional minima at incommensurate angles, e.g. 30°. These unexpected features can be explained by looking at patterns in the distribution of local interlayer registries. If we take the coordinates of
Figure 3.4. The nanotube-graphene interaction energy, measured in meV per atom in the tube, as a function of angle about an axis perpendicular to the substrate. The zero angle is shifted to match the helical angle $\theta_0$. Here we assume matching bond lengths in the nanotube and graphene substrate. The distributions of interlayer registries for the minima marked “[2a]” and “[2b]” (at 30.0 and 38.2 degrees respectively) are shown in Fig. 3.5.

The atoms in the tube which lie above each unit cell of the substrate and translate them back into a single substrate unit cell, we construct a convenient representation for the global distribution of local interlayer registries, as shown in Fig. 3.5. In commensurate systems, the points cluster into a finite set whose energetics can be optimized through center of mass motion of the nanotube relative to the substrate or spinning about the tube axis. In normal incommensurate systems, the points cover the entire unit cell uniformly, yielding a smooth interaction. However, the finite transverse extent of the nanotube allows a third possibility, which is seen here: the points spread uniformly, but only over a discrete set of parallel lines that slice through the substrate unit cell.

In this third case, even though the system is axially incommensurate, the energy can
Figure 3.5. Solid points and lines represent the distribution of nanotube atoms in the area of contact with the substrate, mapped back into a single substrate unit cell, which is depicted by open circles and dashed lines. (a) and (b) are for equal bond lengths in tube and substrate. In (c) and (d) the bonds on the tubes are extended by 0.35% along the axis and 0.45% perpendicular to it.

be lowered by spinning the tube about its axis and/or shifting its center of mass parallel to the substrate and perpendicular to the tube axis. The effect is largest for the high-symmetry zigzag nanotubes with short axial unit cells, at angles \( \theta = 30.0, 41.4^\circ \) (41.4 = \( \cos^{-1}(3/4) \)). This case is absent in normal two-dimensional thin-film growth, since it is a nanoscale finite-size effect. The system is part-way between the commensurate and incommensurate states, so we term it a “pseudo-commensuration”. The effect is purely geometric. It would appear also in a Lennard-Jones treatment or for a hexagonal non-graphitic (or doped graphite) substrate.

The separation of the primary minima for the (10,0) tube of Fig. 3.4 into a broad valley and a sharp spike results from similar physics. The sharp minimum occurs as the set of projected registries condenses from lines into isolated points. The broader valley follows the extension of these lines across the unit cell and their subsequent smearing into a uniform and fully incommensurate coverage. The angular widths \( \Delta \theta \) of the primary minima depend on the tube length as \( 1/\sqrt{L} \). For zigzag tubes \( \Delta \theta \approx 1/\sqrt{N} \), where \( N \) is the number of axial unit cells.

Careful ab initio studies [66] show that the carbon-carbon bonds lengthen slightly when graphene is rolled into a nanotube. For tubes of the diameter studied here, the
mismatch is about 0.40%, meaning that even for the perfect angular alignment at $\theta_0$, registry is destroyed over about 40 nm (barring axial relaxation of the tube, to be discussed later). Do any substantial commensurate valleys survive this bond-length mismatch in long tubes? To answer this question, we treat next the case of mismatched bonds. Bonds parallel to the tube axis ($a_\parallel$) are extended by 0.35%. Bonds perpendicular ($a_\perp$) are extended by 0.45% [66]. The other bonds follow a linear interpolation. Strikingly, rather strong valleys survive, as seen in Fig. 3.6. The primary peaks shrink by less than 20% (except for the armchair tube) compared to the ideal case of Fig. 3.4. Several new minima appear, none of which is a standard commensuration: for mismatched bonds, essentially all preferred angles result from pseudo-commensurate sampling. For example, in the (10,0) tube minima are now at 0.0, 0.116, and 4.8 degrees. The second angle solves the equation $\sqrt{3} = (a_\parallel \cos \theta - a_{\text{graphene}})/(a_\parallel \sin \theta)$, which
yields \( \theta = \sin^{-1}\left( (a_\parallel - a_{\text{graphene}}) / (\sqrt{3}a_\parallel) \right) \) for small \( \theta \)'s. The 4.8° valley occurs when the projected parallel bond in the nanotube matches the substrate: \( \theta = \cos^{-1}\left( a_{\text{graphene}} / a_\parallel \right) \).

For chiral tubes this happens when the angle \( \theta \) satisfies \( a_\parallel \cos(\theta + \theta_0) = a_{\text{graphene}} \cos(\theta_0) \).

Although the C-C bond is extremely stiff, deformations due to the tube-substrate interaction are possible. Thus, we investigate the role of tube deformations using the Tersoff-Brenner potential [67] to relax the tubes’ internal coordinates. Because these calculations are computationally intensive, they are performed for only a subset of the systems described above. The attraction between tube and substrate flattens the tube slightly to increase the contact area. However, even for a relatively large (10,10) tube, the tube-substrate binding energy increases by less than 10%, and the angular dependence of the binding energy shows no substantial changes. The tube also can relax along its axis, as in the Frenkel-Kontorowa model [68]. A sufficient axial compression could restore equal bond lengths in tube and substrate. Compressing an isolated small-diameter nanotube from equilibrium to a length corresponding to the C-C bond length of flat graphene costs about 0.3 meV per atom [66], which is comparable to the difference in the tube-substrate interaction energy for matched and unmatched bonds (0.1 – 0.2 meV per atom). Since these energies are comparable, the nanotubes could deform along their axis and lock into commensurations with the substrate. To study axial relaxation, we simulated a 10 nm long nanotube with free hydrogenated ends on a graphene sheet. The nanotube typically does not escape the local minimum, meaning that (1) a slightly compressed bond-matched tube at the main commensurate angle remains compressed due to interactions with the substrate [69], and (2) a bond-mismatched (i.e. equilibrium length) tube at a pseudo-commensurate angle retains mismatched bonds. The energies of states (1) and (2) for the same tube are nearly indistinguishable, except for the armchair tubes, where the commensurate state is more favorable.

### 3.2.3 Possible separation process of nanotubes based on their helicity

Fig. 3.7 summarizes the typical (i.e. plateau) and maximal binding energies for several tubes. The displacement between the maximal and typical binding curves, even in the realistic mismatched-bond-length case, suggests that an ensemble of tubes could be distinguished, or perhaps even separated, based on their angle-dependent interactions with a substrate of hexagonal symmetry [70]. Such a separation would have to satisfy at least three requirements. First, the nanotube free energy within an ambient fluid [71] must be comparable to the average tube-substrate interaction to facilitate equilibration. Second, the tube sample must have a fairly narrow diameter distribution [72, 73, 74, 75].
Third, a preferred orientation must be imposed on the nanotube population, for example by pre-aligning them through extensional fluid flow or external fields [76] or by patterning the substrate [77]. Fig. 3.7b demonstrates that e.g. (8, 4) tubes could be separated out if a graphite substrate is oriented correctly to a heterogeneous mixture of aligned tubes.

To summarize, we describe a novel registry-dependent effect in the interaction of a one-dimensional system on a two-dimensional substrate: certain tube/substrate orientations are preferred due to finite size effects, even though they are manifestly incommensurate. We also identify a possible mechanism to distinguish carbon nanotubes by helical angle, which could improve our ability to produce samples with precisely defined structure and properties.

This work was done in collaboration with Monika H. Schleier-Smith and James C. Ellenbogen (Nanosystems Group, The MITRE Corporation, McLean, VA). It was published in Physical Review Letters, 92, 085503 (2004).
Chapter 4

Doping of multiwalled carbon nanotubes

4.1 Introduction: band structure of nanotubes

In order to simulate charge transfer in a system one needs to build a good model of the systems’ band structure. Since extra charge changes the energy levels of the doped system a first-principle self-consistent calculation is usually required. At low doping levels though the relative position and shape of the bands do not change much, so a tight-binding rigid-band approach is a good starting approximation which allows to capture the essential physics of the charge transfer process.

To derive the band structure of a nanotube in the tight-binding approximation we start with the band structure of graphite. For a single, flat graphene sheet, symmetry forbids coupling of \( \pi \) bands to \( \sigma \) bands that are well below and well above the Fermi level.

Consideration of the first neighbor \( \pi \) orbital overlaps is enough to accurately describe most of physical properties of the two-dimensional structure. In this simple model the dispersion relation for graphite can be found analytically [78]:

\[
E_{2D}(k_x, k_y) = \pm \gamma_0 \left\{ 1 + 4 \cos \left( \frac{\sqrt{3} k_x a}{2} \right) \cos \left( \frac{k_y a}{2} \right) + 4 \cos^2 \left( \frac{k_y a}{2} \right) \right\}^{1/2}, \tag{4.1}
\]

where \( a = 1.42 \times \sqrt{3} \) Å is the lattice constant and \( \gamma_0 \) is nearest-neighbor transfer integral (the overlap integral is set to zero for simplicity).

When the sheet is wrapped into a cylinder we assume that in the first approximation the bands are unchanged from graphene and the only effect is a selection of states made
by the boundary conditions: the wave vector in the circumferential direction becomes quantized while the wave vector along the tube axis remains continuous. Therefore the nanotube energy bands will be a set of one-dimensional cross sections of the two-dimensional energy surface of graphite. If \( N = 2(n^2 + m^2 + nm)/d_R \) is the number of hexagons of the graphite honeycomb lattice that lie within the nanotube unit cell (\( d_R \) is the greatest common divisor of \( 2(n+m) \) and \( 2(m+n) \)) we have \( N \) pairs of one-dimensional dispersion relations for an \((n,m)\) nanotube:

\[
E_\mu(k) = E_{2D} \left( k \frac{K_2}{|K_2|} + \mu K_1 \right), \quad (\mu = 0, \cdots, N - 1, -\frac{\pi}{T} < k < \frac{\pi}{T}),
\]

(4.2)

where the magnitude of the translational vector \( T \) is \( |T| = a \sqrt{3(n^2 + m^2 + nm)/d_R} \).

The reciprocal lattice vectors are now

\[
K_1 = \frac{2\pi}{2a(n^2 + m^2 + nm)} \left( \sqrt{3(n + m), n - m} \right)
\]

(4.3)

\[
K_2 = \frac{2\pi d_R}{2a(n^2 + m^2 + nm)} \left( -\frac{(n - m)}{\sqrt{3}}, n + m \right)
\]

The vector \( K_1 \) defines the direction and spacing between the \( N \) cuts of the potential surface of graphite (see Fig. 4.1). The properties of a nanotube depend critically on whether one of the allowed lines of continuous vector \( k \) crosses the special point \( K \) (as was discussed earlier in Chapter 1 at this point there is a band crossing exactly at the Fermi level). If this happens, there will be a similar crossing of \( \pi - \pi^* \) bands at the Fermi level of the nanotube. However, this structure is one-dimensional and the density of states at the Fermi level is not zero as in graphene; therefore the nanotube will be truly metallic. If none of the cutting lines catches the \( K \) point there will be a finite energy gap between the valence and conduction bands and such nanotubes are semiconducting.

A general expression for the band energy of an \((n,m)\) nanotube can be obtained by plugging the reciprocal vectors \( K_1 \) and \( K_2 \) into (4.2). It is convenient to decompose the vector \( k \frac{K_2}{|K_2|} + \mu K_1 \) into \( k_x \) and \( k_y \):

\[
k_x = k \frac{m - n}{2\sqrt{n^2 + m^2 + nm}} + \mu \frac{\pi}{a} \frac{\sqrt{3(n + m)}}{n^2 + m^2 + nm}
\]

(4.4)

\[
k_y = k \frac{\sqrt{3(m + n)}}{2\sqrt{n^2 + m^2 + nm}} + \mu \frac{\pi}{a} \frac{n - m}{n^2 + m^2 + nm}
\]

In two special cases these vector components and the corresponding one-dimensional band energies reduce to a simple form. For an armchair \((n,n)\) tube the \( x \) and \( y \) projec-
Figure 4.1. Directions of the continuous wave vectors in nanotubes slicing through the Brillouin zone of graphene. Nanotubes are metallic if the allowed lines go through the K point of graphene, as shown on the left, and semiconducting otherwise, as shown on the right.

The superscript \( a \) refers to armchair and \( k \) is the one-dimensional vector in the direction of the vector \( \mathbf{K}_2 \). This direction corresponds to the \( \Gamma \) to \( K \) line in the two-dimensional Brillouin zone of graphite (see Fig. 1.3). The dispersion relations \( E^a_\mu(k) \) in this model for the \((5,5)\) armchair nanotube are shown in Fig. 4.2(a), where there are six dispersion relations for the occupied bands and an equal number for the empty bands. In each case, two bands are nondegenerate (thin lines) and four are doubly degenerate (thick lines), leading to 10 levels in each case, consistent with the 20 carbon atoms in the unit cell of the \((5,5)\) nanotube. For all armchair nanotubes, the energy bands show a large
Figure 4.2. One-dimensional energy dispersion relations for (a) armchair \((5,5)\) nanotube, (b) \((9,0)\) zigzag nanotube, and (c) \((10,0)\) zigzag nanotube.

degeneracy at the zone boundary, where \(ka = \pi\), so that (4.1) becomes

\[
E_{\mu}^a(\pi/a) = \pm \gamma_0
\]  

(4.7)

for the 2D graphene sheet, independent of zone folding and independent of \(n\). The valence and conduction bands in Fig. 4.2(a) for the armchair nanotube cross at \(k = \frac{2}{3}\pi/a\). The crossing takes place at the Fermi level and the energy bands are symmetric for \(\pm k\) values.

For an \((n,0)\) zigzag nanotube the \(k_x\) and \(k_y\) are

\[
k_x = -\frac{k}{2} + \mu \frac{\pi \sqrt{3}}{a} \frac{n}{n},
\]

(4.8)

\[
k_y = \frac{k\sqrt{3}}{2} + \mu \frac{\pi}{a} \frac{1}{n}
\]

and the corresponding band energies \(E_{\mu}^z(k)\) are

\[
E_{\mu}^z(k) = \pm \gamma_0 \left\{ 1 + 4 \cos \left( \frac{\mu \pi}{n} \right) \cos \left( \frac{\sqrt{3}ka}{2} \right) + 4 \cos^2 \left( \frac{\sqrt{3}ka}{2} \right) \right\}^{1/2},
\]

(4.9)
\(- \frac{\pi}{\sqrt{3}} < ka < \frac{\pi}{\sqrt{3}}\), \quad (\mu = 1, \cdots, 2n).

where superscript \(z\) denotes a zigzag tube. Dispersion relations \(E^z_{\mu}(k)\) for the metallic \((9,0)\) and semiconducting \((10,0)\) zigzag nanotubes are shown in Figs. 4.2(b) and (c), respectively.

In general, the band crossing at the Fermi level for an \((n,m)\) nanotube happens if the following condition is satisfied:

\[ n \equiv m \pmod{3} \quad (4.10) \]

Therefore, all armchair tubes and one third of the zigzag and chiral tubes are metallic.

With this knowledge about the band structure properties of nanotubes and the explicit expression for the band energies we can simulate charge transfer effects in doped carbon nanotubes.

### 4.2 Raman study of phonon frequency shifts of DWCN under doping

Raman spectroscopy is a powerful tool for studying vibrational properties of carbon nanotubes. Due to the strong electron-phonon coupling in carbon nanotubes one can also deduce valuable information about their electronic structure by examining phonon modes of the nanotube. Using Raman spectroscopy P. Eklund’s group at Pennsylvania State University produced interesting experimental data on phonon frequency shifts of DWCN under doping with bromine. I will first describe their experimental results and then introduce our theoretical model of doping effects.

It has been indicated previously that bromine can either chemically dope SWNTs and MWNTs [80] or form a C-Br complex [81]. Eklund’s group acquired \textit{in situ} Raman spectra of bundled SWNT and DWNT samples during vapor phase bromine doping. The SWNT comparison spectra came from the same starting material that was used to synthesize the DWNTs. The Raman spectra in Fig. 4.3 show the low frequency radial breathing modes (RBMs) (Fig. 4.3a) and the high frequency tangential (T) modes (Fig. 4.3b) for pristine and bromine-doped DWNT samples. Consistent with previous Raman results [82] and band structure calculations [83, 84], the incident photon (1064 nm) can resonate with \(E^{s}_{22}\) in the outer semiconducting tubes and \(E^{s}_{11}\) in the inner semiconducting tubes simultaneously, where \(E_{ii}\) is the energy difference between filled and
Figure 4.3. Unpolarized Raman scattering spectra (T=300K) of DWNTs using 1064 nm laser excitation [88]: (a) low frequency radial modes and (b) high frequency tangential modes. The thin solid line is the fit to the spectrum, with Lorentzian components represented by dashed lines.

empty van Hove singularities [85]. In the pristine DWNT sample, three low-frequency and three high-frequency RBMs are observed and assigned to the most dominant outer ($d \approx 1.45, 1.33$ and 1.26 nm) and inner shells ($d \approx 0.75, 0.70$ and 0.68 nm) using the expression $\omega_{RBM}(\text{cm}^{-1}) = 234/d \text{[cm}^{-1}]$ [82]. The proportionality constant (234 cm$^{-1}$ nm) is an average over several literature values [82]. Small changes in $\omega_{RBM}$ due to $sp^2/sp^3$
hybridization in the inner shell would induce small deviations from this $1/d$ form, without affecting the main conclusions. Although the scattering from the RBMs of the outer tube is extinguished by bromine doping, the signal from the inner tube is essentially unaffected. Note that the 260 cm$^{-1}$ peak, reported earlier in bromine saturation doped SWNT bundles [80], is not observed here.

In the present work, $\approx 100$ spectra were taken, one every few minutes during an in situ experiment. As the RBMs begin to upshift with bromine doping, the intensity drops quickly. The extinction of the outer tube radial breathing band may be due to the loss of Raman resonance via the depression of the Fermi level by the bromine, which depopulates the initial states of the $E_{22}^s$ transitions. A Lorentzian line shape analysis of the 1064 nm T-band spectra (Fig. 4.3b) also suggests that most of the charge is transferred from the bromine shell to the outer layer of the DWNT. For convenience, a single Lorentzian represents the three unresolved tangential modes in the strongly curved inner tube (1581 cm$^{-1}$) [86] and the outer tube (1590 cm$^{-1}$). After bromine doping (Fig. 4.3), the inner tube T-band remains at 1581 cm$^{-1}$, whereas the outer tube T-band vanishes (consistent with the results for the SWNT comparison sample).

The effects of bromine doping on the Raman T-bands of the inner and outer CNTs can also be observed with 514.5 nm excitation (Fig. 4.4). The RBM region is omitted here, because the stretching frequency of bromine vapor at $\approx 320$ cm$^{-1}$ overlaps with the RBM frequencies of the inner tubes. The upper (lower) pair of spectra shows the SWNT (DWNT) results before and after doping. The details of the SWNT T-band are better resolved with 514.5 nm excitation, fitted by a sum of four Lorentzians. The T-band of the doped SWNT sample is well fit by a simple 18 cm$^{-1}$ rigid upshift of the undoped spectrum. The DWNT T-band can be fitted similarly: the pristine DWNT spectrum is fit by a relatively weak, unresolved T-band Lorentzian at 1580 cm$^{-1}$ for the inner tube, plus the same four Lorentzians used to fit the 514.5 nm SWNT spectrum. After Br doping, the T-bands shift and become very weak (note the 40x scaling of the signal). To fit the doped DWNT T-band spectrum, consistent with the analysis of the 1064 nm results, the 1580 cm$^{-1}$ Lorentzian band for the inner tube is superimposed with an outer tube T-band represented by the four-line fit to the undoped SWNT spectrum, but rigidly upshifted by 16 cm$^{-1}$. Freezing the liquid bromine in the attached bulb with liquid $N_2$ only partially reverses the doping. Consistent with previous Raman scattering [80] and resistivity results [87] on bromine-doped SWNTs, the T-band downshifts to 1603 cm$^{-1}$ (compared to 1592 cm$^{-1}$ in pristine material).
Figure 4.4. Unpolarized Raman scattering spectra (T=300K) of DWNTs and SWNTs using 514.5 nm laser excitation [88]. The upper (lower) pair of spectra are for the SWNT control (DWNT) sample. The thin solid line is the fit to the spectrum. Lorentzian components for the outer (inner) tubes are represented by dashed thin (dotted thick) lines.
4.3 Theoretical model of charge transfer in DWCN

We model the system as a three-layer cylindrical capacitor with the bromine anions forming a shell around the outer nanotube. The total energy contains three terms: the band structure energy of the inner and outer tubes and the electrostatic energy $E^{es}$ of the tri-layer charge distribution (see Fig. 4.5):

$$E = \sum_{i,k} E^{inner}_i(k) + \sum_{i,k} E^{outer}_i(k) + E^{es}. \quad (4.11)$$

The indices $i$ and $k$ label the occupied bands and wavevectors for the inner or outer tubes. For simplicity, we use a one-orbital tight-binding model with rigid bands and a nearest-neighbor $\pi$ overlap $\gamma_0 = 2.90$ eV [83]. This model ignores the small $1/R^2$ curvature-induced bandgap of $n \neq m$ tubes and non-rigid band effects. However, the model captures the primary physical effects relevant here, namely: (1) the distinction between metallic and large-bandgap semiconducting tubes with radius dependent bandgaps and (2) the cylindrical electrostatics. The smallest-diameter tube considered has $d = 0.63$ nm, approaching the size where re-hybridization becomes relevant [89]. Sub-band filling is relatively modest, so a rigid band approximation is a reasonable starting point for the essential physics.

We assume that the excess charge on each shell is uniformly distributed in an infinitely thin wall at the nuclear radius of that shell. The resulting electrostatic energy of the

---

Figure 4.5. Schematic representation of the double-walled nanotube as a cylindrical capacitor. The potential difference between the tubes depends only on the charge on the inner tube.
The triple-walled capacitor is

\[ E^{es} = \frac{1}{2} \frac{e^2 L}{2 \pi \epsilon_0} n_{inner}^2 \ln \left( \frac{R_{outer}}{R_{inner}} \right) + \frac{1}{2} \frac{e^2 L}{2 \pi \epsilon_0} n_{Br}^2 \ln \left( \frac{R_{Br}}{R_{outer}} \right), \]

where \( e \) is the electron charge; \( \epsilon_0 \) is the permittivity of free space; \( L \) is the unit cell length of the outer tube; \( n_{inner}(n) \) is the linear density of excess holes for inner tube (both tubes). \( R_{inner} \), \( R_{outer} \) and \( R_{Br} \) are the respective radii. We fix the charge per unit length in the Br shell (i.e., a theoretical dopant coverage with an attendant average charge per Br atom), and we minimize the total energy with respect to \( n_{inner} \).

We consider DWNTs with inner and outer diameters in the ranges \((0.63 \text{ nm} < d < 0.79 \text{ nm})\) and \((1.3 \text{ nm} < d < 1.5 \text{ nm})\). The resulting wall indices are \((5,5)@(10,10), (5,5)@(13,7), (8,2)@(14,6), (6,4)@(10,10), (6,4)@(11,9), (8,0)@(12,7), (7,3)@(12,8), (7,2)@(13,6), (8,1)@(16,2), and (9,2)@(11,11)\), where the metallic tubes are in boldface. Two physical effects act in concert to segregate most of the holes onto the outer nanotube. First, the bandgaps of the smaller-diameter tubes tend to be larger, so they empty last. Second, the cylindrical geometry preferentially raises the electrostatic potential at the inner tube (see Fig. 4.6). Only the charge on the inner tube affects the potential difference between the two layers (see Fig. 4.6).

Fig. 4.7 shows the results of the energy minimization. Two families of curves for \( n_{inner} \) vs \( n \) are obtained, one when the inner tube is semiconducting, and the other when it is metallic. The vertical scale \( n_{inner} \) is ten times smaller than the horizontal scale \( n \); most of the positive charge appears on the outer tube, in agreement with the experimental results. The lower family of curves is associated with two nested semiconducting tubes, except for \((6,4)@(10,10)\) and \((9,2)@(11,11)\), which have a metallic outer tube. In this
family, the outer tubes alone are doped until the Fermi energy $E_F$ cuts the first valence band in the inner tube and then a cusp appears. For the $(6,4)@(10,10)$ tube, the first cusp is at $n \approx 0.14$ holes/Å where $E_F$ first cuts the valence band of the $(6,4)$ layer, and the next cusp occurs at $n \approx 0.21$ holes/Å when $E_F$ reaches the second valence band of the outer $(10,10)$ layer.

Within the second (upper) family of curves, all of the inner tubes are metallic. The $(5,5)@(10,10)$ and $(5,5)@(13,7)$ are metallic pairs, so both tubes dope simultaneously. For $(8,2)@(14,6)$, the outer tube is semiconducting. Even though the inner tube dopes first in this case, the charge distribution rapidly begins to favor the outer tube as electrostatic effects begin to dominate. For the metallic $(5,5)@(10,10)$ and $(5,5)@(13,7)$ DWNTs, a cusp appears as $E_F$ approaches the second valence band of the outer tube at $n \approx 0.22$ holes/Å while for $(8,2)@(14,6)$ the cusp occurs at a smaller value of $n \approx 0.0126$ holes/Å when $E_F$ cuts the first valence band of the outer tube. Although some of these tubes are really small-gap semiconductors due to curvature-induced hybridization, the electrostatic effects dominate and therefore the small curvature-induced gap would have only a small effect on the overall charging behavior.

The charge transfer between the dopant (bromine) and the DWNT may be defined as the charge per host carbon atom transferred to the dopant, denoted as $f$. We can estimate $f$ using the relation $\Delta \omega (cm^{-1}) \approx 460 \times f$ obtained previously for the upshift of the Raman-active $E_{2g}$ modes ($\approx 1582$ cm$^{-1}$) during electrochemical charging of graphite-$H_2SO_4$ [80]. For the saturation doped DWNT-bromine system, we observe a $\approx 16$ cm$^{-1}$ upshift of the T-band in the outer tube relative to the undoped system (Fig. 4.4). Consistent with our tight binding model, we assume that about 90% of these holes reside on the outer tube. Therefore, $f \approx 1/(29 \times 0.9) \approx 1/26$; saturation bromine doping creates one hole in the DWNT per 26 carbon atoms, or 0.6 holes/Å. $E_F$ is then depressed by 1.2 eV to 1.4 eV (depending on $(n,m)$), uncovering states at the top of the third valence band of the semiconducting outer tubes. As a result, resonant Raman scattering from these outer tubes using 514.5 nm (or 2.41 eV) excitation should be dramatically reduced, in agreement with experiment. Theory (i.e., less than 10% of the charge on the inner tube) predicts a 1-2 cm$^{-1}$ upshift of the inner tube T-band, near the limit of our experimental resolution.

Similar issues arise in higher stage index ($n > 3$) graphite intercalation compounds (GICs), where the bounding carbon layers adjacent to the charged intercalate layers are thought to contain most of the compensating charge. For example, optical reflectance studies of potassium-GICs, showed that $\approx 90\%$ of the charge in the carbon layers was
Figure 4.7. Calculated number of holes per Å on the inner tube versus the total number of holes on the double-walled pair. Metallic tubes are in boldface. Alternation of grey and black curves is solely for clarity in distinguishing adjacent curves.

on the two bounding layers, and only 10 % on the interior layers [90]. This charge distribution is similar to that obtained here in a double-tube cylindrical geometry, even though the electronic band structure for cylindrical graphene is very different from that
of planar GICs. Recent results on DWNTs with BN outer walls and C inner walls [91] should exhibit a similar interplay of electrostatics and band filling, with an electrostatic favoritism to doping the outer layer.

The results of this joint study of multiwalled nanotube doping were published in 2003 [88].
Chapter 5

Simulation of Nanoporous Carbon

5.1 Synthesis and graphitization of NPC

As was discussed in Chapter 1, nanoporous carbons (NPC) are solid materials composed mostly of \( sp^2 \) hybridized carbons with pore sizes narrowly distributed around 5 Å [92], which makes them suitable for separation of small molecule mixtures. The structure of NPC can be very different, depending on the synthesis conditions. For NPC synthesized at 400°C the porosity is underdeveloped and the structure is highly disordered even on very short length scales, while at 1200°C a real nanostructure with measurable and controllable properties can be produced. However, the x-ray diffraction indicates that the overall structure is highly ordered only on a length scale of 15 Å, with carbon atoms forming aromatic nanodomains. These domains have local bonding similar to graphite, but lack the three dimensional structure associated with crystalline graphite [93].

A way to induce graphitization of already synthesized NPC is to expose the structure to alkali metals. It has been reported [94] that carbon nanotubes and polyhedra (Fig. 5.1b) can be formed spontaneously at temperatures as low as 50°C, in a transition from disordered (Fig. 5.1a) to ordered carbon through the interaction of elemental cesium with non-crystalline NPC. Similar structures, as well as graphitic like sheets, were observed in the samples treated at 350 (Fig. 5.1c) and 500°C (Fig. 5.1d). The authors explained this intriguing result as due to electron donation from the cesium to carbon nanodomains: the donated electrons may migrate through the NPC and act as a catalyst leaving on its place the ordered structures.

Our group (E.R. Margine, D. Stojkovic, myself, J. Sofo, and V. Crespi) has been investigating possible graphitization mechanisms theoretically. In the following sections I present a summary of this ongoing research.
5.2 Theoretical model of nanoporous carbon: the Wormhole

5.2.1 Introduction

Theoretical study of the NPC graphitization process is a challenging problem. On the one hand, a proper description of charge transfer in the system implies application of \textit{ab initio} methods. On the other hand, the simulation of disorder-to-order transitions is generally done on large systems, which are presently beyond the capabilities of first-

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{Figure5.1.png}
\caption{HREM images of (a) NPC prior to Cs treatment, but after heating in vacuum for 18 h at $T=450^\circ C$. (b) NPC after treatment with Cs and post-treatment at $T=50^\circ C$ [94]. (c) Post-treatment at $T=350^\circ C$. (d) Post-treatment at $T=500^\circ C$.}
\end{figure}
principles methods. Therefore the first task is to find a system that reflects the main structural properties of the NPC, yet is small enough to be tackled with first-principles methods. First I review existing models of NPC, then introduce our model of a nanopore, the *wormhole*, and describe a possible mechanism of its annihilation.

**Schwarzites**

One of the first realistic theoretical models of NPC was proposed by T. Lenosky *et al.* [95] in 1992. They constructed fully-connected three-dimensional forms of graphite called *schwarzites* after the mathematician H.A. Schwarz who first studied the periodic minimal surfaces these structures resemble [96]. T. Lenosky *et al.* proposed an algorithm for generating an $sp^2$ network of carbon on an arbitrary smooth surface with a given number of carbon rings. For example, they created carbon structures on $P$- and $D$-minimal surfaces. These types of three-dimensional surfaces are approximately represented by level sets of periodic functions constructed from low-order Fourier modes:

\[
\begin{align*}
    f_P(x, y, z) &= \cos(x) + \cos(y) + \cos(z), \\
    f_D(x, y, z) &= \sin(x)\sin(y)\sin(z) + \sin(x)\cos(y)\cos(z) \\
    &\quad + \cos(x)\sin(y)\cos(z) + \cos(x)\cos(y)\sin(z),
\end{align*}
\]

where $x, y,$ and $z$ range from 0 to $2\pi$. When the surfaces are tiled with carbon rings, each structure has 216 carbon atoms per unit cell with 80 hexagons and 24 heptagons [95]. The presence of seven-membered rings is not accidental. Much as pentagons make a graphitic surface close on itself and create closed structures with positive curvature (fullerenes), heptagons introduce negative curvature into the structure. The relationship between different numbers or rings is determined by the topology of the structure. If a closed polyhedral structure tiles a surface with $g$ holes (for example, torus has one hole), then, according to Euler’s theorem,

\[
V - E + F = 2 - 2g,
\]

where $V$, $E$, and $F$ are the number of vertices, edges and faces, respectively; $g$ is also known as the *genus* of the surface. Hexagons by themselves do not change the topology of the structure, since we have $6/3 = 2$ vertices and $6/2 = 3$ edges per face and $2 - 3 + 1 = 0$. We restrict our considerations to systems made out of five-, six- and seven-membered rings only. If a structure has $n_5$ pentagons and $n_7$ heptagons the genus can be expressed
as in [95]:

\[ n_5 - n_7 = 12 - 12g. \] (5.3)

For example, the genus of a fullerene is zero, which requires exactly 12 pentagons (if no heptagons are present). Nanotubes and graphite have genus one, and can therefore be made of hexagons exclusively. The \( P \) and \( D \) minimal surfaces have a hole in each dimension (\( g = 3 \)), therefore exactly 24 heptagons are needed to tile them. The number of hexagons can vary; by inserting extra hexagons one can relieve some of the curvature energy. According to T. Lenosky et al. [95] their 216-atom systems are energetically more favorable than the \( C_{60} \) fullerene (by 0.2 eV per atom). The holes in the structures can be considered as pores (their sizes range from 6 to 12 Å). Notice that the effective pore size is different from the geometrically intuitive carbon-to-carbon distance at the bottleneck of the hole. For such small distances one needs to take into account a finite extent of the electron clouds from carbon atoms, which reduces the effective diameter of the pore by about 3 Å. The considered \( P \) and \( D \) surfaces have pores large enough to be good models of the NPC. They have been recently used for calculation of electronic states in NPC [97]. Yet, these three dimensional connected surfaces are too constrained to give clues for possible graphitization mechanisms.

**Radiation-damaged graphite**

A different approach to modeling the graphitization in NPC can be based on structures whose geometries are already close to planar. Radiation-damaged graphite at low dosage of radiation and after subsequent annealing develops defects but retains its layered structure [98]; thus it can be considered a possible candidate. However, the energies involved in the irradiation process suggest that graphite undergoes dramatic structural changes (possibly, large-scale breakdown of the \( sp^2 \) network) which manifest themselves in changes of the physical properties of the system. First, the structure exhibits considerable expansion along the \( c \)-axis (for highly irradiated graphite it can reach 300\%) and contraction in the basal plane. Second, the conductivity normal to the basal plane increases, indicating that the structure becomes connected in three dimensions. It is not clear [98] how this rearrangement occurs. The bridging of the layers can be purely \( sp^2 \) or involve other mechanisms (\( sp^3 \), dangling bonds, etc.). One possibility is that layers break up into little patches under the bombardment of high-energy particles. There may be a lot of dangling bonds created in the system, especially at the edges of the flakes,
which recombine to form interlayer bridges. Telling et al. [99] proposed the formation of a point defect, which connects neighboring layers, leaving vacancies in the basal plane. The interstitial carbon atoms are then in a mixture of $sp^2$ and $sp^3$ states and the defect energy is quite high: about 5 eV. While NPC also has some fraction of $sp^3$ bonds, this local defect is too small in comparison to what we expect to see in NPC.

5.2.2 The wormhole

Our approach to build a model defect in an $sp^2$ carbon structure is somewhere between the two models described above. We also start with two graphitic layers, but we bridge them with a smoothly curving neck, not breaking the $sp^2$ network. Thus, we create a continuous structure with a hole in just one dimension, so it has genus two. According to Eq. 5.3 this structure should have at least 12 heptagons. A simple way to introduce them is shown in Fig. 5.2. Two hexagons in each layer are protruded out of the plane towards one another. Now we have six atoms with two dangling bonds sticking out of the layers. The problem is that these atoms cannot be connected directly without rearranging other bonds, but we can link them by inserting two additional sets of six carbons each in the middle (this middle part can be viewed as a portion of the (6,0) nanotube). The resulting neck, consisting of 6 hexagons and 12 heptagons, is rather narrow (less than 5 Å in diameter) and long (the interplanar distance is now 7.3 Å, more than double of that in graphite). This structure is not exactly what we were looking for. Even if we found a way to restore the original graphitic layers, 12 extra carbon atoms

Figure 5.2. A neck (132 atoms), connecting two graphitic layers, consists of 12 heptagons and 12 additional carbons inserted in the middle.
A way around it is not to introduce extra atoms in the system at all. We go back to the stage when six atoms are sticking out of each layer, but instead of inserting the additional rings we rotate the layers by 90 degrees around the vertical axis. Then the two sides lock as pieces of a jigsaw puzzle. The resulting pore, or what we call a wormhole, is shown in Fig. 5.3a. It is interesting to compare our structure to a fullerene (shown in Fig. 5.3b). We relax the structures with \textit{ab initio} methods (details are below) and compare dimensions. The two objects are similar in size: the diameter of the fullerene is about 7 Å, while the wormhole stands 6 Å high and measures 5 Å in diameter at the bottleneck. We also calculate the energies of the structures taking the energy of graphite as a reference point. The total energies turn out to be a little different: $E_{\text{fullerene}} = 23.4$ eV and $E_{\text{wormhole}} = 18.8$ eV. However, the excess energy in the wormhole is distributed mostly among the 48 atoms which make up the 12 heptagons, so the excess energies per atom for the wormhole and the fullerene are almost identical: 0.39 eV/atom. Finally, both structures have 12 non-hexagons, only they are five-membered rings in fullerenes and seven-membered rings in the wormhole. This makes the fullerene have a genus one less than graphite and the wormhole one greater. One can view these objects as topological opposites.

When we constructed the wormhole one problem was left unsettled: when we rotate the two layers to connect the dangling bonds we violate the boundary conditions of the periodic two-dimensional structure. Due to the incommensurability of the two lattice vectors in the $x$ and $y$ directions, the periodicity cannot be restored without enforced
matching of the two periods. Fortunately, we have the flexibility to add extra hexagons without changing the topology. We find that if each layer consists of 112 atoms the $x$ and $y$ dimensions are very close and by small distortion the unit cell can be made square. The deformation cost is estimated to be rather small: just 1 eV per unit cell (5 meV per atom). It is plausible that such domains (see Fig. 5.4) could be realized in NPC.

As discussed above, the effective diameter of the wormhole is smaller than 5 Å, due to the finite extent of the electron orbitals in carbon. To probe the effective size, we put different alkali atoms inside the wormhole. Lithium can move through the hole relatively easily, and it actually prefers to be in the middle. Potassium has a local minimum 2.5 Å away from the center of the wormhole and is prevented from passing through by a 1.5 eV barrier. Cesium is far too big to squeeze through the hole, and it prefers to be about 3 Å from its center. Nevertheless, this wormhole is a good starting point for investigation of the defect healing in NPC. One can study trends on this model and generalize them on larger structures.

5.3 Wormhole annihilation

5.3.1 Global reaction path

Now that we know the initial (wormhole) and final (two flat sheets) states, a wormhole annihilation mechanism needs to be found. Eventually, all the bonds between the heptagons should be broken and the atoms returned to their original sites in the graphitic layers (see Fig. 5.4). A lot of energy would be released in such a reaction: 19 eV from the wormhole and another 5 eV (224/2 atoms $\times$ 50 meV/atom [27]) from additional inter-layer binding if the sheets are allowed to settle at the optimal distance from each other (3.34 Å). However, bond breaking reactions have high activation barriers; a carefully constructed reaction path should be introduced.

An example of large-scale recombination of carbon bonds is the merging of fullerenes in a series of transformations [100]. The basic reaction step is a Stone-Wales (SW) transformation, originally observed in graphite. It involves rotation of one carbon bond by 90 degrees (see Fig. 5.9 below), which creates a pair of pentagons and a pair of heptagons instead of four hexagons (notice that according to Eq. 5.3 the genus is conserved). In this reaction bonds do break, and in the intermediate state the rotated atoms go partially in an $sp^3$ state, but at the end the $sp^2$ bonding is restored.

We use this bond rotation as a first step (see Fig. 5.5,1) to help bring atoms closer to the layers. The SW transformation in the wormhole transforms four heptagons into
Figure 5.4. A full square unit cell (224 atoms) containing a wormhole. The lattice vectors are 13.1 Å in each direction. The distance between the layers is 6 Å. The flat sheets can be reconstructed from the wormhole if the two sets of carbons (six green and six blue) retract back into the respective layers.

A pair of octagons and a pair of hexagons. As a result we end up with a weaker bond shared by two octagons and two atoms closer to the upper and lower sheets. We expect that it should then be easier to break the first bond in the central part of the wormhole (step two in Fig. 5.5). The resulting configuration will now have two dangling bonds, which is very costly energetically. One expects that further bond breaking would make the situation even worse. Fortunately, the next two broken bonds recombine with the currently dangling ones, so the next metastable state will still have two dangling bonds (Fig. 5.5,3). Thus, we can go around the neck, unzipping it (steps 3-6 in Fig. 5.5) and have no more than two dangling bonds at a time, until we are done. This important conclusion suggests that the barriers along this reaction path should stay relatively low.
Figure 5.5. A series of transformation leading to the destruction of the wormhole and restoration of the flat $sp^2$ network. The energetics of these transformations are shown in Fig. 5.6

First, we evaluate the energies of the relaxed metastable configurations, which we call step zero (the wormhole), step one (SW), step two (first bond breaking), and so
Figure 5.6. The energies of the wormhole annihilation steps calculated in LDA and tight-binding. The corresponding configurations before relaxation are shown in Fig. 5.5 (step zero is the wormhole, step six is graphite). The reference point is the energy of the wormhole. For the first three configurations the relaxed states are close to the starting structure. For step three the two methods find different metastable configurations: the one produced by the LDA does not have dangling bonds because some atoms are in the $sp^3$ configuration (see Fig. 5.7). For some configurations after step three the relaxation leaves the system in the graphitic state.

We use both tight-binding and ab initio methods in these simulations. For the first three steps the relative energies in both treatments match to within 0.1 eV; the total energy of the wormhole above graphite also agrees very well: $E_{\text{wormhole}}^{\text{tight-binding}} = 19.2$ eV, $E_{\text{wormhole}}^{\text{LDA}} = 18.8$ eV. As expected, the second step has the highest energy: about 2 eV above the wormhole level. After that the energy quickly drops and some of the steps shown in Fig. 5.5, constructed geometrically and pre-converged with Tersoff-Brenner potential [67], become unstable. For example, if step three is relaxed in LDA, the structure prefers to put two atoms in an $sp^3$ configuration to avoid having dangling bonds (see Fig. 5.7b). For this case, the tight-binding result is different: this method still leaves the system in the $sp^2$ state with dangling bonds. Step four is also described by the two methods differently: according to the LDA, this configuration is not stable and the system relaxes directly to graphite; the tight-binding finds a local minimum with $sp^3$-connected carbons. Finally, step five does not occur in either method: even if there
Figure 5.7. Step three relaxed with tight-binding (a) and LDA (b) methods. The structure on the right does not have dangling bonds; instead some atoms are in the $sp^3$ configuration.

are local minima nearby, they must be too shallow and the structure ends up as two flat sheets. This indicates, that once the structure goes over the hump during the first few steps, none of the remaining intermediate states in the zipper proposed represent even metastable states and it goes directly to the final configuration (graphite). Therefore we focus on the first two local transitions: the SW and the bond breaking transformations.

5.3.2 Local reaction paths

At this stage it is not enough to know the intermediate metastable states, we need to find a reaction path connecting them, or at least the maximum energy along it. The configuration corresponding to the maximum energy is a saddle point on the potential surface around the two states. There are a number of techniques that can find the saddle point(s) [101] as well as the full Minimum Energy Path (MEP) [102]. We use the Nudged Elastic Band (NEB) method [102] for our problem. The crux of this method is the creation of a chain of intermediate configurations, images, between the initial and final states and rearrangement of the positions of atoms in them till the chain converges to the MEP. In one of modifications of the NEB, the Climbing Image NEB, one image is treated differently: it still stays on the MEP, but it is forced to climb up the energy hill till it reaches a saddle point. While the determination of the MEP requires many images, the saddle point can be located accurately with a shorter chain.

Before we apply this method to the reaction paths in the wormhole we test it on a simpler system, graphite. Fig. 5.8 shows a path found with the NEB for the Stone-Wales transformation in graphite (32 atoms). In this example we use 25 images, but even 7 images are enough to calculate the barrier to within 0.1 eV. We use a fixed unit cell
Figure 5.8. The Stone-Wales transformations in the wormhole, graphite, and the bell (see Fig. 5.10). The initial, saddle point, and final states for graphite are shown in Fig. 5.9.

In these calculations; the adjustment of the lattice vectors is expected to change the results by no more than 0.2-0.3 eV. The final state and the barrier are very high: 6.6 and 9.7 eV above graphite, respectively, which is consistent with calculations of others [103]. The high energies might seem bad news for the wormhole annihilation path we proposed. However, the setup for the bond rotation in the wormhole and planar graphite are rather different. Indeed, the large barrier for graphite is probably caused by the unfavorable planar geometry of the structure: Fig. 5.9b shows that in the configuration, corresponding to the saddle point, the rotated bond sticks out of the plane to stay far away from the neighboring atoms. In the wormhole the local environment of a carbon

Figure 5.9. The Stone-Wales transformation in graphite. a) the initial state; b) saddle point; c) final state.
Figure 5.10. Two finite-size structures used, in particular, for studying the effects of the periodic boundary constraints. a) The bell (from [105]) consists of 34 carbons and 12 hydrogens; in one of the tests we rotate the bond, shown in red. b) A wormhole after the SW transformation and the first bond-breaking (the starting configuration is as in Fig. 5.3, only with hydrogen-terminated ends).

atom is curved, therefore the bonds are slightly weaker. In addition, there is more room for atoms to swing past their neighbors in the intermediate stages of the bond rotation. When we apply the NEB to find the barrier in the SW transformation in the wormhole we do obtain a much smaller value: 5.8 eV, which is 4 eV lower than that in graphite (see Fig. 5.8).

The significant decrease of the barrier in the SW transformation has been observed before in other non-planar structures. J. Yi *et al.* found a 5.4 eV value for the barrier for the fullerene [104]. Z. Slanina *et al.* [105] simulated bond rotation in another system with positive curvature. Their structure, the *bell*, has 34 carbons; 12 hydrogens are added for termination of the dangling bonds (see Fig. 5.10). The top part of this molecule, consisting of two adjacent hexagons and two neighboring pentagons, locally resembles the structure of the fullerene. However, the bell has a 90-degree symmetry axis and the rotation of the bond (shown in red) does not change the structure. The barrier 5.4 eV for the bell [105] was calculated in the B3LYP/6-31G* flavor [106] of DFT. We recalculated this value in the LDA and observed a similar value of 5.7 eV (see Fig. 5.8). Interestingly, the SW barrier is almost identical to that for the wormhole. This could be the result of similar radii of curvature near the point of the bond rotation.

The next stage, the bond-breaking transformation, is more straightforward. It involves an almost linear drift of one atom from the bottom part of the neck to the top one (see Fig. 5.5,1 and 5.5,2). The barrier is much lower than in the SW transformation:
just 3 eV.

Our latest tests reveal that the changes of the structural constraints in the system can further lower the energies of the metastable states and the reaction barriers along with them. Our wormhole lives in a square matrix, with the $x$ and $y$ dimensions not allowed to adjust. Moreover, buckling in the $z$ direction is also restrained by the in-plane periodic boundary conditions. To estimate the effects from the boundaries we consider a free-standing wormhole, shown in Fig. 5.10b, with hydrogen-terminated dangling bonds. We recalculate the energies of steps one and two relative to the energy of step zero. The additional flexibility of this system allows the wormhole to find lower saddle points and local minima. The energy of step one goes down by 1.5 eV and this state is actually more favorable (by 1.3 eV) than the starting symmetric configuration. The energy of step two also goes down by 1.5 eV and Fig. 5.10b shows why: a severe distortion makes one side of the wormhole (where the bonds have been broken) almost flat. This model is probably too soft, compared to the real situation in the NPC, but it reveals interesting trends.

Figure 5.11. Reaction paths found with the NEB for the first two transformations in the wormhole: the SW (from 0 to 1) and the bond breaking (from 1 to 2). These calculations are done for the wormhole in the square periodic unit cell; the pictures of the steps 0, 1, and 2 show only the central part of the wormhole for convenience.
For example, if some flattening in the structure happens on one end of the sample, the wave of graphitization, enhanced by the interlayer attraction, might help squash defects.

5.4 Summary and future study

We have constructed a fully-$sp^2$ model of NPC, the wormhole, small enough to be tackled with first-principle methods. We propose a possible path of the wormhole annihilation, which is broken down into well-defined stages. We calculate the barriers for the key steps (the SW transformation and the bond-breaking) and demonstrate that these barriers strongly depend on the structural constraints of the system.

In the next stage we plan to investigate the effects of doping on the stability of this defect. It has been shown [105] that the SW transformation in one carbon structure, the bell, is facilitated when the system is doped with an alkali atom. This might be caused by formation of an $sp^3$ bond in the saddle-point configuration with the help of the donated electron. However, our preliminary results show that this decrease in the barrier most likely results from the electrostatic interaction of the ionized alkali atom with the system. This effect depends on the particular placement of the dopant and is therefore sensitive to the choice of the system. So far we have not observed large variations in barriers for the SW or the bond-breaking transformations in the wormhole.
Chapter 6

Development of empirical potentials with Data Mining methods

6.1 Introduction

Development of empirical interatomic potentials follows two general trends. In the first approach a functional form of the potential is assumed based on the physical knowledge about the system. A good example is the family of powerful classical potentials developed by Tersoff and Brenner [67], and the environment-dependent interatomic potential (EDIP) [107]. The main advantage of this method is a small number of fitting parameters; on the downside, the limited set of functions often cannot fully reproduce the dynamical matrix. In the second approach the potential is constructed from a more complete set functions. This technique requires a large database to define all the parameters, which becomes problematic for large systems. As the speed and efficiency of first-principles calculations grow, data mining methods for ab initio data utilization attract more and more attention [108]. We have been investigating how Neural Networks (NN) could be used for this purpose and trying to work around the main problems that arise in this approach. One of the most important steps is a proper choice of a NN. In the following section I will discuss the design and function of main types of Neural Networks. Then I will formulate our algorithm on how to adapt the network for its application to solid states problems. Finally, I will present tests of our method on a real ab initio database generated for sp² carbon structures.
6.2 Review of Neural Network methods

A neural network is a powerful learning machine able to capture and represent complex input/output correlations. Much like a human brain, a NN consists of simple processing elements, neurons that are connected into a network by a set of (synaptic) weights. Among the attractive features of a NN are the parallel operation of neurons (input variables are processed simultaneously) and the capability to learn (adjust the weights) as they process information. To understand how a NN works we first look at the function of its basic unit, a neuron. Each element $j$ takes a number of input variables $x_p$, weights them ($w_{j,p}$), sums them up, and uses the result as the argument for a scalar-variable activation function $f_j$ to produce the output $h_j$. The following formula summarizes these steps:

$$ h_j = f_j \left( \sum_{p=1}^{D} w_{j,p} x_p + w_{j,0} \right) $$

(6.1)

A few examples of activation functions are

$$ f(x) = x $$

$$ f(x) = \tanh(x) $$

$$ f(x) = sgn(x). $$

(6.2)

The first activation function is linear; if a NN consisted of only linear transformations it would not be able to handle nonlinear mapping between input and output. Thus, for many problems nonlinear functions need to be applied; a popular choice is the hyperbolic tangent. NNs with such activation functions are used for control and multi-dimensional function approximation. The last function in Eq. (6.3), the step function, is of particular importance to networks for categorization, e.g., problems, that require a yes/no answer. The overall function of a NN is determined by the architecture of the network, the magnitude of the weights and the mode of neuron operation. Two of the most commonly used types of NNs are the Multilayer Perceptron (MLP) and the Radial Basis Functions (RBFs).

6.2.1 Multilayer Perceptron

The basic MLP is constructed by ordering neurons in layers, letting each neuron take input from previous layers or external inputs. The information in such a perceptron is
Figure 6.1. A two-layer perceptron. The first and second layers have three and two neurons respectively. Each layer has an additional bias input. Formatted, the two-layer network consists of five neurons, three input variables, and two outputs. Notice, that each layer has an additional input, bias, which serves as a constant offset. In a decision-making network bias shifts the centerline where the threshold lies; in a multi-dimensional function mapping network it defines the average value of the function.

The full expression for an output of a two-layer network is

\[
y_m(x) = F_m \left[ \sum_{j=1}^{n_2} W_{m,j} f_j \left( \sum_{p=1}^{n_1} W_{j,p} x_p + w_{j,0} \right) + w_{i,0} \right].
\]  

This mathematical expression is not easy to interpret. To help visualize the function of a MLP let’s examine, step by step, how the three-layer perceptron in Fig. 6.2 maps two-dimensional space \((x_1, x_2)\) onto a scalar field \(y(x_1, x_2)\). In this example the network has eight neurons arranged in three layers; it takes two input variables, and produces one output value. The first two layers have a hyperbolic tangent as the activation function; the last layer performs a linear transformation. The left pair of neurons in the first layer creates a ridge along the \(x_1\) direction; the right pair creates a similar ridge along \(x_2\) (see...
Figure 6.2. An example of a three-layer perceptron with two inputs and one output, which creates a localized peak at \((0,0)\). The first and second layers have the hyperbolic tangent as an activation function; the last layer performs a linear transformation. The resulting weights are shown below (the units in each layer are ordered from left to right except that biases are labeled as zero as in Fig. 6.1; first index denotes the unit in a lower layer). First layer: \(w_{01} = 1, w_{02} = 1, w_{03} = 1, w_{04} = 1, w_{11} = 1, w_{12} = -1, w_{13} = 0, w_{14} = 0, w_{21} = 0, w_{22} = 0, w_{23} = 1, w_{24} = -1\); second layer: \(w_{01} = 0, w_{03} = 0, w_{04} = 0, w_{11} = 0.5, w_{12} = 0.5, w_{13} = 0, w_{14} = 0, w_{21} = 0.5, w_{22} = 0.5, w_{23} = 0, w_{31} = 0.5, w_{32} = 0, w_{33} = 0.5, w_{34} = 0.5, w_{41} = 0.5, w_{42} = 0, w_{43} = 0.5\); third layer: \(w_{01} = 0, w_{11} = -10, w_{21} = 10, w_{31} = 10\). The intermediate steps and the resulting peak are shown in Fig. 6.3.

Fig. 6.3a,b). When we combine the results we obtain a localized bump in the middle (Fig. 6.3c), which is still accompanied by the secondary ridges. To suppress them we employ the second layer: the first neuron acts on the whole output of the first layer, while the other two neurons act on individual ridges only. When we finally combine all outputs in the top layer the result is a nice localized peak, shown in Fig. 6.3d. If we add more neurons we can place more peaks in the region \((x_1, x_2)\) of interest. The important feature of the MLP is that the position and the shape of these bumps are controlled by the weights and can be adjusted as needed. Thus, only a small number of parameters
Figure 6.3. An example of a three-layer perceptron with two inputs and one output, which creates a localized peak at \((0,0)\). First, two ridges are created along \(x_1\) (a) and \(x_2\) (b). When combined, they form a peak in the middle (c), while the secondary ridges are still present. The last two layers of the MLP suppress the ridges leaving the localized peak only.

might be required to achieve a good fit to a nonlinear function. Intuitively, it is clear that with enough peaks one can approximate any smooth function \(f(x,y)\) to any degree of accuracy. This result follows from a general theorem on the representation of continuous multivariable functions by the superposition of functions of one variable, proved by A. N. Kolmogorov in 1957 [109]. The mapping of any multidimensional function to any degree of accuracy can also be done with a two-layer perceptron [110] and they are a NN of choice for various problems.

Now comes the important question of how to adjust the parameters of the MLP to tune it to a specific problem. Usually one wants to minimize the error function (from now on I will focus on networks with just one output; generalization on multiple-output networks is straightforward):

\[
Error = \frac{1}{2} \sum_{i=1}^{N} \|y_i - t_i\|^2
\]

where \(N\) is the number of data available, and \(y_i\) and \(t_i\) are the \(i\)th NN output and target value from the database respectively.

As in most minimization problems, one needs to know partial derivatives of the minimized function with respect to all of the parameters in the system. An efficient
algorithmic formulation for computing error derivatives with respect to the weights of a NN was proposed by Rumelhart and McClelland in 1986 [111]. The algorithm is called backpropagation, reflecting the order of the derivative calculation: from the output layer down to the first one. With the partial derivatives known one can apply different minimization techniques, from the simple gradient descent to the advanced Lenenberg-Marquardt algorithms [112]. Due to the complicated issues related to the minimization of a nonlinear function a careful choice of the minimization technique is required. Still, no method guarantees finding the global minimum of a nonlinear system and very often one gets stuck in local minima, which results in a suboptimal performance of the NN. To avoid this problem, an alternative way to do the nonlinear multidimensional mapping can be applied, which is based on the use of Radial Basis Functions.

6.2.2 Radial Basis Functions

The analysis of the MLP function in the previous section showed that most of the NN weights are used to construct localized peaks and only at the last stage are their heights adjusted to fit to a particular multidimensional surface. As a shortcut, we can introduce a special neuron, which by itself represents a localized bump at some point of the $D$-dimensional configuration space. A convenient choice of the localized function is a Gaussian

$$G(||r - \mu_j||) = e^{-\frac{||r - \mu_j||^2}{2\delta_j^2}},$$

(6.5)

where $||r - \mu_j||^2$ is the Euclidean distance to the center of the Gaussian function in the $D$-dimensional space. Then the output of such a network is a linear combination of localized peaks:

$$y(r) = \sum_{j=2}^{M} C_j G(||r - \mu_j||) + C_1 = \sum_{j=2}^{M} C_j e^{-\frac{||r - \mu_j||^2}{2\delta_j^2}} + C_1,$$

(6.6)

where we again add a bias parameter $C_1$. The positions of the Gaussians ($\mu_j$) and their widths ($\delta_j$) are usually determined independently from the set of parameters $C_j$. They can be tuned by so called unsupervised training [112], where only the input vectors are analyzed. In this approach, RBF centers are placed in regions where most of the data points cluster and the widths are set according to the distances between the clusters. A random choice of the Gaussian parameters also works well for many problems. After
these parameters are chosen in one way or another, training of the RBF network becomes a linear problem. Advanced numerical methods of linear algebra (see below) provide a fast and robust solution for the weights $C_j$. This is the main advantage of the RBF network over the MLP.

Of course, some of the flexibility of a network is sacrificed when its nonlinear activation functions are not allowed to adjust. Fig. 6.4 demonstrates limitations of a network, which consists of spherically symmetric radial functions. In this example, the two-dimensional surface is a wave, propagating in one direction and only slightly modulated in the other (the little ripples represent noise of input data). The surface is approximated by 40 Gaussians, randomly scattered in the region of interest. Notice, that the quality of the fit critically depends on the ratio of the widths of the functions to the average distance between them: $\alpha = \delta/\langle d \rangle$. If the parameter $\alpha$ is too small, as in Fig. 6.4b, there is little overlap between the Gaussians and the mapping is too lumpy. For a large value of $\alpha$, as in Fig. 6.4c, the slowly varying radial functions are too smooth to describe the rapid variations of the wave. An intermediate value of $\alpha$ (Fig. 6.4d) provides a better surface approximation. Yet the architecture of the network is too rigid to adapt to the specific property of the function, namely, the much stronger dependence on $x_1$ than on $x_2$. A MLP would eventually learn that one of the variables is not very important, while the RBF network treats both directions on equal footing. As a result, many more centers are required to approximate this surface. This is a general trend for RBF networks: the number of required functions grows exponentially with the dimensionality of the input vector (this phenomenon is known as the curse of dimensionality [112]). One way to alleviate this problem is to use a more general definition of the distance:

$$\|r - \mu\|^2 = (r - \mu)^\Lambda(r - \mu),$$

(6.7)

where $\Lambda$ is a covariance matrix, which now defines the principal axes and the shape of the $G(\|r - \mu\|)$ [112]. But this makes the training a nonlinear problem and the main attractive feature of the RBF network is lost. All these considerations should be taken into account when one constructs a NN for a specific problem.
6.3 Application of Neural Networks to solid state problems

6.3.1 Choice of Degrees of Freedom

Our goal is to represent the energy of a solid state system empirically, given its set of atomic coordinates $\mathbf{R} = (\mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_{N_a})$. The systems of interest can have millions of atoms, so the full $3N_a$-dimensional vector $\mathbf{R}$ is obviously not suitable as an input vector for our approximator. Besides, the number of atoms in the system can vary. So, the first step to make the problem treatable is to break the energy $E$ into individual contributions from each atom $k$:

$$ E(\mathbf{R}) = \sum_{k=1}^{N_a} \epsilon(\mathbf{A}_k), \quad (6.8) $$

where the atomic energy $\epsilon$ depends on the vector $\mathbf{A}_k$, denoting the positions of other atoms in the system with respect to atom $k$. The size of the input vector depends on the range of the interaction. Fortunately, in many cases, such as covalently bound solids, the correlations between atoms are short-ranged. Therefore, one needs to take into account only the local environment of each atom to achieve a good accuracy for the empirical potential. One way to limit the number of neighbors considered is to introduce a cut-off
radius, so that all atoms outside this radius are ignored. This is not very convenient, because atoms can go in and out of the range during a simulation. It is preferable to have a constant number of input variables, so we will consider a constant number of neighbors.

As a test system we choose an \( sp^2 \)-bonded carbon network. As was discussed in Chapter 1, carbon-carbon covalent bonds are very strong, so the \( sp^2 \) network retains its connection structure under a wide range of conditions. The orbital overlaps for a carbon atom in this structure happen primarily with its first and second neighbors. Thus, we model the interaction considering three atoms in the first ring of neighbors and six atoms in the second one, as shown in Fig. 6.5a. The atomic energy depends on a 24-dimensional \( (3 \times 9 - 3 = 24) \) vector \( A_k = (r_{k1}, r_{k2}, \cdots, r_{k9}) \). The necessity to consider so many input variables makes the multidimensional fitting a daunting task. Indeed, even if we did a very coarse sampling of the space with three points per dimension, the database would require \( 3^{24} \sim 10^{11} \) points. Our early tests had little success in describing the interaction with this many input parameters. Dimension reduction is a key step to solving this problem.

Figure 6.5. a) Local environment of an \( sp^2 \)-bonded carbon atom consisting of nine nearest neighbors. b) A branch includes only five atoms: 1,2,3,4, and 5; two similar branches have atoms 1,2,3,6,7 and 1,2,3,8,9 respectively. c) The position of atoms in a branch can be described by 12 parameters \( (x_1, x_2, y_2, x_3, y_3, z_3, x_4, y_4, z_4, x_5, y_5, z_5) \), if the first bond is put along the \( x \) direction and the second bond lies in the \( (x, y) \) plane.
We do so on the basis of physical knowledge about the system under consideration. The degrees of freedom are not equivalent: naturally, the first ring of atoms matters more than the second one. The bonding in graphitic structures has very strong angular dependence, therefore one expects strong correlations between the positions of the three nearest atoms. In the Tersoff-Brenner potentials [67] this correlation is incorporated into three-body terms. The second neighbors are farther away, so the correlations between their positions should be far less important. Therefore, we can decouple the input variables corresponding to the second neighbors. These degrees of freedom do not have to be present in the input vector \textit{simultaneously} anymore, which leads to a reduction of the configuration space. Similar to the decomposition of the total energy of the system into the sum of atomic energies $\epsilon(A_k)$, the atomic energy is now a sum of the energies of three \textit{branches}, shown in Fig. 6.5b:

$$\epsilon(r_{k1}, r_{k2}, \cdots, r_{k9}) \equiv \sum_{n=1}^{N_b} \theta(B_{kn}),$$

(6.9)

where $N_b$ is the number of first nearest neighbors and

$$\begin{align*}
B_{k1} &= (r_{k1}, r_{k2}, r_{k3}, r_{k4}, r_{k5}), \\
B_{k2} &= (r_{k1}, r_{k2}, r_{k3}, r_{k6}, r_{k7}), \\
B_{k3} &= (r_{k1}, r_{k2}, r_{k3}, r_{k8}, r_{k9}),
\end{align*}$$

(6.10)

The branch input vector $B_{kn}$ has just 12 variables ($12 = 5 \times 3 - 3$), which is a significant improvement in comparison with the full 24-dimensional vector $A_k$ we introduced originally. Even though we have cut the input vector in half, it is not feasible to accurately approximate an arbitrary function on a 12-dimensional space. In classical models the way to go around this problem is to carefully choose input variables and set of basis functions, and then tune the parameters for equilibrium states of the studied system. Such states correspond to minima in the potential surface, so the function exhibits a quadratic behavior around them. With a proper choice of basis function one can achieve a good description of the function relatively far away from equilibrium. Additional tuning is required to provide adequate description of transition states; usually it is done by sewing functions along possible transition paths.

In the present work we focus primarily on the description of near-equilibrium states. Here the RBF network comes in handy, since the Gaussians are quadratic at small distances from their centers. Yet, they are not limited to just parabolic form and with
sufficient number of RBFs the anharmonic behavior should be picked up automatically. Another attractive feature is that all input variables are used in the calculation of the distance in the configuration space. Therefore, there is no need to break down the functional form into two-body, three-body or higher terms. In our test system many-body correlations are covered up to the fifth term. Finally, the placement of centers and the tuning of the Gaussian widths are not so important because these parameters are effectively readjusted in the process of the linear training. A simple example in one dimension with two Gaussian functions and a bias demonstrates this situation, assuming that we do not go far away from the centers:

\[
C_1 + C_2 e^{-\frac{||x-\mu_2||^2}{2\delta^2}} + C_3 e^{-\frac{||x-\mu_3||^2}{2\delta^2}} \approx C_1 - \frac{1}{2\delta^2} C_2 (x - \mu_2)^2 - \frac{1}{2\delta^2} C_3 (x - \mu_3)^2 = C' - C'' (x - C''')^2. \tag{6.11}
\]

The two Gaussians merge to produce one parabolic function whose center and curvature are expressed in terms of the adjustable parameters \(C_1, C_2,\) and \(C_3\). The original values of \(\mu\) and \(\delta\) are therefore important only for higher terms in the expansion of the exponential.

It should be stressed that the condition \(\|x - \mu_2\|^2 \ll 2\delta^2\) does not correspond to the normal function of a RBF network. As was discussed earlier, in this regime the network poorly describes rapid changes on potential surfaces. But the RBF network offers a number of advantages for the description of near-equilibrium states. In the next chapters we explore the scope of applicability of the RBF networks as multidimensional function approximators.

### 6.3.2 Training method

#### The Singular Value Decomposition (SVD)

In Section 6.2.1 we formulated the concept of training of a RBF network: given a database of \(N\) input vectors \(r_i\) and corresponding \(N\) target values \(t_i\), we want to minimize the error function in Eq. 6.4 by adjusting the \(M\) weights \(C_j\) in Eq. 6.5 (it is assumed that the Gaussian parameters \(\mu_j\) and \(\delta_j\) are defined beforehand). It is convenient to rewrite
the outputs of the RBF network (Eq. 6.5) $y_i$ in a vector form:

$$
\begin{pmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_N
\end{pmatrix}
= 
\begin{pmatrix}
  1 & \Phi_{12} & \cdots & \Phi_{1M} \\
  1 & \Phi_{22} & \cdots & \Phi_{2M} \\
  \vdots & \vdots & \ddots & \vdots \\
  1 & \Phi_{N2} & \cdots & \Phi_{NM}
\end{pmatrix}
\begin{pmatrix}
  C_1 \\
  C_2 \\
  \vdots \\
  C_M
\end{pmatrix},
$$

(6.12)

or, in short,

$$
y = \Phi C.
$$

(6.13)

The elements $\Phi_{ij}$ of matrix $\Phi$ are calculated by plugging an $i$th input $r_i$ into the $j$th Gaussian $G_j(r)$. Thus, the magnitude of the matrix element $\Phi_{ij}$ represents how far this data point is from the particular Gaussian center. In a well-constructed network with properly spaced centers each data point should generate at least a few non-zero matrix elements. For smooth functions this number does not have to be large, thus the matrix will be sparse and can be handled more easily.

Now let’s rewrite the error function in the vector form as well:

$$
Error = \frac{1}{2} \sum_{i=1}^{N} \|y_i - t_i\|^2 = \frac{1}{2} \|y - t\|^2 = \frac{1}{2} \|\Phi C - t\|^2
$$

(6.14)

The least squares problem can be tackled in different ways. We choose a convenient and numerically robust method based on the singular value decomposition (SVD) of matrix $\Phi$ [113]. Due to the importance of this method for the training of our network I will briefly go over its main features.

It can be proved that for any real matrix $\Phi$ ($N \times M$) there exist orthogonal matrices (in the literature the matrix dimensions are usually denoted as $m \times n$)

$$
U = [u_1, \cdots, u_N] \in \mathbb{R}^{N \times N}, \quad V = [v_1, \cdots, v_M] \in \mathbb{R}^{M \times M}
$$

(6.15)

such that, for $M \leq N$

$$
U^T \Phi V = \Sigma = \text{diag}(\sigma_1, \cdots, \sigma_M) \in \mathbb{R}^{N \times M}
$$

(6.16)
The SVD reveals a great deal about the structure of a matrix. If we define \( r \) by

\[
\sigma_1 \geq \sigma_2 \geq \sigma_{r+1} = \cdots = \sigma_M = 0,
\]

then \( \text{rank}(\Phi) = r \). For \( N = M \), even if \( r = M \), the matrix can be close to singular. Then the inversion problem becomes ill-conditioned and many matrix inversion algorithms will give a numerically unstable solution. The analysis of the smallest singular values can detect this problem.

The SVD is particularly important for the least squares problem. It provides a neat expression for the vector \( \mathbf{C}_{LS} \), which minimizes the error function in Eq. 6.14:

\[
\mathbf{C}_{LS} = \sum_{i=1}^{r} \frac{\mathbf{u}_i^T \mathbf{t}}{\sigma_i} \mathbf{v}_i
\]

and gives the residual error:

\[
\text{Error} = \frac{1}{2} \| \Phi \mathbf{C}_{LS} - \mathbf{t} \|^2 = \frac{1}{2} \sum_{i=r+1}^{N} (\mathbf{u}_i^T \mathbf{t})^2.
\]

Indeed, for any \( \mathbf{C} \in \mathbb{R}^M \)

\[
\| \Phi \mathbf{C}_{LS} - \mathbf{t} \|^2 = \| (U^T \Phi V) (V^T \mathbf{C}) - U^T \mathbf{t} \|^2 =
\]

\[
\| \Sigma \alpha - U^T \mathbf{t} \|^2 = \sum_{i=1}^{r} (\sigma_i \alpha_i - u_i^T \mathbf{t})^2 + \sum_{i=r+1}^{N} (u_i^T \mathbf{t})^2
\]

where \( \alpha = V^T \mathbf{C} \). Clearly, if \( \mathbf{C} \) solves the LS problem, then \( \alpha_i = (u_i^T \mathbf{t} / \sigma_i) \) for \( i = 1 : r \). If we set \( \alpha(r + 1 : M) = 0 \), then the resulting \( \mathbf{C} \) has minimal 2-norm.

In our network training problem some data points can be almost collinear, or Gaussian centers may be placed too close to one another. The SVD allows to construct a robust solution, which avoids the possible numerical instabilities caused by this issue.

**Generalization of the training process**

So far we have assumed that for each input vector \( \mathbf{r}_i \) there is a target value \( t_i \) of the atomic energy. Unfortunately, quantum calculations provide us only with an integral value: the total energy of the system. This is a big setback. First, it reduces the amount of input data. Second, it does not allow us to directly use the training methods, discussed above. We work around this problem by using the linear property of our RBF network.
Indeed, the total energy of the system in the \(i\)th configuration can be written as

\[
E(R_i) = \sum_{k=1}^{N_a} \epsilon(A_{ik}) = \sum_{k=1}^{N_a} \left( \sum_{j=2}^{M} C_j G_j(A_{ik}) + C_1 \right) = \sum_{j=2}^{M} C_j \sum_{k=1}^{N_a} G_j(A_{ik}) + N_a C_1. \tag{6.21}
\]

What we achieve by switching the order of the summation is that now the total energy of the system is a linear combination of new nonlinear functions \(\Phi_{ij} = \sum_{k=1}^{N_a} G_j(r_{ik})\). So, instead of using simple Gaussians for calculation of the matrix elements \(\Phi_{ij}\), we use a sum of the Gaussians, over all atoms in the system. A similar transformation can be performed if we further decompose the atomic energy into branch energies (Eq. 6.9. In this case the matrix elements become

\[
\Phi^E_{ij} = \sum_{k=1}^{N_a} \sum_{n=1}^{N_b} G_j(B_{ikn}). \tag{6.22}
\]

With this new definition of nonlinear activation functions we can use the total energy of the system as a target value and proceed with the linear training of the network as usual. One of the concerns is whether the local energies really reflect the local environment or they are adapted so that their sum matches the total energy. We investigate this question in the following section.

At the beginning of this section I pointed out that as far as the energy of the system is concerned, only the total value can be extracted from quantum calculations. However, local information about the interatomic interaction is also available in the form of the forces on the atoms. So, we go one step further and try to use this additional information to help tune the network. The linearity of our RBF layer allows us to naturally incorporate the force data into the training.

The \(q\)th component of the force on the \(k\)th atom is

\[
F_{kq} = -\frac{\partial E}{\partial r_{kq}} = -\frac{\partial}{\partial r_{kq}} \left( \sum_{k'=1}^{N_a} \sum_{n=1}^{N_b} \left( \sum_{j=2}^{M} (C_j G_j(B_{ik'n}) + C_1) \right) \right) = \sum_{j=2}^{M} C_j \sum_{k'=1}^{N_a} \sum_{n=1}^{N_b} \frac{\partial G_j(B_{ik'n})}{\partial r_{kq}}. \tag{6.23}
\]

Again we are able to express the output of the network as a linear superposition of...
activation functions (note that the bias is eliminated after taking the derivatives). Even though the summation runs over all atoms \((k')\), nonzero terms appear only when the vector \(B_{ik'p}\) contains the coordinates of atom \(k\), which happens if \(k'\) is \(k\) or a (first or second) neighbor of atom \(k\). Therefore the forces can be expressed in an analytic form.

The corresponding matrix elements are now

\[
\Phi^F_{i(i,k,q)} = \sum_{k'=1}^{N_a} \sum_{n=1}^{N_b} \frac{\partial G_j(B_{ik'n})}{\partial r_{kq}}.
\]

We would like to have a linear ordering of the forces in our database, so we use the index \(i(i,k,q) = i + N(k-1) + N_aN(q-1)\) to denote the components of the \(3N_aN\)-dimensional data vector \(F\). Then the total output of the network \(y\) for each point in the database can be expressed as \(y = (E, F)\), and in the matrix form

\[
\begin{pmatrix}
E_1 \\
E_2 \\
\vdots \\
E_N \\
F_1 \\
F_2 \\
\vdots \\
F_{3N_aN}
\end{pmatrix}
= \begin{pmatrix}
1 & \Phi^E_{12} & \cdots & \Phi^E_{1M} \\
1 & \Phi^E_{22} & \cdots & \Phi^E_{2M} \\
\vdots & \vdots & \ddots & \vdots \\
1 & \Phi^E_{N2} & \cdots & \Phi^E_{NM} \\
0 & \Phi^F_{12} & \cdots & \Phi^F_{1M} \\
0 & \Phi^F_{22} & \cdots & \Phi^F_{2M} \\
\vdots & \vdots & \ddots & \vdots \\
0 & \Phi^F_{(3N_aN)2} & \cdots & \Phi^F_{(3N_aN)M}
\end{pmatrix}
\begin{pmatrix}
C_1 \\
C_2 \\
\vdots \\
C_M
\end{pmatrix},
\]

(6.25)

Note, that there are many more force data points than energy points in the database. However, not all of them are independent: according to third Newton’s law \(\sum_{k=1}^{N_a} F_{kq} = 0, \quad q = 1, \cdots, 3\). Even in a small subsystem of the whole structure the forces are somewhat correlated and data points might be almost collinear. Therefore, when we write down the error function, we need to choose a subset of force data points. They also have different units than the energy data, so a conversion factor must be introduced. By varying this parameter we can control what we want our network to be primarily tuned to: the energy, the forces, or both.
6.3.3 Application to graphitic structures

Database and parameters of the RBF network

We choose a 32-atom graphitic unit cell and generate an ab initio database by running molecular dynamics (MD) simulations with VASP [28, 29, 30, 31]. First, we use a low accuracy (PREC=LOW) and just one Γ point to obtain sample configurations around the equilibrium, and then we recalculate the energy accurately at PREC=HIGH and a $4 \times 4$ k-mesh in the $x-y$ directions. The temperature of the system in the MD run is set to 400 K and maintained by the Nosé-Hoover thermostat [114]. The system starts from the equilibrium planar configuration with the bond length of 1.41 Å. Thermal fluctuations of the system on the order of 400 K turn out to be modulated by low-frequency buckling of the structure in and out of the plane. Due to the buckling, for some configurations the potential energy of the system reaches 2500 K. The average energy of the 850 structures in the database is 1050 K and the standard deviation is 450 K.

Next we build a RBF network as follows. The input vector contains 12 variables, which describe the positions of five atoms with respect to atom $k$ as shown in Fig. 6.5. The coordinates are cartesian; use of spherical coordinates produces similar results. The ranges of values for each variable found in the database are

\[
\begin{align*}
    x_1 & \in (1.242, 1.664) \\
    x_2 & \in (-1.012, -0.366) \\
    y_2 & \in (0.928, 1.547) \\
    x_3 & \in (-1.034, -0.370) \\
    y_3 & \in (-1.481, -0.876) \\
    z_3 & \in (-0.858, 0.856) \\
    x_4 & \in (1.777, 2.432) \\
    y_4 & \in (0.850, 1.545) \\
    z_4 & \in (-0.823, 0.842) \\
    x_5 & \in (1.776, 2.472) \\
    y_5 & \in (-1.485, -0.857) \\
    z_5 & \in (-0.661, 0.626)
\end{align*}
\]

where the distances are given in Å. The input variables are scaled to be zero and one on the low and the upper ends of the ranges indicated above. In a branch, one of the three
nearest neighbors is special: two second nearest neighbors are attached to it. So, there is a choice of how to pick the neighbor number two. The choice is important since this bond needs to be put in the $x-y$ plane and the whole branch has to rotate with it. Tests show that if this bond is chosen as the longest of the two candidates, the other variables are limited to a smaller range and the network performs better. The energy of the system should not depend on this choice, or on whether we reflect of the branch with respect to the $x-y$ plane. So, in general we can quadruple the database if needed. The total number of input vectors (branches) in the present database is $(\text{data structures}) \times (\text{atoms per structure}) \times (\text{branches per atom}) = 850 \times 32 \times 3 = 81600$.

The choice of the RBF centers and their widths turns out to be not very important for the function of the network. Positioning of the centers in the regions where data points cluster works as well as simple placement of centers on random branches; we use the latter. If we place 400 centers on branches the average distance between centers $\langle d \rangle$ is 0.021; the average maximum distance $\langle d_m \rangle$ is 0.85 and each center has 225 centers within 0.5 on average (the distance here is calculated with the factor $\lambda = 0.6$ introduced at the end of the paragraph). The network demonstrates similar performance for the values of the Gaussian widths $\delta_j$ in the range from 1.0 to 3.0. We set $\delta_j = 2.0$, ($j = 2, \cdots, M$), so the parameter $\alpha = \langle d \rangle / \delta = 0.01$, which defines the regime of function of the RBF network, is much smaller than one. The weak dependence of the results on the RBF parameters must be caused by the near-parabolic behavior of the interatomic potential around the equilibrium, as discussed above. Another issue, demonstrated in Section 6.2.2, is the relative importance of the degrees of freedom. A RBF network has problems describing functions where some input variable are much less important than the others. Use of the covariance matrix (Eq. 6.7) helps solve this problem but, finding the matrix elements is a difficult task. However, we know that the atomic energy should be more sensitive to the positions of the first neighbors than the second neighbors. We choose a simple, diagonal form of the covariance matrix $\Lambda = diag(1, 1, 1, 1, 1, \lambda, \lambda, \lambda, \lambda, \lambda)$ and tune just one parameter $\lambda$. The optimal value for this parameter is found to be 0.6; with this adjustment the errors in the description of the energies and the forces by the network decrease by about 20% for the same size RBF network.

Results

Figures 6.6 and 6.7 show the main results on how the network approximates the total energy and the forces for the graphitic structures in the database. In Fig. 6.6 the network is trained on 200 data structures and the number of RBFs is increased from 30
to 180. After the network is tuned it is tested on data structures not included in the training process. The error for the energy is calculated as $\left( \frac{\langle \Delta E^2 \rangle}{\langle E^2 \rangle} \right)^{1/2}$. The two curves represent different training conditions: the black curve corresponds to training on the total energy (E-type) only and the red one stands for training on the energy and force data points together (EF-type). Notice that the errors in the first type of training increase, when we use more than 120 RBFs (which is about half of the training set size). This is a known effect in the fitting procedures: a large number of parameters allows us to get a good fit to the training set of data (the residual error decreases monotonically and at $M = N$ it should go down to zero), but the over-trained network loses its generality and performs poorly on new sets of data. We find that the optimal number of RBFs for the E-type training is about a third of the size of the training set. At the optimal value of $M = 100$ the network predicts the total energy with just a 1.1% error. The forces are also described quite accurately (to within 7.7%) even though the training set contains the total energy only. A significant improvement is achieved when force data are also incorporated in the training (EF-type). To avoid correlations between data points we include the three components of the forces on every fourth atom. Therefore the data set becomes much larger: $N' = (3 \times N_a/4 + 1)N$. The prediction errors go down by a factor of two. The over-training is not observed at $M = N = 200$ here because the number of data $N'$ is much larger than that. If more structures are considered (see Fig. 6.7), the E-type training tunes the network much better: the errors decrease to 0.5% and 5.1% for the energy and forces, respectively. The expansion of the training set leads to small improvement for the EF-type training because the network has already achieved nearly-optimal performance. From now on we use a network with 400 RBFs EF-type trained on 600 data structures. The corresponding errors are 0.6% for the energy and 2.8% for the forces. One expects 2-3 meV per atom errors in our LDA calculations for this system; these errors are about 0.3% of the total energy of the system at 1000 K. So, the saturation in the accuracy of our network could be caused by numerical noise in the data set. Another possible reason for the plateau is the decision to consider the first and second neighbors only and decouple the second-neighbor degrees of freedom in our empirical model. Still, the model significantly outperforms the Tersoff-Brenner potential [67], which gives a 9% and 31% errors for the energy and forces respectively, and the tight-binding method [115] with 18% errors for both the energy and the forces. The big discrepancy between these methods and the LDA data is quite surprising. An analysis of the errors shows that the tight-binding method consistently overestimates the energies and the forces in this case. The Tersoff-Brenner potential provides good description of
Figure 6.6. Performance test of the RBF network trained on 200 data structures. The black and red curves correspond to the E-type and EF-type training (see text) respectively. The top panel shows the error in the total energy calculated as \( \left( \langle \Delta E^2 \rangle / \langle E^2 \rangle \right)^{1/2} \). The error in the force on the bottom panel is \( \left( \langle \Delta F^2 \rangle / \langle F^2 \rangle \right)^{1/2} \). The tests are performed on 200 configurations, not included in the training set.

the interaction around equilibrium, which deteriorates as the system moves away from
Figure 6.7. Same as in Fig. 6.6, only the training set has 600 structures.

it. The 31% force error comes primarily from large mismatches of the forces far away from the equilibrium.

To investigate the accuracy of our model we compare it with a tight-binding method and the Tersoff-Brenner potential. The latter is designed primarily to describe energies of different stable structures (hydrocarbons) and is not expected to describe the force...
constants very accurately. In the following test we start with a 128-atom graphitic unit cell at equilibrium and shift one of the atoms towards the center of a hexagon, as shown Figures 6.8, 6.9, 6.10. Our goal is to see how different models describe the force constants and anharmonic behavior of the potential at large displacements. The first plot (Fig. 6.8) shows the variation in energy of the distorted structure. The nearly perfect fit of the LDA with the RBF network is very satisfying but not surprising: the network is trained in part on the total energy of the system and is expected to describe it accurately. The tight-binding and the Tersoff-Brenner models overestimate the energy by about 20%. A more informative test of our model is the examination of the forces in the structure. The top panel in Fig. 6.9 shows the force on the displaced atom. Again, the RBF network output matches the LDA data very well, while the other two methods overestimate the forces by more than 20%. The bottom panel shows the force on a third neighbor of the displaced atom. In our model the branch energy (Eq. 6.9) does not include the third or higher neighbors, so it might not be obvious how the correlations between atoms are carried over beyond second neighbors. The answer is that when the force is calculated as a partial derivative of the total energy with respect to the distance (Eq. 6.23), the range of the interaction effectively expands up to the fourth neighbors. This is a subtle effect.
Figure 6.9. The same setup as in Fig. 6.8. This plot shows the forces on the displaced atom label ‘1’ (top panel) and its third neighbor labeled ‘2’ (bottom panel).

(the LDA predicts that the force on the third neighbor is 20 times smaller than that on the displaced atom); it is interesting to see how our model handles it. Surprisingly, the force on the third neighbor is also in excellent agreement with the LDA. The tight-
binding force is off by about 50% and the Tersoff-Brenner potential does not reproduce this longer-ranged correlation well at all: the force is too small and has the wrong sign.

The last piece of information in this test concerns the deviation of the interaction potential from the harmonic behavior. For each displacement $x_i$ we crudely estimate the derivative of the force as $(F_i - F_{i-1})/(x_i - x_{i-1})$ and normalize it to be one at $x = 0$. Fig. 6.10 demonstrates that at large displacements the anharmonic terms become important: the force constant in the LDA drops by 30% (at $x = 0.2 \text{ Å}$). All methods reflect the decrease of the force constant, but the Tersoff-Brenner potential underestimates and the tight-binding method overestimates the effect. Our model follows the LDA data exactly up to 0.15 Å, impressively outperforming the other two models. The small deviation after that must be caused by the absence of sampling points so far from the equilibrium. That conclusion is based on another test, in which the RBF network was trained on the same database but with discarded high-energy structures. In that case our model perfectly matched the LDA curve up to 0.1 Å. Therefore, one can expect that expansion of the database will make the description of these effects with the RBF network even more accurate if needed.

Overall, the tests validate our decomposition of the total energy into branches and
Figure 6.11. Reaction path for the Stone-Wales transformation found with the NEB. The calculations are done in LDA with VASP at PREC=HIGH (black dots). The unit cell is kept constant for all steps. Different curves show the energy of the system calculated with the RBF (in red), Tersoff-Brenner (in blue), and the tight-binding methods.

Our training method \textit{a posteriori}. Not only does the RBF network match the total energy of the system, but it also accurately predicts the forces on the atoms. Hence, our model could be used as a valuable tool for description of mechanical properties of solids near equilibria with \textit{ab initio} accuracy. One direct application would be the calculation of phonon modes. In the future it would be interesting to compare our model with the force-constant model [116], primarily used for the computation of the vibrational properties of solids.

In our last test we investigate how far from the equilibrium our model remains valid. We consider a Stone-Wales transformation (see Section 5.3.2) in a 32-atom graphitic unit cell (Fig. 6.11). The transition path is calculated in the LDA with the NEB method [102] as implemented in VASP. The final configuration with the rotated bond is found to be 6.6 eV above graphene; the reaction barrier is 9.8 eV (the size of the unit cell is kept constant during the reaction), see Section 5.3.2. We take the structures from the \textit{ab initio} calculations and calculated energies with different methods. This is an extreme test for our model because the transformation involves bond breaking. The RBF network has no sampling points along this path, so it extrapolates the energies around the saddle point. The RBFs continue to grow near-parabolically, overestimating the reaction barrier by
a factor of two. Some kind of cutoff function needs to be introduced to overcome this problem. At this point it is not clear how to do it, further development of the model is required. The tight-binding curve follows the reaction path very well. The energies given by the Tersoff-Brenner potential are also in good agreement with the LDA results. However, the forces around the saddle point are not smooth. This appears to result from the cutoff functions used in the potential. When this potential is used for finding reaction paths the convergence of the NEB is poor due to the large changes of the forces. The final state with the rotated bond has the same \( sp^2 \)-type bonding but has two pentagons and two heptagons. This configuration is far from what our RBF network was trained on: for example, the rotated bond is much shorter (1.27 Å) than the equilibrium one (1.42 Å). Nevertheless, our model predicts the energy of this state slightly better than the other two methods. Therefore, one expects good transferability of our model on other \( sp^2 \)-connected carbon networks such as carbon nanotubes, nanoporous carbon and fullerenes. The RBF network can be additionally tuned to the new structures by including the new configurations in the database.

\[ \text{6.4 Summary} \]

We have formulated an algorithm for applying NNs to description of the interatomic interaction in solids. The key step in this algorithm is the use of a RBF network, which has the advantage over MLP for describing configurations around equilibria. A careful choice of the input variables is also required to reduce the size of the configuration space. A training technique has been proposed that makes use of the total binding energy of the system and the forces on each atom. We have implemented the RBF network, trained it on an \( ab \text{ initio} \) database and tested it on different structures. For structures around the equilibrium our method significantly outperforms the tight-binding and Tersoff-Brenner models in the description of the total energy, force constants and anharmonic behavior of the interaction potential. This makes our model suitable for high-accuracy calculations of vibrational properties of solids; further tests and comparison to the force-constant methods should be conducted. Our RBF network has many more parameters than other empirical methods, but the computation time still grows linearly with the system size. So, this is a small cost for the \( ab \text{ initio} \) accuracy provided by our model. Transition states are currently beyond the capabilities of the RBF network; further development of the model is needed.
Bibliography


[5] Quantization of electronic motion in the circumferential direction means that the electronic levels are discrete, with an average level separation that varies as $1/R^2$. Therefore small-diameter nanotubes are insulating in the transverse direction.


[38] Y. Baskin and L. Mayer, Phys. Rev. 100, 544 (1955)
We are interested in the change in the equilibrium spacing, so for convenience, we slightly adjusted the LJ potential parameters to have the equilibrium spacing of AB stacking at exactly 3.34 Å.

The corrugation does not go to zero in the infinite system but remains finite. Therefore infinite incommensurate systems do not have zero barrier to sliding, but have zero barrier to sliding per atom.

One-dimensional electrons within a single layer of a multiwalled tube can also experience a weak external incommensurate potential from adjacent layers.

In our paper we have ‘a few MPa’; that estimate of the sheer strength was miscalculated.


[69] The (10,0) tube does extend slightly, since the commensuration turns into a pseudo-commensuration under axial extension for this special case.

[70] The (n,n) tubes have unexpectedly large binding energies for matching bond lengths. In a narrow (n,n) tube, a central axial row of atoms forming the line of contact with the substrate can be placed in the most energetically favorable position: above the hexagonal centers in the graphite substrate (as in AB stacked graphite). Neighboring rows of atoms, which are positioned in energetically less favorable positions directly above atoms in the substrate, curve away from the substrate, so have less important interactions. If the bonds on the (n,n) nanotube are mismatched to the substrate, then this effect disappears, since atoms on the nanotube are uniformly spread along the substrate unit cell. This effect is therefore restricted to short and narrow armchair nanotubes.


Vita
Aleksey N. Kolmogorov

Education

2004  Ph.D., Condensed Matter Physics, The Pennsylvania State University
2004  Ph.D. Minor, Computer Science, The Pennsylvania State University
1995-1997 Research Assistant, P.L. Kapitza Institute for Physical Problems
1995  M.S., Condensed Matter Physics, Moscow Institute of Physics and Technology

Awards

1997-1999 Braddock Graduate Fellowship, PSU
1997-1998 Materials Research Institute Fellowship, PSU
1995-1996 Landau Scholarship, P.L. Kapitza IPP
1995  Soros Scholarship, P.L. Kapitza IPP