

NANOSCALE DEVICE MODELING:
FROM MOSFETS TO MOLECULES

A Thesis

Submitted to the Faculty

of

Purdue University

by

Prashant Subhash Damle

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

May 2003

To *aai*, *baba*, *ajoba* and *apu*.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge my thesis advisor Prof. Supriyo Datta for readily providing many helpful ideas and suggestions throughout the course of this work. I would like to thank Dr. Avik Ghosh for important contributions to this work. I would also like to thank Prof. Mark Lundstrom for his guidance and critical reading of my work, Dr. David J. Frank for mentoring me during my summer (2001) internship at the IBM Thomas J. Watson research center in Yorktown Heights, New York, and Prof. James A. Cooper, Jr. for his guidance during my first year at Purdue.

I am grateful to Dr. Zhibin Ren for sharing with me his computer program to solve the Poisson's equation, to Dr. Magnus Paulsson for painstakingly going through my computer programs and suggesting improvements, to Dr. Sebastien Goasguen for helping me run large simulations on Prof. Lundstrom's Beowulf Linux compute cluster, and to Dr. Uday Savagaonkar for giving me helpful tips related to computational issues. Special thanks go to Ramesh Venugopal, Titash Rakshit, Ferdows Zahid and Albert Liang for many useful discussions.

Parts of Chapter 2 in this dissertation are reproduced with permission from a forthcoming book by Prof. Supriyo Datta entitled "Quantum Phenomena: From Atoms To Transistors", and from M. Paulsson et al. in "Handbook of Nanoscience, Engineering and Technology" (W.A. Goddard III et al., Ed.). CRC Press, Boca Raton, Florida. © 2003 by CRC Press LLC.

This work was supported by the National Science Foundation (NSF) and the US Army Research Office (ARO) under grants number 9809520-ECS and DAAD19-99-1-0198.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vii
LIST OF FIGURES	viii
ABSTRACT	xviii
1 INTRODUCTION	1
1.1 The Self-consistent Field Method	1
1.2 Step-by-step Solution Procedure	4
1.3 Outline Of This Dissertation	5
2 THE NEGF FORMALISM	9
2.1 Current Flow Through A Single Discrete Energy Level	10
2.2 Current Flow Through A Single Broadened Energy Level	13
2.3 NEGF Equations As A Multi-level Generalization	13
2.4 Origin Of The Contact Self-energy Matrix	16
2.5 Non-orthogonal Basis	19
3 DUAL-GATE NANOTRANSISTOR: EFFECTIVE MASS MODEL	21
3.1 Hamiltonian	22
3.1.1 Modified Fermi function	25
3.1.2 Basis transformation to mode space	26
3.2 Self-energy	29
3.3 Self-consistent Potential	32
3.4 Results	33
3.5 Summary	36
4 MOLECULAR TRANSISTOR: SEMI-EMPIRICAL ATOMIC ORBITAL MODEL	37
4.1 Introduction	37

	Page
4.2 Hamiltonian	41
4.3 Self-energy	43
4.3.1 Where is the Fermi energy?	45
4.4 Self-consistent Potential	45
4.4.1 Capacitance model	46
4.4.2 Numerical solution	48
4.5 Results	49
4.5.1 Ideal gate control, on state	49
4.5.2 Ideal gate control, off state	52
4.5.3 Estimate of gate control	53
4.5.4 Comparison: capacitance model versus numerical solution	55
4.6 Summary	56
5 MOLECULAR WIRES: AB INITIO ATOMIC ORBITAL MODEL	59
5.1 Introduction	59
5.2 Hamiltonian And Self-consistent Potential	62
5.3 Self-energy	64
5.4 Results	69
5.4.1 Equilibrium	69
5.4.2 Non-equilibrium	72
5.5 Summary	80
6 MOLECULAR WIRES II: THE “EXTENDED” MOLECULE	83
6.1 Advantages Of Using An Extended Molecule	83
6.2 Locating The Fermi Energy	85
6.2.1 Choice of the “window”	87
6.2.2 Rigorous approach	89
6.3 I-V Characteristics	94
6.3.1 Quantum point contact	95
6.3.2 Xylyl Dithiol: symmetric coupling	97

	Page
6.3.3 Xylyl Dithiol: asymmetric coupling	99
6.4 Summary And Outlook	101
LIST OF REFERENCES	104
APPENDIX A: SELF-ENERGY - IMPOSING THE GOLD FCC SYMMETRY	111
APPENDIX B: DENSITY MATRIX - ANALYTICAL INTEGRATION	115
APPENDIX C: GOLD SURFACE DOS WITH MORE NEAR NEIGHBORS .	117
APPENDIX D: MODIFICATIONS TO GAUSSIAN '98	121
D.1 Link 302	121
D.2 Link 401	123
D.3 Link 502	124
D.4 MATLAB Script	126
VITA	131

LIST OF TABLES

Table	Page
6.1 Results of test for the choice of window to determine the location of E_b . <i>PDT</i> denotes Phenyl Dithiol (Figure 6.1) while <i>QPC</i> denotes quantum point contact, which is a linear chain of six gold atoms (Figure 6.7). Au_n denotes that n gold atoms were present on each side in the extended molecule. E_b was adjusted till $E_f - E_b^H$ and $E_f - E_c^H$ came within about 0.5 eV of each other, more accuracy needs a bigger gold cluster on each side in the extended molecule. The results of the calculation with the isolated extended <i>PDT</i> molecule (Figure 6.1) depend on whether we use $n = 12$ (row (1) below) or $n = 13$ (row (2) below) gold atoms. We expect this odd-even discrepancy to become smaller and smaller as the number n of gold atoms increases. The final adjusted E_b does lie slightly below the lowest upper levels (Figure 6.3), thus giving us confidence in our choice of the window width.	94

LIST OF FIGURES

Figure	Page
1.1 Schematic view of a device coupled to contacts and the solution procedure. H is the device Hamiltonian and U_{SC} is the self-consistent potential. The contact Fermi levels μ_1 and μ_2 are equal at equilibrium and differ by an amount qV when a bias of V volts is applied across the device. The effect of the contacts on the device is <i>exactly</i> accounted for by the contact self-energy matrices Σ_1 and Σ_2 . The self-energy matrix Σ_p may be used to model non-coherent scattering processes inside the device. The word “Poisson’s” is in quotes as a reminder that more sophisticated schemes like Hartree-Fock (HF) or density functional theory (DFT) may also be used to obtain the self-consistent potential. The Non-equilibrium Green’s Function (NEGF) formalism is used to obtain the electron density ρ and current I	2
2.1 Current flow through a single energy level. One contact keeps filling it up and the other keeps emptying it causing a net current flow from one contact to the other.	10
2.2 Schematic description of resonant current conduction through discrete molecular levels. H and L denote the highest occupied and the lowest unoccupied molecular levels respectively. An electron in any of these levels has a finite lifetime before it escapes into one of the contacts. Onset of current corresponds to crossing of a molecular level by a contact chemical potential under bias.	12
2.3 The complete system consisting of device and contact may be partitioned into a device subspace and a contact subspace. The effect of the infinitely large contact is then incorporated through a self-energy matrix Σ which is of the <i>same</i> size as the device Hamiltonian.	17
3.1 Schematic view of a dual-gate MOSFET.	23
3.2 Choice of basis: We use the eigenstate representation for the transverse (z) dimension but a discrete real space lattice for the longitudinal (x - y) dimensions. The coordinate axes are defined in Fig. 3.1.	23
3.3 y -confinement in a dual-gate MOSFET gives rise to discrete sub-bands. .	26

Figure	Page
3.4 The MOSFET is coupled to <i>two</i> semi-infinite contacts and not just one. Thus two self-energy matrices are needed, one for the source contact and another for the drain contact.	31
3.5 Drain current vs. drain and gate voltages. (a) linear scale, (b) log scale. The on and off currents are about $1800 \mu\text{A}/\mu\text{m}$ and $0.03 \mu\text{A}/\mu\text{m}$ respectively. An ideal subthreshold slope of 60 mV/dec is observed.	34
3.6 Potential vs. position along the length of the MOSFET, (a) low V_{GS} (b) high V_{GS} . In (b) the source to drain barrier height stops decreasing for high drain bias leading to current saturation as seen in Fig. 3.5a.	35
4.1 (a) Schematic view of a Phenyl Dithiol molecule coupled to gold source and drain contacts. A third (gate) terminal modulates conductance of the molecule. The phenyl ring is shown in the plane of the paper for clarity, in the actual simulation the phenyl ring is perpendicular to the plane of the paper, “facing” the gate electrode. (b) and (c) are reproduced for convenience from Figure 1.1. The only difference between (b) and Figure 1.1 is the presence of the third (gate) terminal.	40
4.2 Comparison of the simple π orbital based model with an ab initio model (density functional theory with atomic orbital basis set) for the PDT molecule. The energy levels can be divided in two sets: occupied levels (analogous to the valence band) and unoccupied levels (analogous to the conduction band). The energy gap (analogous to the bandgap) is the energy difference between the Highest Occupied Molecular Orbital (HOMO) level (analogous to the top of the valence band) and the Lowest Unoccupied Molecular Orbital (LUMO) level (analogous to the bottom of the conduction band). The simple π model agrees very well with the ab initio calculation in both the energy gap (middle) and HOMO and LUMO wavefunctions (left and right). The energy levels obtained from the simple model were equally shifted in energy so as to make the HOMO level coincide with the ab initio HOMO level.	42
4.3 The discrete levels of an isolated molecule broaden into a continuous density of states (DOS) upon coupling to contacts. The gold FCC (111) contacts are modeled using a basis of one 's' type orbital on each gold atom. The LUMO wavefunction is localized on the phenyl ring (Fig. 4.2) and gives a sharp peak in the DOS. The HOMO is delocalized and gives a comparatively broadened peak in the DOS. The equilibrium Fermi energy E_f ($\sim -5.1 \text{ eV}$ for bulk gold) lies just above the HOMO level.	44

Figure	Page	
4.4	Equivalent circuit model used to obtain the homogeneous (or zero charge) solution to the Poisson's equation. The molecular potential is controlled by the gate if the capacitive coupling C_{MG} between the molecule and the gate is much greater than the capacitive couplings C_{MS} and C_{MD} between the molecule and the source and drain respectively (see Eqs. 4.5, 4.6 and related discussion).	46
4.5	Three terminal molecular drain current vs. drain to source voltage characteristic with (a) Ideal gate control ($\beta = 1$), negative drain bias, (b) Ideal gate control, positive drain bias, (c) No gate control ($\beta = 0$), negative drain bias and (d) no gate control, positive drain bias. With ideal gate control the IV is asymmetric with respect to drain bias. Good saturation and gate modulation is seen for negative drain bias but not for positive drain bias. With no gate control the IV is symmetric with respect to drain bias. For an explanation of the underlying mechanism for each of these IV curves, see Fig. 4.6.	50
4.6	Gate induced current saturation mechanism: Assuming that the gate is very close to the molecule (ideal gate control, $\beta = 1$), the gate holds the molecular DOS fixed relative to the source Fermi level μ_1 because the gate is held at a fixed potential with respect to the source. When a negative drain bias is applied (top left), the drain Fermi level μ_2 moves up relative to the molecular DOS. Since the DOS dies out in the gap, for sufficiently high drain bias, no more DOS comes in the μ_1 - μ_2 window and the current saturates. When a positive drain bias is applied (top right), μ_2 moves down relative to DOS and eventually crosses the HOMO. The IV is thus asymmetric. If the gate is far away (no gate control, $\beta = 0$), the DOS lies roughly halfway between the source and drain Fermi levels. In this case, for negative drain bias (bottom left), μ_1 crosses HOMO while for positive drain bias (bottom right) μ_2 crosses HOMO. No gate modulation is seen as expected, and the current is symmetric with respect to drain bias. . .	51
4.7	Subthreshold IV characteristic assuming ideal gate control. The <i>temperature independent</i> subthreshold slope is ~ 300 mV/decade even with an ideal gate. This is because the DOS in the HOMO-LUMO gap dies out very slowly as a function of energy. This slow fall of the DOS may be attributed to the gold metal-induced gap states. Thus a molecular FET with a rigid molecule acting as the channel is a very poor switch.	52

Figure	Page	
4.8	Estimate of gate control using 2D numerical Laplace's solution with a single gated geometry (solid line) and a double gated geometry (dashed line). β (which is a measure of gate control, see Eqs. 4.5, 3.5, 4.8 and related discussion) is plotted as a function of the gate oxide thickness T_{ox} . The length of PDT molecule (equal to the channel length of the molecular FET) is about 1 nm. Thus in order to get good gate control ($\beta > 0.8$) the gate oxide thickness has to be about one tenth of the channel length, or about 1 Angstrom! We have used 3.9 and 2 as the dielectric constants for silicon dioxide and the self-assembled monolayer (SAM) respectively. For a double gated geometry, good gate control can be obtained with more realistic oxide thicknesses (of about 7 Angstroms), as expected simply because two gates can better control the channel than one.	54
4.9	Comparison of the numerical Poisson solution with the capacitance model. We see reasonable agreement in the two solutions despite the simplifications made in the capacitance model. (a) A realistic case with $t_{ox} = 1.5$ nm which yields $\beta = 0.28$ and $U_0 = 1.9$ eV. No current saturation is seen, but some gate modulation is present. (b) An "absurd" case with $t_{ox} = 1$ Angstrom which yields $\beta = 0.82$ and $U_0 = 1$ eV (U_0 is less for this case because the gate is closer to the molecule; screening effect of the gate electrode is larger). The IV for this case looks like that for a MOSFET. Similar IV may be obtained with $t_{ox} = 1$ nm, provided one uses a gate insulator with a dielectric constant about ten times that of silicon dioxide.	56
5.1	Schematic view of a molecule (Phenyl Dithiol) connected to two semi-infinite gold (111) contacts. The self-energy matrices Σ_1 and Σ_2 <i>exactly</i> account for the contacts and are of the same size as the molecular Fock matrix F . The self-consistency scheme is shown for (a) isolated molecule in equilibrium and (b) contact-molecule-contact system under bias.	61
5.2	Device coupled to a semi-infinite contact (top: 1-D contact, bottom: 3-D contact). The contact self-energy depends on the device-contact coupling τ and the contact surface Green's function g which is recursively solved for using the contact on-site and coupling matrices α and β (see text). In case of a 3-D contact, the in-plane periodicity is used to obtain an equivalent 1-D picture with k-dependent α_k and β_k	65
5.3	Central 13 atom part of the 28 atom gold cluster in FCC (111) geometry used to calculate the coupling matrices S_{mn} and F_{mn} in Eq. 5.2. A reference atom (atom 1) and all it's nearest neighbors are shown in this central 13 atom part. 28 atoms are used so as to reduce the 'edge' effects that tend to destroy the symmetries associated with the Fock matrices F_{mn} (see the discussion following Eq. 5.2 and Appendix A).	66

Figure	Page
5.4 Comparison of non-self-consistent evaluations of DOS for phenyl dithiol with and without a constant self-energy approximation. The good correspondence is a consequence of the relatively flat DOS of Au(111) near the Fermi energy.	68
5.5 An appropriate self-energy can take care of bonding even if the ‘molecule’ we define has an unsatisfied valence. The energy levels (B3PW91/ LANL2DZ) of phenyl dithiol (leftmost panel) differ considerably from those of the hydrogen-desorbed molecule (left center), but are restored on coupling to gold (111) clusters (right center). Hybridized electrons from the gold triangle replace those missing from the H atoms, restoring the valency of the sulfur atoms and also give rise to additional gold based levels, shown in dashed lines. Rightmost panel shows the self-consistently calculated density of states (DOS) of a hydrogen-desorbed phenyl dithiol (PDT-2H) chemisorbed on gold contacts (described by self-energy matrices). A sharp LUMO peak (PDT LUMO is localized on the phenyl ring) and a well-broadened HOMO peak (PDT HOMO is delocalized over the molecule) is seen in addition to a non-zero DOS in the gap attributed to metal induced gap states.	70
5.6 Charge transfer and self-consistent band lineup upon coupling a gold nanowire to two gold contacts. The charge neutrality level (CNL) of the wire lies below the Fermi energy of gold, leading to electron flow from contacts to the wire. The DOS (dashed) of the charge-neutral wire redistributes and floats up (solid) self-consistently due to charging energies associated with the electron flow.	71
5.7 (a) I-V of a Gold nanowire with six gold atoms forming a 1-D chain connected to two gold contacts. An energy-independent self-energy $\Sigma(E_f)$ is used to model the contacts. A quantum unit conductance of $2e^2/h$ signifies that $\Sigma(E_f)$ accounts for the perfect transmission by seamlessly joining the nanowire to the contacts. (b) I-V of the same gold nanowire but with reduced contact coupling which decreases level broadening and a resonant tunneling type conduction is seen. The I-V shows step-like behavior with the current jumping up when the contact Fermi levels μ_1 and/or μ_2 cross a molecular level (See Figs 5.8, 5.9).	73
5.8 At equilibrium ($V = 0$) $\mu_1 = \mu_2 = E_f$ and no current flows. For non-zero V the contact Fermi energies separate by an amount eV and a significant current flows only when a molecular level lies in between μ_1 and μ_2 (also see Fig 5.9).	74

Figure	Page
5.9 Energy levels of a gold nanowire weakly coupled to contacts as a function of applied bias. The I-V for this case is shown in Fig 5.7 b. At zero bias the contact Fermi level is close to LUMO. At non-zero bias μ_2 wants to fill up the LUMO and add extra electrons in the device, so charging effects tend to float the levels up till μ_1 comes close to HOMO and wants to empty it. At this point, more electrons are lost than gained and charging effects make the levels float down parallel to μ_1 in order to minimize the loss of electrons.	75
5.10 Left and right LDOS for a wire with a defect in the center for 21 equidistant voltage values from 0 to 2 volts, laterally shifted equally for clarity. The electrons in the left and right segments of the wire are separately in equilibrium with the left and right contacts, and their respective LDOS follow the corresponding contact electrochemical potentials (dashed lines) because of charging effects similar to those seen in Fig 5.9. As the two LDOS slide past each other within the μ_1 - μ_2 window (shown magnified in the inset of the bottom figure at the four voltage points circled in the I-V graph) their peak values (tail of the peak for the left LDOS) come in and out of resonance, producing thereby a weak NDR in the I-V characteristic.	76
5.11 I-V of phenyl dithiol connected to two gold contacts. The underlying mechanism is resonant tunneling - much like that in a gold nanowire weakly coupled to the contacts (see Fig 5.7 b). The gap depends on the proximity of the Fermi energy to the zero bias HOMO level if we neglect charging effects, while the maximum current at the onset of conduction is proportional to the parallel combination of broadenings at the molecular energy (see text). For phenyl dithiol, the predictions from the above estimates are about 4 volts and 40 μ A respectively.	77
5.12 Electrostatics of (a) a weakly contacted gold wire, and (b) a strongly contacted wire with a substitutional defect (a stretched bond at the center), in response to a 2 volt applied bias across the contacts. Top panel: schematic, middle panel: voltage drop along the wire (dashed lines show a linear drop for comparison), bottom panel: density of electrons along the wire. In all plots, the corresponding quantities at equilibrium have been subtracted out in order to remove peaks near the positions of the nuclei for clarity.	79

6.1	Schematic picture of a Phenyl Dithiol molecule coupled to gold contacts. The term “extended molecule” is used to denote the molecular subsystem that includes a few contact atoms. For the results shown in this chapter we include 12 gold atoms on each side of the molecule. The rest of the contacts are treated using self-energy matrices. The contact surface Green’s function is evaluated on the 19 gold atoms adjacent to the 12 “molecular” gold atoms as shown. The coupling between the 12 and 19 gold atoms is used as the molecule-contact coupling, and is evaluated once using unmodified Gaussian ’98 and stored for use in subsequent calculations. As an example we have shown above the Phenyl Dithiol extended molecule, with the following color scheme: gold for gold atoms, red for sulfur atoms, green for carbon atoms and white for hydrogen atoms. We will use this color scheme for the rest of this chapter.	84
6.2	When a molecule is coupled to contacts, there is some charge transfer and the system relaxes to a common Fermi energy E_f at equilibrium. The discrete levels of the molecule broaden into a continuous density of states. When a bias is applied across the molecule, the broadened level closest to the equilibrium Fermi energy carries most of the current, and the conductance gap depends on how far this level is from E_f (see Section 2.1). If the molecule is used in a three-terminal configuration, the location of E_f relative to the HOMO or LUMO band decides if it will function as a p-type or n-type transistor (see Chapter 4).	86
6.3	In our analytical (but approximate) method of calculating the density matrix (see appendix B), we replace the energy dependent self-energy $\Sigma(E)$ with an energy independent $\Sigma(E_f)$ for energies lying in the “window” between E_f and E_b . E_b is fixed between one to two electron-volts below the lowest “upper” level. Since the location of E_b affects the location of E_f relative to the molecular density of states, an independent test for our choice of window necessary. The core levels are expected not to influence the transport properties of the molecule, and are accounted for by using a small (one micro-electron-volt) broadening. Typically the lowest upper level is found near -20 eV, while the core levels are found at -60 eV and -110 eV. The HOMO band of levels is denoted by H while the LUMO band is denoted by L	88

Figure	Page	
6.4	Independent scheme to test the choice of window width. (a) Isolated molecule, (b) Isolated extended molecule and (c) Extended molecule coupled to contacts. For a sufficiently large cluster of contact atoms in the extended molecule, the two cases (b) and (c) are <i>exactly</i> equivalent. Calculation of energy levels in case (b) does not involve any E_b , while that in case (c) does. Knowing the HOMO wavefunction Ψ_a^H of the isolated molecule (a), wavefunctions Ψ_b^H and Ψ_c^H that retain the characteristics of Ψ_a^H can be identified. E_f lies between the HOMO and LUMO levels of the isolated extended molecule (b). Knowing the relative distance between E_f and the energy E_b^H of Ψ_b^H , the relative distance between E_f and the energy E_c^H of Ψ_c^H can be obtained for any given E_b . The two distances should be equal for a sufficiently large extended molecule.	90
6.5	(a) Real space contour plot of the HOMO wavefunction Ψ_a^H of the isolated molecule shown in Figure 6.4a. (b) Real space contour plot of that wavefunction Ψ_b^H of the isolated extended molecule (Figure 6.4b) which most resembles Ψ_a^H . Ψ_b^H is identified by visual inspection of wavefunctions of the isolated extended molecule.	91
6.6	Identifying that wavefunction of the extended molecule coupled to contacts which most resembles the HOMO wavefunction of the isolated molecule. (a) Absolute value of the inner product of Ψ_b^H (see Figure 6.5) with all wavefunctions (eigenvectors) Ψ_c of the matrix $F + \Sigma_1 + \Sigma_2$ (this matrix represents the extended molecule coupled to contacts). Of all the wavevectors Ψ_c , one wavevector Ψ_c^H maximizes the inner product, implying that Ψ_c^H most resembles Ψ_b^H (as shown in (b) above), which in turn is similar to the HOMO wavefunction Ψ_a^H of the isolated molecule (see Figure 6.5). This inner product method of identification is unambiguous because the eigenvectors Ψ_b^H and Ψ_c^H have the same number of components, unlike Ψ_a^H , which is defined on a smaller Hilbert space. The inner product method is necessary because visual identification in real space (like the one shown in Figure 6.5) of wavefunctions of an open system is currently not feasible. .	92
6.7	Schematic picture of a quantum point contact (a chain of six gold atoms) with twelve contact atoms included in the extended molecule.	95
6.8	Transmission of a quantum point contact as a function of energy. Blue line is a self-consistent calculation with $\Sigma(E_f)$, while the magenta line is a non self-consistent calculation (Fock matrix obtained from the $\Sigma(E_f)$ self-consistent calculation) using $\Sigma(E)$. The Fermi energy E_f is at -5.1 eV. The transmission is close to one over a range of energies around E_f , giving rise to a quantized conductance of $2e^2/h$ as expected (Figure 6.9). The close agreement with $\Sigma(E)$ calculation reemphasizes the validity of the $\Sigma(E_f)$ approximation (also see Figure 5.4).	96

Figure	Page
6.9 I-V characteristic of a quantum point contact (compare with Figure 5.7a). The close agreement with the perfect quantum conductance $2e^2/h = 80 \mu A/V$ provides support for the accuracy of the calculation of the self-energy matrices $\Sigma_{1,2}(E_f)$	96
6.10 Schematic picture of the Xylyl Dithiol extended molecule. For an explanation of the color scheme see the caption to Figure 6.1.	98
6.11 I-V characteristic of Xylyl Dithiol. Small asymmetry in the current is due to small numerical asymmetries in the computation of the self-energy matrices Σ_1 and Σ_2 . The polynomial fit smoothens the data thereby reducing noise in the calculation of the derivative dI/dV (see Figure 6.12).	98
6.12 Conductance of Xylyl Dithiol as a function of applied voltage. The conductance is calculated by numerically differentiating the polynomial fit to the I-V (see Figure 6.11). Symbols represent experimental data. To obtain a fit to the experiment, all calculated conductance values were scaled down by one thousand. This orders of magnitude discrepancy in the theoretical and experimental conductance values is well known in the literature but no satisfactory explanation exists yet, at least for STM experiments.	99
6.13 Schematic picture of a Xylyl Dithiol molecule asymmetrically coupled to two contacts. The strongly coupled contact is modeled using the self-energy matrix $\Sigma_1(E_f)$, while the weakly coupled contact is modeled using an artificial small ($50 meV$) broadening on the sulfur and hydrogen orbitals on the right side.	100
6.14 I-V and G-V characteristics of a Xylyl Dithiol molecule asymmetrically coupled to two contacts. (a) I-V characteristic, and (b) G-V characteristic obtained by numerically differentiating the polynomial fit to the I-V. When a positive voltage is applied to the strongly coupled contact, current is suppressed due to charging effects.	101
6.15 Origin of I-V asymmetry lies in charging effect. One side (substrate) is strongly coupled to the molecule, while the other is weakly coupled. Although the same HOMO level is crossed by the contact electrochemical potentials μ_1 and μ_2 , for positive substrate voltage (b), the HOMO level is emptied by the strongly coupled contact, which positively charges the molecule and shifts the energy levels down. Such a shift, not present for negative substrate voltage (a), postpones the onset of conduction and effectively stretches out the voltage axis in the direction of positive substrate voltage.	102

Figure	Page
C.1 A 258 gold atom cluster used to obtain the couplings between nearest neighbor three-atom unit cells. Using this cluster ensures that we account for third nearest neighbor atom (out of plane) and fifth nearest neighbor atom (in plane) coupling. Ideally, we need about five to six near neighbors for an accurate calculation of the surface Green's function with the LANL2DZ basis set.	118
C.2 Gold surface density of states (LANL2DZ basis). The three-atom unit cell calculation accounts for many more near neighbor atoms as compared to the one-atom unit cell calculation, and the results for the two cases are qualitatively similar, but differ quantitatively. This quantitative difference is expected in light of the fact that the LANL2DZ basis functions are relatively delocalized.	119

ABSTRACT

Damle, Prashant Subhash. Ph.D., Purdue University, May, 2003. Nanoscale device modeling: from MOSFETs to molecules. Major Professor: Supriyo Datta.

This thesis presents a rigorous yet practical approach to model quantum transport in nanoscale electronic devices. As conventional metal oxide semiconductor devices shrink below the one hundred nanometer regime, quantum mechanical effects are beginning to play an increasingly important role in their performance. At the same time, demonstration of molecular switches has generated considerable interest in the emerging field of molecular electronics. Understanding electronic transport in nano-devices and molecules using rigorous physics based models is critical in order to design and fabricate such structures. Our approach is based on the non-equilibrium Green's function (NEGF) formalism which is rapidly gaining acceptance as the method of choice to treat quantum transport in nanostructures. We use our method to treat a few nanoscale devices of current interest, namely, a dual-gate silicon nanotransistor (effective mass model), a three terminal molecular device (semi-empirical atomic orbital model) and two terminal molecular wires (rigorous *ab initio* atomic orbital model). The results provide useful insights into the underlying physics in these devices. Several important features like charge transfer, self-consistent band lineup, I-V characteristics, voltage drop etc. are analyzed and explained.

1. INTRODUCTION

This thesis presents a rigorous yet practical approach to model quantum transport in nanoscale electronic devices. As conventional metal oxide semiconductor devices shrink below the one hundred nanometer regime, quantum mechanical effects are beginning to play an increasingly important role in their performance [1]. At the same time, demonstration of molecular switches has generated considerable interest in the emerging field of molecular electronics [2, 3, 4]. Understanding electronic transport in nano-devices and molecules using rigorous physics based models is critical in order to design and fabricate such structures.

Nanometer scale dimensions (1-100 nm) are comparable to (if not smaller than) the mean free path (~ 100 nm) as well as the de Broglie wavelength (~ 10 nm) of electrons in typical semiconductors and metals. Hence standard macroscopic models (for example, the drift diffusion model of current flow) which include scattering effects using lumped parameters (like mobility) and neglect quantum effects (like the wave nature of electrons) are not applicable to nanostructures, and new models and methods need to be developed.

In this chapter we discuss the self-consistent field approach used to model nano-devices (Section 1.1), followed by an overall self-consistent procedure (Section 1.2) and an outline of this dissertation (Section 1.3).

1.1 The Self-consistent Field Method

Figure 1.1a shows a schematic picture of a nanoscale device coupled to two contacts. The device is described using a Hamiltonian H . The contacts are identified by their respective Fermi levels μ_1 and μ_2 . Upon coupling to the contacts, the discrete energy levels of the isolated device broaden into a continuous density of states. The

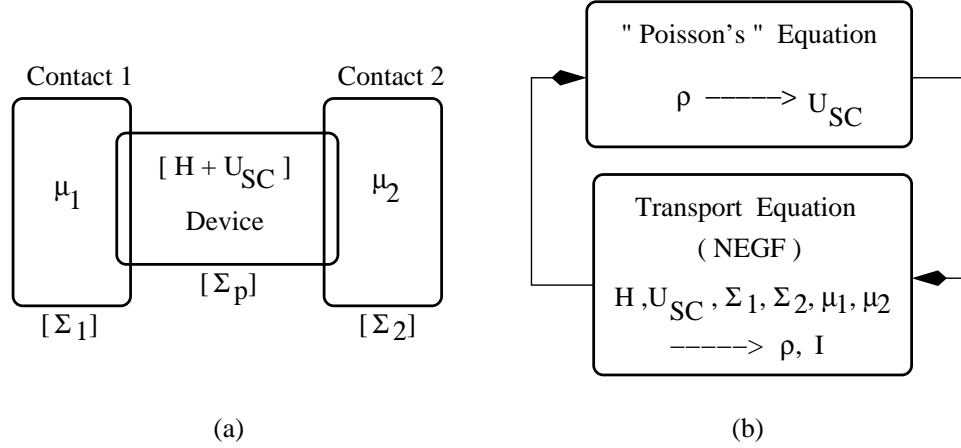


Fig. 1.1. Schematic view of a device coupled to contacts and the solution procedure.

H is the device Hamiltonian and U_{SC} is the self-consistent potential. The contact Fermi levels μ_1 and μ_2 are equal at equilibrium and differ by an amount qV when a bias of V volts is applied across the device. The effect of the contacts on the device is *exactly* accounted for by the contact self-energy matrices Σ_1 and Σ_2 . The self-energy matrix Σ_p may be used to model non-coherent scattering processes inside the device. The word ‘‘Poisson’s’’ is in quotes as a reminder that more sophisticated schemes like Hartree-Fock (HF) or density functional theory (DFT) may also be used to obtain the self-consistent potential. The Non-equilibrium Green’s Function (NEGF) formalism is used to obtain the electron density ρ and current I .

level broadening is related to the lifetime of an electron in the device - electrons can, and do, flow in and out of the device. This effect of the contact coupling on the device can be *exactly* accounted for using self-energy matrices Σ_1 and Σ_2 (see Section 2.4 for a more detailed discussion on the self-energy matrices). Although the self-energy matrices describe the effect of the large contacts, they are still of the same size as the device Hamiltonian. Incoherent scattering processes inside the device may be modeled using another self-energy matrix Σ_p , but in this dissertation, we will confine our attention to coherent or ballistic transport.

When the device is coupled to contacts, there is some charge transferred in or out of the device, which gives rise to a potential $U_{SC}(\vec{r})$. This potential in turn modifies the electronic charge density $\rho(\vec{r})$ inside the device and the process continues till things settle down and U_{SC} and ρ attain some self-consistent values. This is the

equilibrium situation when the occupancy or otherwise of the electronic states inside the device and the contacts is dictated by a single Fermi level E_f ($\mu_1 = \mu_2 = E_f$). If a voltage drop of V volts is now maintained across the device, the contact Fermi levels separate by an amount equal to qV (q is the electronic charge, $q = 1.6 \cdot 10^{-19}$ Coulombs). One contact now tries to pump electrons into the device while the other contact tries to pull them out of the device, and a current flows. In steady state U_{SC} and ρ again attain some self-consistent values. The device is now said to be in a state of non-equilibrium.

In order to model such equilibrium as well as non-equilibrium situations, we need a procedure like the one shown schematically in Fig. 1.1b. We start with some initial guess for U_{SC} . Knowing H , U_{SC} , $\Sigma_{1,2,p}$, and $\mu_{1,2}$, we can compute ρ using the well-defined relations of the non-equilibrium Green's function (NEGF) formalism (Chapter 2) which is rapidly gaining acceptance as the method of choice to treat quantum transport in nanostructures. Knowing ρ we obtain U_{SC} by solving the Poisson's equation:

$$\vec{\nabla} \cdot (\epsilon \vec{\nabla} U_{SC}) = -q^2 \rho + \text{other charges} \quad (1.1)$$

Other charges in the above equation refer to ionized impurities in case of semiconductor devices, and nuclear charges in case of molecular devices. If we wish to model the electron-electron interactions (significant in very small devices) we need to augment the solution to Poisson's equation using more sophisticated methods like Hartree-Fock (HF) or density functional theory (DFT). Once we have U_{SC} , we recalculate ρ using the NEGF equations. We iterate the two steps above (ρ to U_{SC} and U_{SC} to ρ) until convergence is reached.

The solution to Poisson's equation is a well-established technique in the field of device simulation [5, 6], and HF and DFT methods are standard in quantum chemistry calculations [7, 8]. For this reason, we will not discuss these methods in any detail in this dissertation. Rigorous derivations of the NEGF formalism are available in the literature, [9, 10, 11], so we will only present a brief intuitive derivation of the NEGF equations (Chapter 2).

One small point before we move on. In the discussion above, we assumed a real space representation and talked about the electron density $\rho(\vec{r})$. A more powerful point of view is that the electron density *is the same* as the *density matrix* ρ expressed in a real space basis. In some cases (for example, molecular devices) it may be more convenient to leave the real space basis and work in an atomic orbital basis (Chapters 4, 5, 6) and in such cases the concept of a density matrix is more useful than the relatively limited concept of electron density. A good introductory discussion about the density matrix can be found in [12]. Without going into details, here we will simply state that obtaining the density matrix is the central problem in quantum transport theory, once known, the density matrix can be used to compute all quantities of interest like the electron density, current etc. [12].

1.2 Step-by-step Solution Procedure

Based on the above discussion, given a nanoscale device in which quantum phenomena are expected to play an important role, the step-by-step procedure we use to treat electronic transport in such a nano-device is as follows:

1. Given a particular type of device and contacts, we first make a choice of an appropriate basis (or representation) in which all the matrices are expressed. For example, we choose a real space discrete lattice basis set for a MOSFET (Chapter 3) and an atomic orbital basis set for molecular devices (Chapters 4,5,6).
2. The next step is to write down a suitable Hamiltonian matrix for the device in the chosen basis. For a MOSFET, we use the effective mass approximation to obtain the Hamiltonian whereas for a molecule we use either a semi-empirical Hamiltonian or, more rigorously, an ab initio Hamiltonian obtained using a standard quantum chemistry software.
3. We then compute the contact self-energy functions Σ_1 and Σ_2 (Section 2.4). The details of this calculation for a dual-gate MOSFET and for a molecule

coupled to contacts are given in chapters 3 and 5 respectively. Non-coherent scattering processes in the device may in principle be modeled (see Chapter 8 of [11]) by computing an appropriate self-energy Σ_p . In this dissertation, however, we mainly focus on ballistic devices ($\Sigma_p = 0$).

4. To begin the self-consistent procedure, we then start with a guess for the self-consistent potential U_{SC} in the chosen basis. In the case of a MOSFET, this guess may be obtained from a standard drift-diffusion simulator. For a molecule, the guess may be obtained by using a standard quantum chemistry software to do a self-consistent solution for an *isolated* molecule without the contacts.
5. Using all the information gathered in the steps above, and the contact Fermi levels μ_1 and μ_2 (see Figure 1.1) we use the NEGF equations (Chapter 2) to obtain the density matrix ρ .
6. Using this density matrix ρ , we solve either the Poisson's equation (MOSFET) or the HF/DFT equations (molecule) to obtain a new self-consistent potential U_{SC} .
7. We iterate steps 5 and 6 till both U_{SC} and ρ converge to an acceptable accuracy.
8. We then use the converged ρ to obtain the self-consistent electron density, current etc.

1.3 Outline Of This Dissertation

An intuitive introductory discussion of the NEGF formalism can be found in Chapter 2. This basic formalism can be used in conjunction with any type of Hamiltonian, ab initio or semi-empirical. Chapters 3, 4, 5 and 6 illustrate the use of the procedure listed in the previous section as applied to various nano-devices of current interest: (1) an effective mass treatment of a ten-nanometer scale dual-gate MOSFET (Chapter 3), (2) a semi-empirical atomic orbital based treatment of a molecular field

effect transistor (Chapter 4) and (3) a much more rigorous ab initio atomic orbital based treatment of different kinds of two terminal molecular wires (Chapters 5, 6). Other examples using different semi-empirical Hamiltonians can be found in [13]. The following paragraphs provide a brief glimpse into the contents of chapters 3, 4, 5 and 6.

Chapter 3 may be viewed as a follow-up to [12] which introduced the central concepts in quantum transport theory using an $n^{++} - n^+ - n^{++}$ resistor as example. Here we present a more complex example involving a dual-gate MOSFET which is widely viewed as a possible alternative to bulk MOSFETs that can be scaled to nanometer dimensions. We hope that combined with [12], this chapter will bring a reader with little or no background in the NEGF formalism upto a level where he or she can simulate devices like a MOSFET using NEGF. The results presented in this chapter were generated using a PC with a Pentium 3 processor. We will be glad to share our MATLAB codes with interested readers. This work is published in [14].

Following the increasing interest in the emerging field of molecular electronics, Chapter 4 addresses the question of whether a “rigid molecule” (one which does not deform in an external field) used as the conducting channel in a standard three-terminal MOSFET configuration can offer any performance advantage relative to a standard silicon MOSFET. We show that even for extremely small channel lengths (about 1 nm), a “well-tempered” molecular FET demands much the same electrostatic considerations as a “well-tempered” conventional MOSFET. In other words, we show that just as in a conventional MOSFET, the gate oxide thickness needs to be much smaller than the channel length (length of the molecule) for the gate control to be effective. Furthermore, we show that a rigid molecule with metallic source and drain contacts has a temperature independent subthreshold slope much larger than 60 mV/decade, because the metal-induced gap states in the channel prevent it from turning off abruptly. However, this disadvantage can be overcome by using semiconductor contacts because of their band-limited nature. This work is published in [15].

Following the recent success of many experimental groups in measuring the current-voltage (I-V) characteristics of individual molecules, Chapters 5 and 6 present a rigorous and computationally efficient method to do a first-principles analysis of molecular wires connected to contacts. Standard quantum chemistry software (Gaussian '98 [16]) is used to calculate the self-consistent potential using density functional theory (DFT). Such close coupling to standard quantum chemistry software not only makes the procedure simple to implement but also makes the relation between the I-V characteristics and the chemistry of the molecule more obvious. We use our method to interpolate between two extreme examples of transport through a molecular wire connected to gold (111) contacts: band conduction in a metallic (gold) nanowire, and resonant conduction through broadened, quasidiscrete levels of benzene ring based molecules like Phenyl Dithiol and Xylyl Dithiol. We obtain several quantities of interest like I-V characteristic, electron density and voltage drop along the molecule. Chapter 6 basically extends the treatment in Chapter 5 to include a chunk of each contact as part of the molecule. Good agreement between theoretical and experimental conductance-voltage (G-V) curves is observed. Part of this work (Chapter 5) is published [17, 18], while the other part (Chapter 6) is to be submitted for publication.

Each of appendices A, B, and C discusses one of the three simplifying approximations made while modeling the gold contacts in Chapters 5 and 6. Appendix D has a listing of the modifications we made to the source code of Gaussian '98 in order to model open systems under bias.

2. THE NEGF FORMALISM

This chapter contains an introductory discussion of the non-equilibrium Green's function (NEGF) formalism. We will not derive any of the NEGF formalism equations, since rigorous derivations are available in the literature (See, for example, [9, 10] and references therein. These references deal with bulk conductors and do not explicitly treat the contacts. The use of self-energy functions to represent the contacts can be traced back to [19] and has been discussed in Chapter 3 of [11]). Instead we will try to present an intuitive derivation of the basic results.

There are several ways in which a device coupled to contacts can be driven into a non-equilibrium situation (see, for example, Chapter 3 of [20]) - by shining light on the device, by maintaining a temperature gradient across the device, by maintaining a potential gradient across the device etc. In this dissertation we will confine our attention to the case where a finite current flows through the device due to a potential gradient. The process of current flow involves a non-equilibrium situation where the different contacts (like the source and the drain in the case of a MOSFET) have different electrochemical potentials μ_1, μ_2 etc. For example if a positive voltage V is applied externally to the drain with respect to the source, then the drain has an electrochemical potential lower than that of the source by qV : $\mu_1 - \mu_2 = qV$. The source and drain contacts thus have different Fermi functions and each seeks to bring the device into equilibrium with itself. The source keeps pumping electrons in hoping to establish equilibrium. But equilibrium is never achieved as the drain keeps pulling electrons out in its bid to establish equilibrium. The device is thus forced into a balancing act between two reservoirs with different agendas which sends it into a non-equilibrium state intermediate between what the source would like to see and what the drain would like to see. To describe this balancing process we need a kinetic equation that keeps track of the input and output from each of the reservoirs.

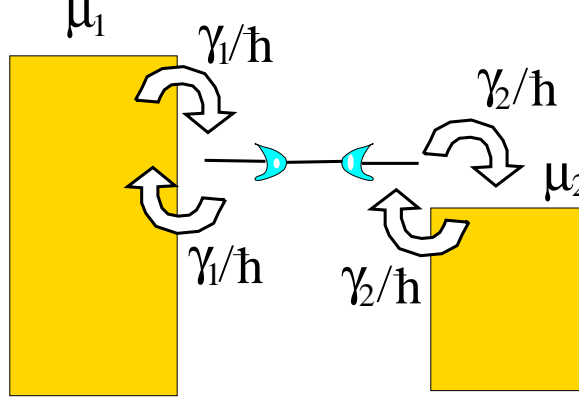


Fig. 2.1. Current flow through a single energy level. One contact keeps filling it up and the other keeps emptying it causing a net current flow from one contact to the other.

2.1 Current Flow Through A Single Discrete Energy Level

The balancing process described above is easy to see if we consider a simple one level system, biased such that its energy lies between the electrochemical potentials in the two contacts (Fig. 2.1). An electron in this level can escape into contacts 1 and 2 at a rate of γ_1/\hbar and γ_2/\hbar respectively. If the level were in equilibrium with contact 1 then the number of electrons occupying the level would be given by

$$N_1 = 2 \text{ (for spin) } f_1(E) \quad (2.1)$$

where f_1 is the Fermi function with Fermi level μ_1 of contact 1:

$$f_1(E) = \frac{1}{1 + \exp\left(\frac{E - \mu_1}{k_B T}\right)} \quad (2.2)$$

Similarly if the level were in equilibrium with contact 2 the number would be

$$N_2 = 2 \text{ (for spin) } f_2(E) \quad (2.3)$$

where

$$f_2(E) = \frac{1}{1 + \exp\left(\frac{E - \mu_2}{k_B T}\right)} \quad (2.4)$$

Under non-equilibrium conditions the number of electrons N will be somewhere in between N_1 and N_2 . To determine this number we write a steady state *kinetic equation* that equates the net current at the left junction

$$I_L = \frac{q\gamma_1}{\hbar} (N_1 - N) \quad (2.5)$$

to the net current at the right junction

$$I_R = \frac{q\gamma_2}{\hbar} (N - N_2) \quad (2.6)$$

Setting $I_L = I_R$, we obtain

$$N = 2 \frac{\gamma_1 f_1(E) + \gamma_2 f_2(E)}{\gamma_1 + \gamma_2} \quad (2.7)$$

so that the current is

$$I = I_L = I_R = \frac{2q}{\hbar} \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} [f_1(E) - f_2(E)] \quad (2.8)$$

Eq. 2.8 follows very simply from elementary arguments, but it serves to illustrate a basic fact about the process of current flow. No current will flow if $f_1(E) = f_2(E)$. A level that is way below both electrochemical potentials μ_1 and μ_2 will have $f_1(E) = f_2(E) = 1$ and will not contribute to the current, just like a level that is way above both potentials μ_1 and μ_2 and has $f_1(E) = f_2(E) = 0$. It is only when the level lies within a few $k_B T$ of the potentials μ_1 and μ_2 that we have $f_1(E) \neq f_2(E)$ and a current flows. Current flow is thus the result of the *difference in opinion* between the contacts. One contact would like to see more electrons (than N) occupy the level and keeps pumping them in, while the other would like to see fewer than N electrons and keeps pulling them out.

Given the level (ϵ_0), broadening (γ_1, γ_2) and the electrochemical potentials μ_1 and μ_2 of the two contacts, we can solve Eq. 2.8 for the current I . But we want to include charging effects in the calculations. Therefore, we add a potential U_{SC} due to the change in the number of electrons from the equilibrium value $2f_0(E_f)$ (f_0 is the Fermi function with the equilibrium Fermi level E_f , see Equation 2.2):

$$U_{SC} = U [N - 2f_0(E_f)] \quad (2.9)$$

where U is the charging energy per electron, which is the amount by which the level floats up (down) if one electronic charge is added (removed). We then let the level ϵ float up or down by this potential:

$$\epsilon = \epsilon_0 + U_{SC} \quad (2.10)$$

Since the potential depends on the number of electrons, we need to calculate the potential using a self-consistent procedure similar to the one shown in Fig. 1.1b. Once the converged solution is obtained, the current is calculated from Eq. 2.8.

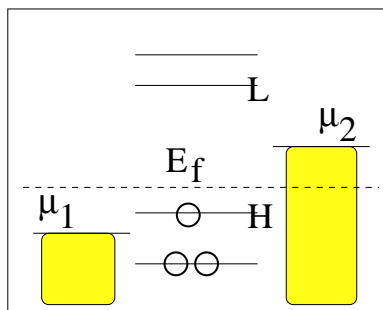


Fig. 2.2. Schematic description of resonant current conduction through discrete molecular levels. H and L denote the highest occupied and the lowest unoccupied molecular levels respectively. An electron in any of these levels has a finite lifetime before it escapes into one of the contacts. Onset of current corresponds to crossing of a molecular level by a contact chemical potential under bias.

The simple discrete one-level model just described captures a lot of the essential physics of conduction through single molecules coupled to contacts. We may think of the level ϵ_0 as a molecular level that is closest to the Fermi energy E_f (Figure 2.2). Typically the Fermi energy lies in the energy gap between the highest occupied molecular orbital (HOMO) level and the lowest unoccupied molecular orbital (LUMO) level, and the molecular conductance gap is then determined by the distance $|E_f - \epsilon_0|$, with ϵ_0 corresponding to the HOMO or the LUMO level, whichever is closest to E_f [21, 22]. At low applied voltage, all molecular levels are either full or empty, and no current flows. Onset of current corresponds to crossing of a molecular level by one of the contact electrochemical potentials. For example, see the molecular current-voltage characteristics shown in Figures 4.5c,d, 5.7b and 6.11.

2.2 Current Flow Through A Single Broadened Energy Level

In Section 2.1, we treated the level ϵ as discrete, ignoring the broadening $\gamma = \gamma_1 + \gamma_2$ that arises due to the coupling with the contacts. To take the broadening into account we may replace the discrete level with a Lorentzian density of states $D(E)$:

$$D(E) = \frac{1}{2\pi} \frac{\gamma}{(E - \epsilon)^2 + (\gamma/2)^2} \quad (2.11)$$

and modify Eqs. 2.7, 2.8 for N and I to include an integration over energy:

$$N = 2 \int_{-\infty}^{\infty} dE D(E) \frac{\gamma_1 f_1(E) + \gamma_2 f_2(E)}{\gamma_1 + \gamma_2} \quad (2.12)$$

$$I = \frac{2q}{\hbar} \int_{-\infty}^{\infty} dE D(E) \frac{\gamma_1 \gamma_2}{\gamma_1 + \gamma_2} [f_1(E) - f_2(E)] \quad (2.13)$$

The charging effect is included as before by letting the center ϵ , of the Lorentzian density of states, float up or down:

$$\epsilon = \epsilon_0 + U_{SC} \quad (2.14)$$

$$U_{SC} = U(N - N_0) \quad (2.15)$$

where the number of electrons at equilibrium is approximated by $N_0 = 2f_0(E_f)$. As before, now we can implement the self-consistent procedure shown in Figure 1.1b. Once convergence is obtained, we can calculate the current using Equation 2.13.

2.3 NEGF Equations As A Multi-level Generalization

The one-level model just described includes the three basic factors that influence nanoscale conduction, namely, $E_f - \epsilon_0$, $\gamma_{1,2}$ and U . However, real nanoscale devices typically have multiple levels that often broaden and overlap. In general we need a formalism that can do justice to multiple levels with arbitrary broadenings and overlaps. The general multi-level Non-Equilibrium Green's Function (NEGF) formalism outlined in this section does just that. A much more detailed discussion of the NEGF formalism may be obtained in [11].

In Section 2.2 we obtained equations for the number of electrons, N and the current, I for a one-level model. It is useful to rewrite these equations in terms of the Green's function $G(E)$ which is defined as follows:

$$G(E) = \left(E - \epsilon + i \frac{\gamma_1 + \gamma_2}{2} \right)^{-1} \quad (2.16)$$

The density of states $D(E)$ is proportional to the spectral function $A(E)$ defined as:

$$A(E) = -2\text{Im} \{G(E)\} \quad (2.17)$$

$$D(E) = \frac{A(E)}{2\pi} \quad (2.18)$$

while the number of electrons, N and the current, I can be written as:

$$N = \frac{2}{2\pi} \int_{-\infty}^{\infty} dE \left(|G(E)|^2 \gamma_1 f_1(E) + |G(E)|^2 \gamma_2 f_2(E) \right) \quad (2.19)$$

$$I = \frac{2q}{h} \int_{-\infty}^{\infty} dE \gamma_1 \gamma_2 |G(E)|^2 [f_1(E) - f_2(E)] \quad (2.20)$$

In the NEGF formalism the single energy level ϵ_0 is replaced by a Hamiltonian matrix $[H]$ while the broadenings $\gamma_{1,2}$ are replaced by complex energy-dependent self-energy matrices $[\Sigma_{1,2}(E)]$ so that the Green's function becomes a matrix given by¹:

$$G(E) = (ES - H - \Sigma_1 - \Sigma_2)^{-1} \quad (2.21)$$

where S is the identity matrix of the same size as the other matrices and the broadening matrices $\Gamma_{1,2}$ are defined as the imaginary (more correctly as the anti-Hermitian) parts of $\Sigma_{1,2}$:

$$\Gamma_{1,2} = i \left(\Sigma_{1,2} - \Sigma_{1,2}^\dagger \right) \quad (2.22)$$

The spectral function is the anti-Hermitian part of the Green's function:

$$A(E) = i \left[G(E) - G^\dagger(E) \right] \quad (2.23)$$

¹In general, scattering processes inside the device can be included in the NEGF formalism through another self-energy matrix Σ_p (see Fig. 1.1 and the discussion in Section 2.4). In this dissertation, we will neglect all scattering inside the device ($\Sigma_p = 0$).

from which the density of states $D(E)$ can be calculated by taking the trace:

$$D(E) = \frac{\text{Tr}(AS)}{2\pi} \quad (2.24)$$

The density matrix $[\rho]$ is given by (compare with Eq. 2.19):

$$\rho = \frac{1}{2\pi} \int_{-\infty}^{\infty} [f_1(E)G\Gamma_1G^\dagger + f_2(E)G\Gamma_2G^\dagger] dE \quad (2.25)$$

from which the total number of electrons, N can be calculated by taking a trace:

$$N = \text{Tr}(\rho S) \quad (2.26)$$

The current is given by (compare with Eq. 2.20):

$$I = \frac{2q}{h} \int_{-\infty}^{\infty} [\text{Tr}(\Gamma_1G\Gamma_2G^\dagger)(f_1(E) - f_2(E))] dE \quad (2.27)$$

Equations 2.21 through 2.27 constitute the basic equations of the NEGF formalism which have to be solved self consistently (Fig. 1.1) with a suitable scheme to calculate the self-consistent potential matrix $[U_{SC}]$ (compare with Eq. 2.10):

$$H \longrightarrow H + U_{SC} \quad (2.28)$$

where U_{SC} is an appropriate functional of the density matrix ρ :

$$U_{SC} = F(\rho) \quad (2.29)$$

This self-consistent procedure is essentially the same as that used for the one level model, except that scalar quantities have been replaced by matrices:

$$\epsilon_0 \rightarrow [H] \quad (2.30)$$

$$\gamma \rightarrow [\Gamma], [\Sigma] \quad (2.31)$$

$$N \rightarrow [\rho] \quad (2.32)$$

$$U_{SC} \rightarrow [U_{SC}] \quad (2.33)$$

The size of all these matrices is $(n \times n)$, n being the number of basis functions used to describe the *device*. Note that the self-energy matrices $\Sigma_{1,2}$ are of this size although they represent the effect of infinitely large contacts. In the remainder of this dissertation, we will describe the procedure used to evaluate the Hamiltonian matrix H , the self-energy matrices $\Sigma_{1,2}$ and the functional “ F ” used to evaluate the self-consistent potential U_{SC} (see Eq. 2.29) using specific examples. But the point to note is that once we know how to evaluate these matrices, Eqs. 2.21 through 2.29 can be used in a straightforward manner to calculate the current.

2.4 Origin Of The Contact Self-energy Matrix

The concept of self-energy is used in many-body physics to describe non-coherent electron-electron and electron-phonon interactions. We could do the same in principle and use a self-energy function Σ_p to describe the effect of non-coherent interactions of the device with its surroundings. In addition to Σ_p , we can use self-energy functions Σ_1 and Σ_2 to describe the interactions of the device with the two contacts respectively.

When an isolated device is connected to contacts, its discrete energy levels broaden into a continuous density of states. This broadening may be interpreted as the inverse residence time of an electron in the device. In other words, the device is not isolated anymore, and electrons (and holes) in the device escape into the contacts in some finite time. In addition to this broadening of levels, the energy levels of the isolated device may also shift a little bit upon coupling to the contacts. The overall effect of shift and broadening of energy levels may be regarded as a complex shift in energy and described by non-Hermitian self-energy matrices $\Sigma_1(E)$ and $\Sigma_2(E)$ corresponding to contacts 1 and 2 respectively. The important point to note is that although the self-energy matrices represent the effect of a very large contact, they are of the *same* size as the device Hamiltonian.

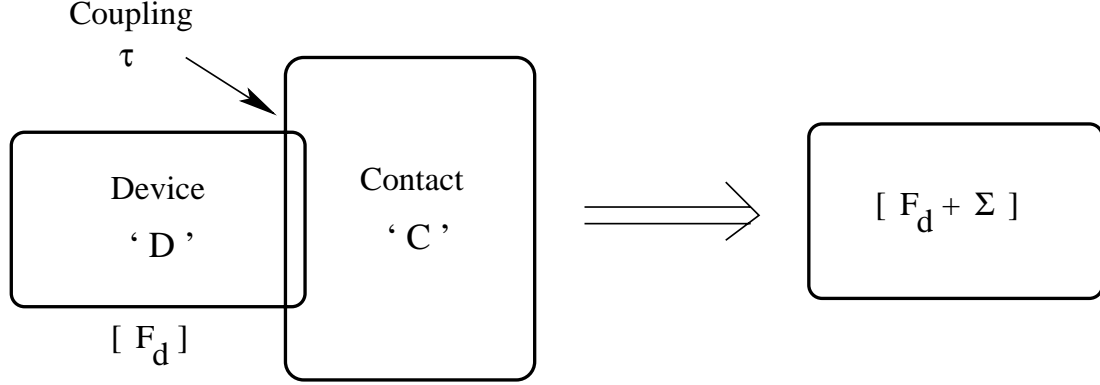


Fig. 2.3. The complete system consisting of device and contact may be partitioned into a device subspace and a contact subspace. The effect of the infinitely large contact is then incorporated through a self-energy matrix Σ which is of the *same* size as the device Hamiltonian.

The self-energy matrices arise formally out of partitioning ² an infinite system consisting of the device and the contacts into a device subspace and a contact subspace [19, 11]. Fig. 2.3 shows a device D coupled to a semi-infinite contact C . The retarded Green's function of the complete system is given by

$$G(E) = \left[(E + i0^+)S - (H + U_{SC}) \right]^{-1} \quad (2.34)$$

where H is the Hamiltonian of the complete system and U_{SC} is the self-consistent potential discussed in the previous section. From this point onwards, we will use the symbol F (short for the term “Fock matrix” commonly used in quantum chemistry literature) to represent the sum of H and U_{SC} :

$$F \equiv H + U_{SC} \quad (2.35)$$

²As pointed out in [19], localized basis functions are needed in order for such partitioning to make good sense. Basis functions with a large spatial extent give rise to some “fuzziness” in the partitioning. For example, the non-orthogonal atomic orbital basis functions used in ab initio quantum chemistry calculations (Chapter 5) are quite delocalized and lead to some unavoidable errors in the calculation of the density matrix, as shown in Section 2.5 below.

The device part (subscripted d) of $G(E)$ is then obtained by partitioning

$$G = ((E + i0^+)S - F)^{-1} = \begin{bmatrix} (E + i0^+)S_{dd} - F_{dd} & \vdots & \tau \\ \dots & & \dots \\ \tau^\dagger & \vdots & C \end{bmatrix}^{-1} \quad (2.36)$$

where τ is the coupling between the device and the contact and C is the contact part of the overall Green's function. Using matrix algebra, it is straightforward to show that the device part of G is given by

$$G_{dd} = [ES_{dd} - F_{dd} - \Sigma(E)]^{-1} \quad (2.37)$$

with

$$\Sigma(E) = \tau C^{-1} \tau^\dagger \quad (2.38)$$

In equation 2.37 we have dropped the usual small imaginary number “ $i0^+$ ” (see Equation 2.34) because the self-energy matrix is complex with a finite imaginary part.

Since the size of C is effectively infinite, it may seem at this point that we have not gained much from this partitioning. However, it is important to note that the coupling τ is non-zero only at the device-contact interface (or the ‘surface’), and only the surface part of C^{-1} matters. This surface Green's function g is of a finite size: if the device Hamiltonian is $(N \times N)$ and the non-zero part of coupling τ is $(N \times n)$, then the size of g is $(n \times n)$. The self-energy is thus given by

$$\Sigma(E) = \tau g \tau^\dagger \quad (2.39)$$

The surface Green's function g only depends on the contact material and geometry and is independent of the device. Given a particular type of contact, there are recursive techniques (See, for example, [23, 24, 25, 26] and references therein) to compute g . We will discuss these techniques in more detail in the following chapters using specific examples.

From here on we will drop the subscripts d for convenience. The above discussion may be easily extended when the device is coupled to *two* semi-infinite contacts. The Green's function for the device (including the effect of contacts) is then given by

$$G(E) = [ES - F - \Sigma_1(E) - \Sigma_2(E)]^{-1} \quad (2.40)$$

with

$$\Sigma_{1,2}(E) = \tau_{1,2} g_{1,2} \tau_{1,2}^\dagger$$

Thus we have shown how to incorporate the effect of infinitely large contacts *exactly* into the device Green's function.

2.5 Non-orthogonal Basis

The matrices appearing above depend on the basis functions that we use. In order to model molecular devices, we need to use non-orthogonal atomic orbital basis functions and the matrix equations 2.21 through 2.29 are still valid, except that the elements of the matrix $[S]$ in Eq. 2.21 represents the overlap of the basis functions $\phi_m(\bar{r})$:

$$S_{mn} = \int d^3r \phi_m^*(\bar{r}) \phi_n(\bar{r}) \quad (2.41)$$

For orthogonal bases, $S_{mn} = \delta_{mn}$ so that S is the identity matrix as stated earlier.

To see the validity of Eqs. 2.21-2.27 in a non-orthogonal representation we transform H , Σ into the orthogonal representation \widetilde{H} , $\widetilde{\Sigma}$ [7]:

$$\begin{aligned} \widetilde{H} &= S^{-1/2} H S^{-1/2} \\ \widetilde{\Sigma} &= S^{-1/2} \Sigma S^{-1/2} \end{aligned} \quad (2.42)$$

$$(2.43)$$

and define a transformed Green's function

$$\widetilde{G} = (EI - \widetilde{H} - \widetilde{\Sigma})^{-1} \quad (2.44)$$

where I is the identity matrix. It is easy to show that:

$$\begin{aligned} G &= S^{-1/2} \widetilde{G} S^{-1/2} \\ \Sigma &= S^{1/2} \widetilde{\Sigma} S^{1/2} \\ G\Gamma_{1,2}G^\dagger &= S^{-1/2} [\widetilde{G}\widetilde{\Gamma}_{1,2}\widetilde{G}^\dagger] S^{-1/2} \\ \Gamma_1 G \Gamma_2 G^\dagger &= S^{1/2} [\widetilde{\Gamma}_1 \widetilde{G} \widetilde{\Gamma}_2 \widetilde{G}^\dagger] S^{-1/2} \\ \text{Tr}(\Gamma_1 G \Gamma_2 G^\dagger) &= \text{Tr}(\widetilde{\Gamma}_1 \widetilde{G} \widetilde{\Gamma}_2 \widetilde{G}^\dagger) \end{aligned} \quad (2.45)$$

The point is that if Eqs. 2.21-2.27 are valid in the orthogonal representation (denoted by the tilde \sim), then Eqs. 2.42-2.45 can be used to show that they will be valid for the non-orthogonal representation as well. The main point to note is that

$$\tilde{\rho} = S^{1/2} \rho S^{1/2} \quad (2.46)$$

This means that when calculating the number of electrons, N :

$$N = \text{Tr}(\tilde{\rho}) = \text{Tr}(\rho S) \quad (2.47)$$

we should multiply ρ by S before taking the trace. This can lead to small errors because we only evaluate the density matrix ρ in the molecule. To see this we write for the entire molecular-contact system (m : molecule, c : contact):

$$N = \text{Tr} \left(\begin{pmatrix} \rho & \rho_{mc} \\ \rho_{cm} & \rho_{cc} \end{pmatrix} \cdot \begin{pmatrix} S & S_{mc} \\ S_{cm} & S_{cc} \end{pmatrix} \right) \quad (2.48)$$

we can write

$$N(\text{molecule}) = \text{Tr}(\rho S) + \text{Tr}(\rho_{mc} S_{cm}) \quad (2.49)$$

where we ignore the second term entirely.

3. DUAL-GATE NANOTRANSISTOR: EFFECTIVE MASS MODEL

In this chapter, we illustrate the procedure given in Section 1.2 (see Fig. 1.1) using the example of a dual-gate nanoscale MOSFET. We present results for a ballistic MOSFET - incoherent scattering processes are ignored. Similar calculations on 10 nm scale ballistic dual-gate MOSFETs have been done recently [27, 28]. Simplified models have been used to account for scattering [12, 18, 29], whereas a proper description of scattering processes is still an area of intense research [30]. This chapter may be viewed as a follow-up to [12] where the central concepts of quantum transport theory were explained by using the NEGF formalism to simulate a simpler case of an $n^{++} - n^+ - n^{++}$ resistor. We hope that combined with [12], this and the previous chapter will bring a reader with little or no background in NEGF upto a level where he or she can feel confident about simulating nanoscale devices using NEGF.

A schematic view of a dual-gate MOSFET is shown in Fig. 3.1. We follow the solution procedure outlined in Section 1.2, where the NEGF equations (Chapter 2) are self-consistently solved with the Poisson's equation. The choice of a suitable basis is discussed in Section 3.1. Section 3.1 also explains how we set up the device Hamiltonian matrix H and the self-consistent potential matrix U_{SC} in this basis. Section 3.2 explains how we set up the source and drain self-energy matrices Σ_1 and Σ_2 in the basis discussed in Section 3.1. The self-energy matrices account for the coupling between the device and the contacts. A brief description of our Poisson's equation solution is contained in Section 3.3. Section 3.4 contains the simulation results for a 10 nm dual-gate MOSFET.

3.1 Hamiltonian

We use the effective mass approximation to obtain the device Hamiltonian. The first step is to identify a suitable basis set. The coordinate axes are shown in Fig. 3.1. The overall effective mass Hamiltonian is then written as the sum of a longitudinal (x-y) part H_L and a transverse (z) part H_T ($H + U_{SC} = H_L + H_T$), m^* being an appropriate effective mass:

$$H_L = -\frac{\hbar^2}{2m^*} \frac{\partial^2}{\partial x^2} - \frac{\hbar^2}{2m^*} \frac{\partial^2}{\partial y^2} + U_{SC}(x, y) \quad (3.1)$$

and

$$H_T = -\frac{\hbar^2}{2m^*} \frac{\partial^2}{\partial z^2} + U_{SC}(z) \quad (3.2)$$

We have added the self-consistent potential U_{SC} to the Hamiltonian for convenience. The procedure to obtain a suitable U_{SC} is discussed in section 3.3.

We assume that the width (z-direction) of the device is large compared to the other dimensions of the device. The MOSFET is thus essentially a 2-D charge sheet. For such devices with effectively infinite widths, we ignore the transverse confining potential $U_{SC}(z)$ and write the transverse eigenstates as plane waves:

$$\chi_{\vec{k}_z}(\vec{z}) = \frac{1}{\sqrt{W}} e^{i\vec{k}_z \cdot \vec{z}}$$

and

$$H_T \chi_{\vec{k}_z} = \epsilon_{\vec{k}_z} \chi_{\vec{k}_z}$$

where $\epsilon_{\vec{k}_z} = \hbar^2 k_z^2 / 2m^*$, and W is the device width.

We use a discrete lattice in real space for H_L . We assume a device structure as shown in Fig. 3.1. Let N_{px} and N_{py} be the number of grid points in the x and y directions respectively (see Fig. 3.2). Then the longitudinal Hamiltonian will be of

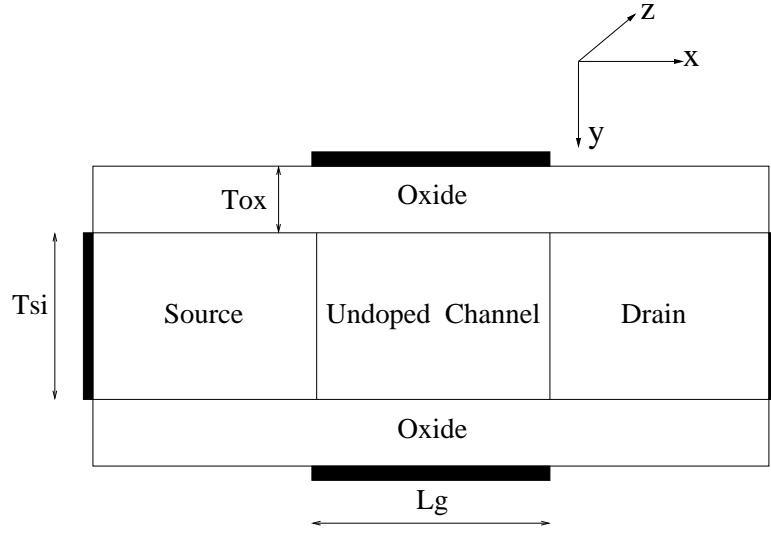


Fig. 3.1. Schematic view of a dual-gate MOSFET.

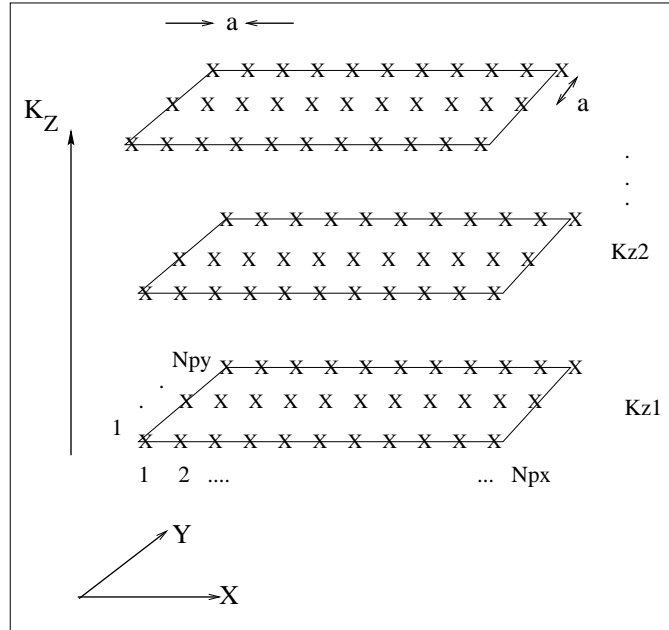


Fig. 3.2. Choice of basis: We use the eigenstate representation for the transverse (z) dimension but a discrete real space lattice for the longitudinal (x - y) dimensions. The coordinate axes are defined in Fig. 3.1.

size $N = N_{px}N_{py} \times N_{px}N_{py}$. Using the finite difference method to discretize Eq. 3.1 we see that the real space Hamiltonian matrix is

$$H_L = U_{SC} + \begin{bmatrix} \alpha & \beta & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ \beta^\dagger & \alpha & \beta & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & \beta^\dagger & \alpha & \beta & \cdot & \cdot & \cdot & 0 \\ & & & \cdot & \cdot & \cdot & & \\ & & & & \cdot & \cdot & \cdot & \\ \cdot & & & & & \beta^\dagger & \alpha & \beta \\ 0 & \cdot & \cdot & & & 0 & \beta^\dagger & \alpha \end{bmatrix}_{N \times N} \quad (3.3)$$

where U_{SC} is a diagonal matrix with $U_{SC}(n, n) = U_{SC}(x_i, y_j)$, and $n = j + (i - 1)N_{py}$ (Our node numbering scheme starts with node 1 as the (x_1, y_1) node, node 2 as the (x_1, y_2) node, etc. and ends with node N as the $(x_{N_{px}}, y_{N_{py}})$ node. α and β are matrices of size $N_{py} \times N_{py}$, with

$$\alpha = \begin{bmatrix} 4t & -t & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ -t & 4t & -t & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & -t & 4t & -t & \cdot & \cdot & \cdot & 0 \\ & & & \cdot & \cdot & \cdot & & \\ & & & & \cdot & \cdot & \cdot & \\ \cdot & & & & & -t & 4t & -t \\ 0 & \cdot & \cdot & & & 0 & -t & 4t \end{bmatrix}_{N_{py} \times N_{py}} \quad (3.4)$$

and

$$\beta = \begin{bmatrix} -t & 0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & -t & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & -t & 0 & \cdot & \cdot & \cdot & 0 \\ & & & \cdot & \cdot & \cdot & & \\ & & & & \cdot & \cdot & \cdot & \\ \cdot & & & & & 0 & -t & 0 \\ 0 & \cdot & \cdot & & & 0 & 0 & -t \end{bmatrix}_{N_{py} \times N_{py}} \quad (3.5)$$

Here, $t = \hbar^2/2m^*a^2$ and a is the grid spacing.

The overall basis functions can be labeled as (\vec{k}_z, n_x, n_y) as in Fig. 3.1. The matrix elements representing $(H + U_{SC})$ then are

$$[H_L + H_T]_{n_x, n_y, \vec{k}_z; n_{x'}, n_{y'}, \vec{k}_z} = ([H_L]_{n_x, n_y, n_{x'}, n_{y'}} + \epsilon_{\vec{k}_z}) \delta_{\vec{k}_z, \vec{k}_z}$$

Thus we ignore scattering between two different transverse modes \vec{k}_z and \vec{k}_z' . So we can think of the transverse modes as separate 2-D devices connected in parallel (see Fig. 3.2).

3.1.1 Modified Fermi function

Each transverse mode \vec{k}_z has an extra transverse energy $\epsilon_{\vec{k}_z} = \hbar^2 k_z^2/2m$ that should be added to the longitudinal energy to get the total energy used in the argument of the Fermi function. Since all the transverse modes are like independent devices in parallel, all we have to do is replace the Fermi function in Eqs. 2.25, 2.27

$$f(E) = \left(1 + \exp \left[\frac{(E - E_f)}{k_B T} \right] \right)^{-1}$$

with

$$F(E) = \sum_{\vec{k}_z} f(E + \epsilon_{\vec{k}_z}) = \sqrt{\frac{2m k_B T}{\pi \hbar^2}} \mathcal{F}_{-1/2} \left(\frac{E_f - E}{k_B T} \right) \quad (3.6)$$

where $\mathcal{F}_{-1/2}(\eta)$ is the Fermi-Dirac integral of order $-1/2$. This accounts for all the transverse modes exactly. With f replaced by F we can now proceed as if the problem

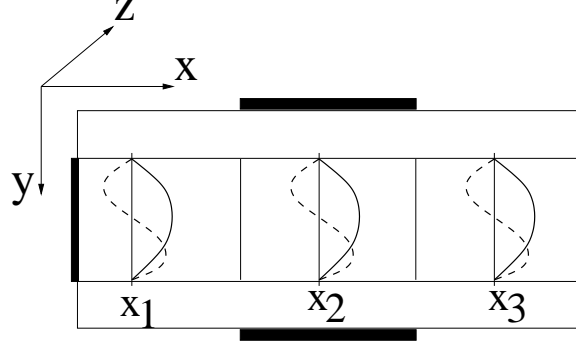


Fig. 3.3. y -confinement in a dual-gate MOSFET gives rise to discrete sub-bands.

was purely two-dimensional. Note that the function F in the above equation includes the sum over spins as well.

3.1.2 Basis transformation to mode space

We have now made a choice of representation (real space discrete lattice in the x - y plane and plane wave eigenstates in the z direction) and we have written down the Hamiltonian matrix in this basis. However, one practical problem remains. The choice of basis explained above seems natural till this point, but it leads to a computational problem. Consider a device like the one shown in Fig. 3.1. Let $L_g = 10$ nm and $T_{si} = 2$ nm. Let us assume that we need to include about 5 nm length of the source and drain regions in our simulation. With a grid spacing (a) of 0.2 nm, we get $N_{px} = 101$ and $N_{py} = 9$. Thus even for such a small device, the size of the Hamiltonian is 909×909 . The computational burden of doing a large number (approximately 100,000 for a full I-V characteristic) of matrix operations (inversions, multiplications etc.) with such a big matrix is too much for any commonly available computer. This forces us to look for another basis where the size of all the matrices is much smaller.

Fortunately, for the case of a dual-gate MOSFET in which electrons are tightly confined in the vertical y direction (see Fig. 3.1), such a choice of a ‘computer friendly’ basis is possible. Elementary quantum mechanics tells us that an electron confined in

a potential well can have a discrete ‘quantized’ set of energy levels or ‘modes’, just like the vibrational modes of a drum. Thus if we solve for the energy levels that arise due to the y -confinement (see Fig. 3.3), we will get a discrete set of N_{py} energy levels (there are infinite energy levels, but we capture the lowest N_{py} because our discrete lattice has N_{py} points in the y direction), or N_{py} modes. The trick is to recognize the fact that most of the carriers occupy only a few lowest modes out of the total N_{py} modes that we can have in our basis. Let us choose to include the lowest r modes. This number r may be obtained by starting with $r = 1$ and increasing r till the electron density ceases to change as we add more modes. Thus our new basis functions are the transverse (y direction) wavefunctions corresponding to the lowest r modes at each x point. In other words, at each point $x = x_n$, we solve the 1-D Schrödinger’s equation in the y direction in real space and use the lowest r wavefunctions as our basis functions at that point $x = x_n$. The transformation matrix V will be of size $N_{px}N_{py} \times N_{px}r$. For example, consider $N_{py} = 3$, $N_{px} = 5$ and $r = 2$. Then the transformation matrix V (size 15×10) will look like

$$V = \begin{bmatrix} X & X & 0 & 0 & \cdot & \cdot & 0 & 0 \\ X & X & 0 & 0 & & & \cdot & \cdot \\ X & X & 0 & 0 & & & \cdot & \cdot \\ 0 & 0 & X & X & & & & \\ 0 & 0 & X & X & & & & \\ 0 & 0 & X & X & & & & \\ \cdot & \cdot & \cdot & \cdot & & & \cdot & \cdot \\ & & & & & & & \\ \cdot & \cdot & \cdot & \cdot & & & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & & & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & & & X & X \\ \cdot & \cdot & \cdot & \cdot & & & X & X \\ 0 & 0 & 0 & 0 & \cdot & \cdot & X & X \end{bmatrix}_{15 \times 10}$$

where X's mark non-zero entries corresponding to the transverse (y) wavefunctions expressed in real space representation. Note that each column of V is normalized (orthonormal wavefunctions used) and is orthogonal to every other column; it is a unitary transformation matrix.

From this point on we use the term '*real space basis*' for the full real space basis explained at the beginning of this chapter and the term '*mode space basis*' for the truncated basis explained above. Note that in both these representations, the z direction is in a plane wave eigenstate representation. The only difference in the two representations is that in the real space basis both x and y directions have a discrete lattice while in the mode space basis only the x direction has a discrete lattice while the y direction has a truncated set of y -eigenstates as the basis functions.¹

Let

$$\overline{H}_{real}(E) = H_L + \Sigma_1(E) + \Sigma_2(E) \quad (3.7)$$

where H_L is given by Eq. 3.3 and $\Sigma_1(E)$ and $\Sigma_2(E)$ are the self-energy matrices in the real space basis. The procedure to set up the self-energy matrices in real space basis is described in detail in the next section. For the purpose of the following discussion, let us assume that we have set up the self-energy matrices in real space basis.

Now we transform the \overline{H}_{real} matrix (eq. 3.7) expressed in the real space basis to the mode space using the following standard formula in linear algebra:

$$\overline{H}_{mode}(E) = V^\dagger \overline{H}_{real}(E) V \quad (3.8)$$

This matrix is much smaller in size than \overline{H}_{real} . The size of \overline{H}_{mode} is $r \times r$ while the size of \overline{H}_{real} is $N_{px} N_{py} \times N_{px} N_{py}$. Now we can compute the mode space Green's function²

$$G_{mode}(E) = [EI - \overline{H}_{mode}(E)]^{-1} \quad (3.9)$$

This inversion is much more tractable because of the smaller size.

¹A similar mode space approach is used in [27] where all the y modes are treated separately (or assumed decoupled). In our approach two different y modes at the same x location may couple due to the self-consistent potential U_{SC} .

²For the orthogonal basis functions used in this chapter, the overlap matrix S in eq. 2.41 is equal to the identity matrix I .

Once we compute the Green's function, all the quantities of interest can be obtained using the standard set of NEGF equations (see eq. 2.21 through 2.27). The only point to remember is that the Fermi function in Eqs. 2.25,2.27 is replaced by the modified Fermi function (eq. 3.6). Once we obtain the mode space density matrix ρ_{mode} , we can get the real space density matrix ρ_{real} (whose diagonal elements give the electron density in real space) using a reverse transformation:

$$\rho_{real} = V \rho_{mode} V^\dagger \quad (3.10)$$

and

$$n = \text{diagonal}(\rho_{real}) \quad (3.11)$$

Finally, the terminal current is calculated using Eq. 2.27.

3.2 Self-energy

In this section we explain the procedure to set up the source and drain self-energy matrices in the real space basis. Consider a device connected to a single large reservoir or contact. The overall Hamiltonian matrix has the form:

$$\left[\begin{array}{c|c} H & \beta \\ \hline \beta^\dagger & H_C \end{array} \right] = \left[\begin{array}{c|ccccccc} \alpha & \beta & 0 & 0 & \cdot & \cdot & \cdot \\ \hline \beta^\dagger & \alpha & \beta & 0 & \cdot & \cdot & \cdot \\ 0 & \beta^\dagger & \alpha & \beta & \cdot & \cdot & \cdot \\ \cdot & 0 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{array} \right] \quad (3.12)$$

where α and β are matrices of size $N_{py} \times N_{py}$ as given in Eq. 3.4 and 3.5.

In Section 2.4 it was shown that the effect of a large contact can be incorporated in the device Green's function using a self-energy $\Sigma(E)$ (see eq. 2.37, 2.38):

$$G(E) = [EI - H - \Sigma(E)]^{-1} \quad (3.13)$$

where

$$\Sigma(E) = \tau C^{-1} \tau^\dagger \quad (3.14)$$

is the self-energy function and C^{-1} is the contact part of the overall Green's function.

Writing out the above equation, we have ($\bar{\alpha} \equiv E + i0^+ - \alpha$),

$$\begin{aligned} \Sigma &= \begin{bmatrix} -\beta & 0 & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \bar{\alpha} & -\beta & 0 & 0 & \cdot \\ -\beta^\dagger & \bar{\alpha} & -\beta & 0 & \cdot \\ 0 & -\beta^\dagger & \bar{\alpha} & -\beta & \cdot \\ \cdot & 0 & \cdot & \cdot & \\ \cdot & \cdot & & \cdot & \cdot \end{bmatrix}^{-1} \begin{bmatrix} -\beta^\dagger \\ 0 \\ \cdot \\ \cdot \end{bmatrix} \\ &= \beta g \beta^\dagger \end{aligned} \quad (3.15)$$

where g is the surface Green's function, which is the matrix formed from the first $N_{py} \times N_{py}$ entries of C^{-1} . Once we solve for g , the above equation gives us Σ . Now since C is a huge (effectively infinite) matrix, C^{-1} may be written recursively [26] in terms of C as

$$C^{-1} = \left[\begin{array}{c|ccc} E + i0^+ - \alpha & -\beta & 0 & \cdot \\ \hline -\beta^\dagger & C & & \\ 0 & & & \\ \cdot & & & \end{array} \right]^{-1} \quad (3.16)$$

We need g , which is a small part of C^{-1} . So we use the following matrix identity : If

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

then $a = [A - BD^{-1}C]^{-1}$. Using this identity in Eq. 3.16, we get

$$g = [E + i0^+ - \alpha - \beta g \beta^\dagger]^{-1} \quad (3.17)$$

Now we need to solve Eq. 3.17 for g and we are done. Eq. 3.17 may be solved in general using an iterative procedure [23, 24, 25, 26] in which we start with an initial

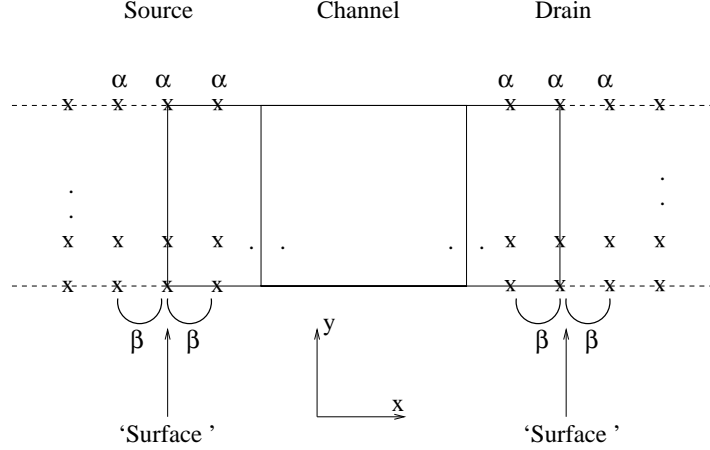


Fig. 3.4. The MOSFET is coupled to *two* semi-infinite contacts and not just one. Thus two self-energy matrices are needed, one for the source contact and another for the drain contact.

guess for g and keep on refining it using Eq. 3.17 till we obtain the solution. But for our special case, since β is a multiple of the identity matrix, the above equation may be written as a matrix quadratic equation as follows

$$\beta^2 g^2 + (\alpha - (E + i0^+)I)g + I = 0$$

It turns out that because all the matrices in the above equation viz. β , α and I commute with each other, the above equation may be solved just as if it were an ordinary scalar quadratic equation, and the solution is

$$g = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad (3.18)$$

where all the arithmetic operations in the above equation are to be carried out in a matrix sense, with $a = \beta^2$ (not to be confused with the symbol a used earlier for the grid spacing), $b = \alpha - (E + i0^+)I$ and $c = I$, the identity matrix.

When the surface Green's function obtained from Eq. 3.17 is substituted in Eq. 3.15, we get the self-energy function for the semi-infinite contact. This self-energy function takes into account the effect of a semi-infinite lead attached to a small device. But we are interested in a device connected to *two* contacts, the source and drain (see Fig. 3.1). The discussion till this point can be applied to this more complex structure

as follows. Consider the device connected to two semi-infinite leads. The only points where the device is coupled to the reservoirs are the points in the columns marked ‘surface’ in Fig.3.4. Thus the effect of the two reservoirs may be included in H_L (see Eq. 3.3) by adding two self-energy matrices $\Sigma_1(E)$ (for the source) and $\Sigma_2(E)$ (for the drain). $\Sigma_1(E)$ and $\Sigma_2(E)$ are both matrices of the same size as H_L i.e. $N \times N$, where $N = N_{px}N_{py}$ (see Fig. 3.2). The only non-zero entries in $\Sigma_1(E)$ will be the *first* $N_{py} \times N_{py}$ entries given by the matrix $\beta g_1 \beta^\dagger$ (see Eq. 3.15). Here g_1 may be found by solving Eq. 3.17 with the appropriate self-consistent potential at the source ‘surface’ points added to the diagonal elements of α . Similarly, the only non-zero entries in $\Sigma_2(E)$ will be the *last* $N_{py} \times N_{py}$ entries given by the matrix $\beta g_2 \beta^\dagger$. g_2 may be found by solving Eq. 3.17 with the appropriate self-consistent potential at the drain ‘surface’ points added to the diagonal elements of α .

3.3 Self-consistent Potential

The electron density $n(\vec{r})$ obtained using equation 3.11 is then used to solve for the self-consistent potential U_{SC} using the Poisson’s equation (ϵ : permittivity, N_D^+ : ionized donor density):

$$\vec{\nabla} \cdot (\epsilon \vec{\nabla} U_{SC}) = q^2 (N_D^+ - n) \quad (3.19)$$

in the active region of the device. The Poisson’s equation is solved on a rectangular grid spanning the device (Fig. 3.1). We use *zero-field* boundary conditions at the source and drain ends of the device and let U_{SC} float to whatever value it chooses to [27]³.

³The reason we use zero-field boundary conditions for the source and drain contacts may be understood from Fig. 3.6 which shows the potential as a function of position along the channel for high gate bias and different drain biases. Note that at the left (source) end the potential is actually pulled down by the applied drain bias. At high drain bias, there is a reduction in the net electronic charge at the source end since the left going (-k) states (supplied by the drain contact at equilibrium) are no longer occupied. Hence to preserve charge neutrality, the potential profile shifts down in order to compensate for the lost -k states with additional +k states. So while solving Poisson’s equation it is inconvenient to fix the potential at both ends. A better way is to use zero-field boundary conditions [27].

The real space self-consistent potential $U_{SC}(\vec{r})$ so obtained is added to the real space Hamiltonian (Eq. 3.1). The NEGF equations are then solved after changing over to the mode space basis. The new electron density is fed back to the Poisson solver and the procedure is repeated till neither n nor U_{SC} change from iteration to iteration.

3.4 Results

We couple our mode space NEGF equations with the Poisson equation to obtain the self-consistent I-V characteristic of a ballistic dual-gate MOSFET similar to that shown in Fig. 3.1. We use $L_g = 10$ nm, $T_{ox} = 1$ nm and $T_{si} = 2$ nm.

The source and drain regions are assumed to be uniformly doped n-type with $N_d = 10^{20} \text{ cm}^{-3}$. We use two different effective masses in the x and y directions, $m_x^* = 0.19 m_0$ and $m_y^* = 0.91 m_0$, because of the particular bandstructure of Silicon (see page 80 of reference [31] for details of the Silicon bandstructure). We choose Aluminum as the gate metal (gate work function = 4.25 eV). This material was chosen so as to achieve suitable on and off currents (positive threshold voltage). The bias range is 0 to 0.5 V for both the gate voltage and the drain voltage.

The linear and log scale I-V is shown in Fig. 3.5. The off current is about $0.03 \mu\text{A}/\mu\text{m}$ and the on current is about $1800 \mu\text{A}/\mu\text{m}$. These current values compare well with those in [27]. The subthreshold slope is the ideal 60 mV/dec as expected for a dual-gate device with an undoped body.

The self-consistent potential profiles averaged along the y-direction (y-axis is defined in Fig. 3.1) for different drain voltages and for low and high gate voltage are shown in Fig. 3.6. In Fig. 3.6, the zero of energy corresponds to the Fermi level deep inside the source. The potential profiles in Fig. 3.6 b correspond to the top I-V curve in Fig. 3.5 a. If we look at Fig. 3.6 carefully, we see that as the drain voltage is increased initially, the height of the energy barrier between the source and drain reduces, but as the drain voltage is further increased, the barrier stops moving down

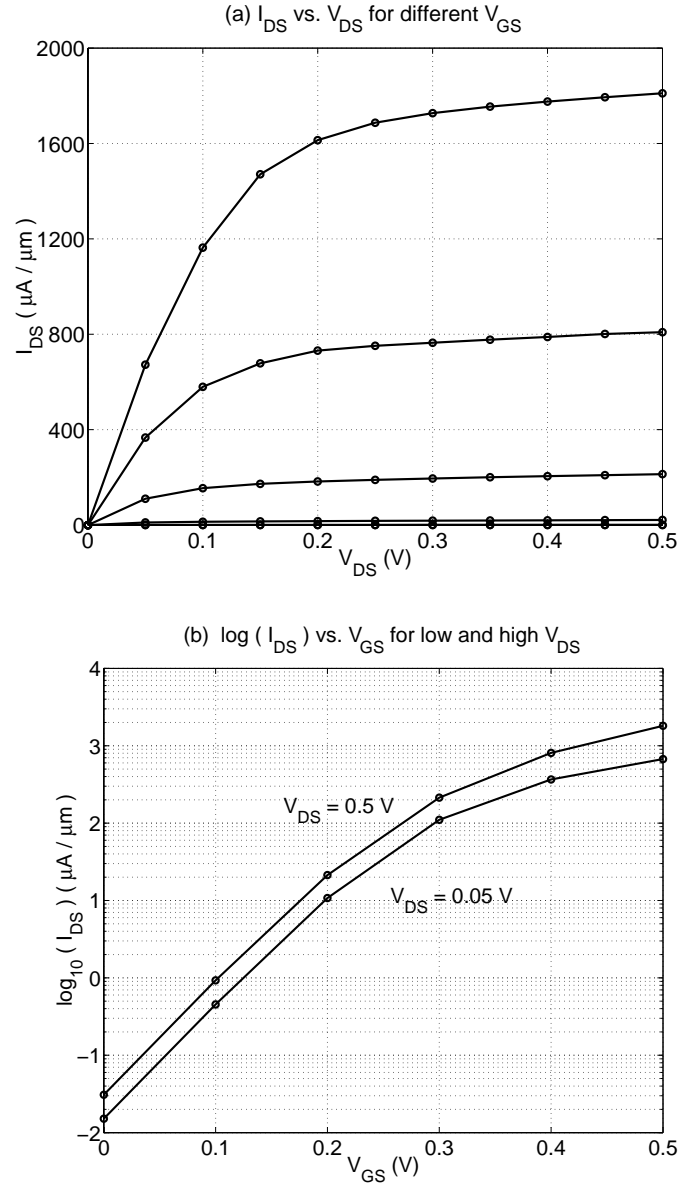


Fig. 3.5. Drain current vs. drain and gate voltages. (a) linear scale, (b) log scale. The on and off currents are about $1800 \mu A/\mu m$ and $0.03 \mu A/\mu m$ respectively. An ideal subthreshold slope of 60 mV/dec is observed.

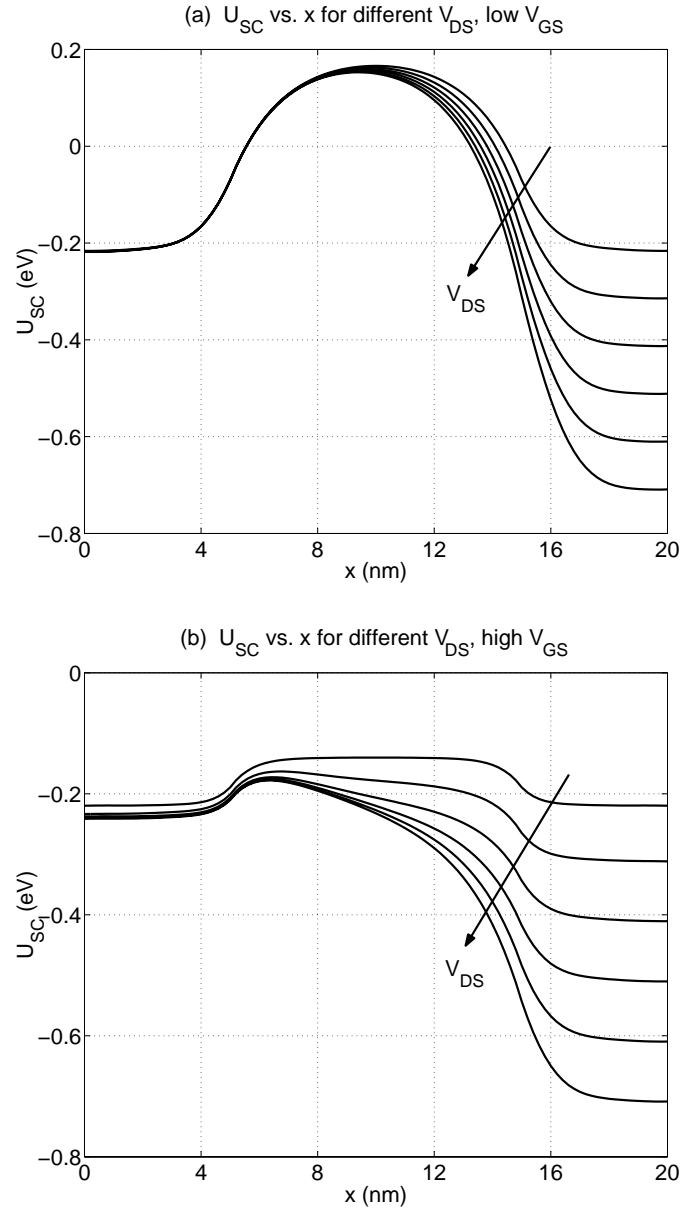


Fig. 3.6. Potential vs. position along the length of the MOSFET, (a) low V_{GS} (b) high V_{GS} . In (b) the source to drain barrier height stops decreasing for high drain bias leading to current saturation as seen in Fig. 3.5a.

because all voltage drops across the carrier depleted region near the drain. This is the reason why the current saturates, as explored in detail in [32, 33].

3.5 Summary

In this chapter we have presented a computationally efficient quantum model for the simulation of ballistic double-gate MOSFETs. We observe a huge gain in simulation time by using a mode space basis instead of the real space basis. This method is most suitable for double-gate MOSFETs because the spatial confinement allows us to use only a few mode space basis functions, thus reducing the size of the matrix used in computations. All the results presented in the previous section were obtained using a PC with a Pentium 3 processor. We will be glad to share our MATLAB codes with interested readers.

This chapter may be viewed as a follow-up to [12]. We hope that combined with [12], this and the previous chapter will serve as a good starting point for readers who wish to use the NEGF formalism to simulate nanoscale devices. The next step would be to add scattering in our model. This is an area of current research [30]. NEGF has clear prescriptions to include scattering processes, but sophisticated computing machinery seems essential to actually do it.

One may ask if the effective mass Hamiltonian (see Eq. 3.1) used in this chapter is appropriate for a nanoscale device. After all, the lattice constant of silicon is 5.43 Å, so a MOSFET with $L_g = 10$ nm and $T_{si} = 2$ nm (see Fig. 3.1) is about 20 atoms long and about 5 atoms thick. In such a situation, the self-consistent potential obviously varies on an atomic length scale and the validity of the effective mass approximation is in serious doubt. An atomic orbital basis set is more appropriate for this case. As a first step towards developing an atomic orbital based model for MOSFETs, one may simulate smaller molecular devices using atomic orbital basis sets. The following chapters discuss the treatment of such molecular devices in more detail.

4. MOLECULAR TRANSISTOR: SEMI-EMPIRICAL ATOMIC ORBITAL MODEL

In the previous chapter we illustrated the self-consistent procedure outlined in Section 1.2 using the example of a dual-gate nanotransistor. In this chapter, we use the same procedure to address an important question in the emerging field of molecular electronics: can a molecule, when used as the conducting channel in a standard three-terminal MOSFET configuration, offer any performance advantage relative to a standard silicon MOSFET.

The self-consistent solution of coupled quantum transport and Poisson’s equations shows that even for extremely small channel lengths (about 1 nm), a “well-tempered” molecular FET demands much the same electrostatic considerations as a “well-tempered” conventional MOSFET. In other words, we show that just as in a conventional MOSFET, the gate oxide thickness needs to be much smaller than the channel length (length of the molecule) for the gate control to be effective. Furthermore, we show that a “rigid” molecule (one which does not deform in an external field) with metallic source and drain contacts has a temperature independent subthreshold slope much larger than 60 mV/decade, because the metal-induced gap states in the channel prevent it from turning off abruptly. However, this disadvantage can be overcome by using semiconductor contacts because of their band-limited nature.

4.1 Introduction

Molecules are promising candidates as future electronic devices because of their small size, chemical tunability and self-assembly features. Several experimental molecular devices have recently been demonstrated (for a review of the experimental work see [2]). These include two terminal devices where the conductance of a molecule

coupled to two contacts shows interesting features such as a conductance gap [22], asymmetry [34] and switching [4]. Molecular devices where a third terminal produces a negative differential resistance [35], or suppresses the two terminal current [36] have been theoretically studied, but most of the work on modeling the current-voltage (IV) characteristics of molecular conductors has focused on two-terminal devices (see, for example, [17, 21, 37, 38, 39, 40] and references therein).

The purpose of this chapter is to analyze a three-terminal molecular device assuming that the molecule behaves essentially like a rigid solid. Unlike solids, molecules are capable of deforming in an external field and it may be possible to take advantage of such conformational effects to design transistors with superior characteristics. However, in this chapter we do not consider this possibility and simply address the question of whether a “rigid molecule” used as the conducting channel in a standard three-terminal MOSFET configuration can offer any performance advantage relative to a standard silicon MOSFET.

Although rigorous *ab initio* models are available in the literature [17, 38, 39, 40], they normally do not account for the three-terminal electrostatics that is central to the operation of transistors. For this reason we have used a simple model Hamiltonian whose parameters have been calibrated by comparing with *ab initio* models. We believe that a simple model Hamiltonian with rigorous electrostatics is preferable to an *ab initio* Hamiltonian with simplified electrostatics since the essential physics of a rigid molecular FET lies in its electrostatics.

The role of electrostatic considerations in the design of conventional silicon MOSFETs (with channel lengths ranging from 10 nm and above) is well understood. For the gate to have good control of the channel conductivity, the gate insulator thickness has to be much smaller than the channel length. Also, for a given channel length and gate insulator thickness, a double gated structure is superior to a single gated one, simply by virtue of having two gates as opposed to one. If a molecule is used as the channel in a standard three-terminal MOSFET configuration, the effective channel length is very small - about 1 nm. Would similar electrostatic considerations apply for

such small channel lengths? In this chapter we answer this question in the affirmative. Specifically we will show that:

- The only advantage gained by using a molecular conductor for an FET channel is due to the reduced dielectric constant of the molecular environment. To get good gate control with a single gate the gate oxide thickness needs to be less than 10% of the channel (molecule) length, whereas in conventional MOSFETs the gate oxide thickness needs to be less than 3% of the channel length [41]. With a double gated structure, the respective percentages are 60% and 20% [27].
- Relatively poor subthreshold characteristics (a *temperature independent* subthreshold slope much larger than 60 mV/decade) are obtained even with good gate control, if metallic contacts (like gold) are used, because the metal-induced gap states in the channel preclude it from turning off abruptly. Preliminary results with a molecule coupled to doped silicon source and drain contacts, however, show a temperature dependent subthreshold slope ($\sim k_B T/q$). We believe this is due to the band-limited nature of the silicon contacts, and we are currently investigating this effect.

Overall this study suggests that superior saturation and subthreshold characteristics in a molecular FET can only arise from novel physics beyond that included in our model. Further work on molecular transistors should try to capitalize on the additional degrees of freedom afforded by the “soft” nature of molecular conductors [42] - a feature that is not included in this study.

Although there has been no experimental report of a molecular FET to date ¹, judging from the historical development of the conventional silicon MOSFET, it is reasonable to expect that a single gated structure would be easier to fabricate than a double gated one. With this in mind, in this chapter we mainly focus on a single

¹We are aware of one experimental claim (J.H. Schön et al., Nature 413, page 713, 2001) reporting superior molecular FET with a single gated geometry. This claim, however, has been strongly questioned (see article by R.F. Service in Science 298, page 31, 2002).

gated molecular FET geometry (see Fig. 4.1). Few key results with a double gated geometry will be shown wherever appropriate to emphasize the differences between the single and double gated structures.

A schematic view of a molecule coupled to gold contacts (source and drain) is shown in Fig. 4.1a. As an example we use the Phenyl Dithiol (PDT) molecule which consists of a phenyl ring with thiol (-SH) end groups. A gate terminal modulates the conductance of the molecule. We use a simple model Hamiltonian H to describe the molecule (Fig. 4.1b). The effect of the source and drain contacts is taken into account using self-energy functions Σ_1 and Σ_2 [11]. Scattering processes may be described using another self-energy matrix Σ_p . However, in this chapter we focus on

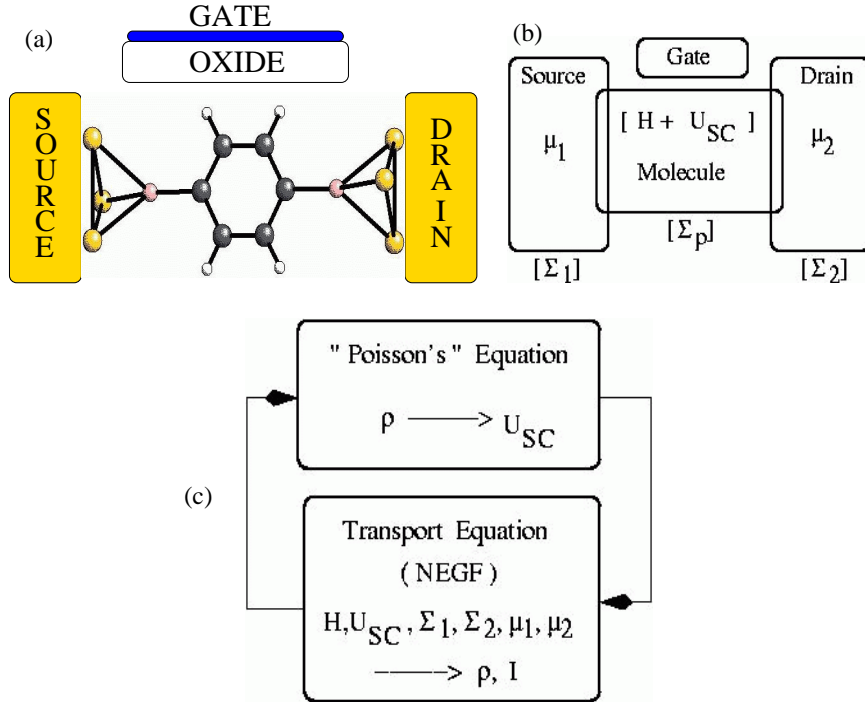


Fig. 4.1. (a) Schematic view of a Phenyl Dithiol molecule coupled to gold source and drain contacts. A third (gate) terminal modulates conductance of the molecule. The phenyl ring is shown in the plane of the paper for clarity, in the actual simulation the phenyl ring is perpendicular to the plane of the paper, “facing” the gate electrode. (b) and (c) are reproduced for convenience from Figure 1.1. The only difference between (b) and Figure 1.1 is the presence of the third (gate) terminal.

coherent or ballistic transport ($\Sigma_p = 0$). The source and drain contacts are identified with their respective Fermi levels μ_1 and μ_2 . The self-consistent procedure outlined in Section 1.2 is briefly revisited here for the sake of completeness.

We iteratively solve a set of coupled equations (Fig. 4.1c) - the Non-Equilibrium Green's Function (NEGF) formalism equations for the density matrix ρ and the Poisson's equation for the self-consistent potential U_{SC} . Given H , U_{SC} , Σ_1 , Σ_2 , μ_1 and μ_2 the NEGF formalism has clear prescriptions to obtain the density matrix ρ from which the electron density and the current may be calculated. Once the electron density is calculated we solve the Poisson's equation to obtain the self-consistent potential U_{SC} . The solution procedure thus consists of two iterative steps:

- **Step 1:** calculate ρ given U_{SC} using NEGF
- **Step 2:** calculate U_{SC} given ρ using Poisson's equation

The above two steps are repeated till neither U_{SC} nor ρ changes from iteration to iteration. It is worth noting that the self-consistent potential obtained by solving Poisson's equation (Eq. 4.4) may be augmented by an appropriate exchange-correlation potential that accounts for many electron effects using schemes like Hartree-Fock (HF) or Density Functional Theory (DFT) [7]. In this chapter we do not consider the exchange-correlation effects.

This chapter is organized as follows. To obtain the density matrix ρ from the self-consistent potential U_{SC} (step 1 above) using the NEGF formalism, we need to know the Hamiltonian H (Section 4.2, the contact self-energy matrices $\Sigma_{1,2}$ and the contact Fermi levels $\mu_{1,2}$ (Section 4.3). To obtain U_{SC} from ρ (step 2 above), we solve the Poisson's equation (Section 4.4). Section 4.5 presents the results along with a discussion of the underlying physical mechanisms.

4.2 Hamiltonian

We use a simple basis consisting of one p_z (or π) orbital on each carbon and sulfur atom. It is well known that the PDT molecule has π conjugation - a cloud

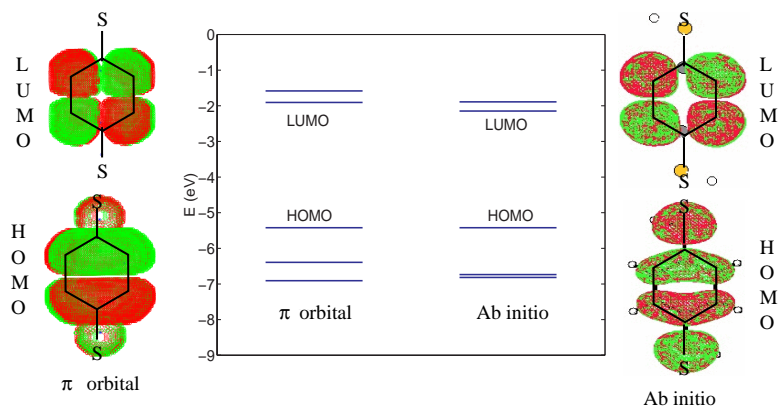


Fig. 4.2. Comparison of the simple π orbital based model with an ab initio model (density functional theory with atomic orbital basis set) for the PDT molecule. The energy levels can be divided in two sets: occupied levels (analogous to the valence band) and unoccupied levels (analogous to the conduction band). The energy gap (analogous to the bandgap) is the energy difference between the Highest Occupied Molecular Orbital (HOMO) level (analogous to the top of the valence band) and the Lowest Unoccupied Molecular Orbital (LUMO) level (analogous to the bottom of the conduction band). The simple π model agrees very well with the ab initio calculation in both the energy gap (middle) and HOMO and LUMO wavefunctions (left and right). The energy levels obtained from the simple model were equally shifted in energy so as to make the HOMO level coincide with the ab initio HOMO level.

of π electrons above and below the plane of the molecule that dictate the transport properties of the molecule [43]. The on-site energies of our p_z orbitals correspond to the energies of valence atomic p_z orbitals of sulfur and carbon (apart from a constant shift of all levels which is allowed as it does not affect the wavefunctions). The carbon-carbon interaction energy is 2.5 eV which is widely used to describe carbon nanotubes [44]. The sulfur-carbon coupling of 1.5 eV is empirically determined to obtain a good fit to the ab initio energy levels obtained using the commercially available quantum chemistry software Gaussian '98 [16] (Fig. 4.2).

Our model is very similar to the well established p_z orbital based Hückel theory used by many quantum chemists. Although we use a simple model Hamiltonian to describe the molecule, we believe that the essential qualitative physics and chemistry of the molecule is captured. This is because both the molecular energy levels and the

wavefunctions closely resemble those calculated from a much more sophisticated ab initio theory (Fig. 4.2).

4.3 Self-energy

As explained in Section 2.4, self-energy formally arises out of partitioning the molecule-contact system into a molecular subsystem and a contact subsystem. The contact self-energy Σ is calculated knowing the contact surface Green's function g and the coupling between the molecule and contact τ . For a molecule coupled to two contacts (source and drain) the molecular Green's function at an energy E is then written as ² (I : identity matrix, H : molecular Hamiltonian, U_{SC} : self-consistent potential):

$$G = [EI - H - U_{SC} - \Sigma_1 - \Sigma_2]^{-1} \quad (4.1)$$

where the contact self-energy matrices are

$$\Sigma_{1,2} = \tau_{1,2} g_{1,2} \tau_{1,2}^\dagger \quad (4.2)$$

We model the gold FCC (111) contacts using one s -type orbital on each gold atom. The coupling matrix element between neighboring s orbitals is taken equal to -4.3 eV - this gives correct surface density of states (DOS) of $0.07 / (\text{eV} - \text{atom})$ for the gold (111) surface [45]. The site energy for each s orbital is assumed to be -8.74 eV in order to get the correct gold Fermi level of ~ -5.1 eV. The surface Green's function g is calculated using a recursive technique explained in detail in [26]. The contact-molecule coupling τ is determined by the geometry of the contact-molecule bond. It is believed [46] that when a thiol-terminated molecule like Phenyl Dithiol is brought close to a gold substrate, the sulfur bonds with three gold atoms arranged in an

²The basis functions in the atomic (p_z) orbital basis set used in this chapter are assumed to be orthogonal. With orthogonal basis functions, the overlap matrix S in eq. 2.41 is equal to the identity matrix I . Non-orthogonal atomic orbital basis sets will be used in the ab initio treatment of molecular wires in the following chapters.

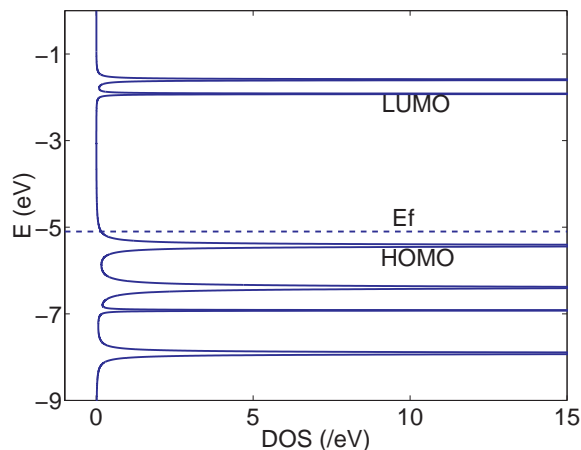


Fig. 4.3. The discrete levels of an isolated molecule broaden into a continuous density of states (DOS) upon coupling to contacts. The gold FCC (111) contacts are modeled using a basis of one 's' type orbital on each gold atom. The LUMO wavefunction is localized on the phenyl ring (Fig. 4.2) and gives a sharp peak in the DOS. The HOMO is delocalized and gives a comparatively broadened peak in the DOS. The equilibrium Fermi energy E_f (~ -5.1 eV for bulk gold) lies just above the HOMO level.

equilateral triangle. For a good contact extended Hückel theory predicts a coupling matrix element of about 2 eV between the sulfur p_z orbital and the three gold s orbitals. However to simulate the bad contacts typically observed in experiments [22, 38] we reduce the coupling by a factor of five (this factor is also treated as a parameter, and our results do not change qualitatively for a range of values of this parameter).

Unlike the Hamiltonian, the self-energy matrices are non-Hermitian. The anti-Hermitian part of the self-energy, also known as the broadening function:

$$\Gamma_{1,2} = i(\Sigma_{1,2} - \Sigma_{1,2}^\dagger) \quad (4.3)$$

is related to the lifetime of an electron in a molecular eigenstate. Thus upon coupling to contacts, the molecular density of states (Fig. 4.3) looks like a set of broadened peaks.

4.3.1 Where is the Fermi energy?

When a molecule is coupled to contacts there is some charge transfer between the molecule and the contacts, and the contact-molecule-contact system attains equilibrium with one Fermi level E_f . A good question to ask is where E_f lies relative to the molecular energy levels. The answer is not clear yet, the position of E_f seems to depend on what contact model one uses. A jellium model [38] for the contacts predicts that E_f is closer to the LUMO level for PDT whereas an extended Hückel theory based model [47] predicts that E_f is closer to the HOMO level (see Fig. 4.2 and the related caption for a description of HOMO and LUMO levels). Our ab initio model [17] seems to suggest that for gold contacts, E_f (~ -5.1 eV) lies a few hundred millivolts above the PDT HOMO. In this chapter we will use $E_f = -5.1$ eV and set the molecular HOMO level (obtained from the π model) equal to the ab initio HOMO level (~ -5.4 eV) (see Figs. 4.2, 4.3). Once the location of the equilibrium Fermi energy E_f is known, we can obtain the source and drain Fermi levels μ_1 and μ_2 under non-equilibrium conditions (non-zero V_{DS}): $\mu_1 = E_f$ and $\mu_2 = E_f - qV_{DS}$.

4.4 Self-consistent Potential

The Poisson's equation relates the real space potential distribution $U(\vec{r})$ in a system to the charge density $n(\vec{r})$. We assume a nominal charge density $n_0(\vec{r})$ obtained by solving the NEGF equations with $U(\vec{r}) = 0$ (at $V_{GS} = V_{DS} = 0$). The Poisson's equation is then solved for the *change* in the charge density ($n - n_0$) from the nominal value n_0 ³:

$$\vec{\nabla} \cdot (\epsilon \vec{\nabla} U(\vec{r})) = -q^2(n(\vec{r}) - n_0(\vec{r})) \quad (4.4)$$

³The potential distribution corresponding to the nominal charge density (when no drain or gate bias is applied) is included in the calculation of the molecular Hamiltonian [21].

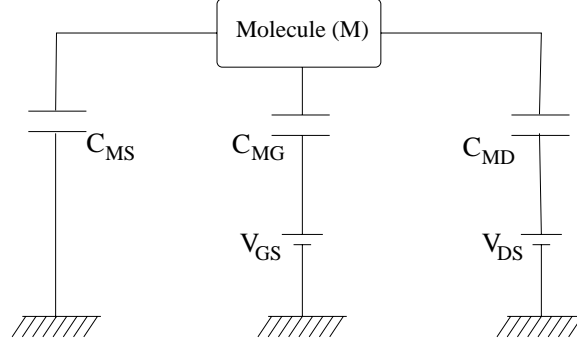


Fig. 4.4. Equivalent circuit model used to obtain the homogeneous (or zero charge) solution to the Poisson's equation. The molecular potential is controlled by the gate if the capacitive coupling C_{MG} between the molecule and the gate is much greater than the capacitive couplings C_{MS} and C_{MD} between the molecule and the source and drain respectively (see Eqs. 4.5, 4.6 and related discussion).

The Poisson (or Hartree) potential U may be augmented by an appropriate exchange-correlation potential U_{xc} . In this chapter, we do not take into account the exchange-correlation effects ($U_{xc} = 0$). We have two schemes to solve the Poisson's equation:

1. simple Capacitance Model, and
2. rigorous numerical solution over a 2D grid in real space.

4.4.1 Capacitance model

We use a simplified picture of the molecule as a quantum dot with some nominal *total* charge N_0 (at $V_{GS} = V_{DS} = 0$) and some average potential U arising because of the *change* $N - N_0$ in this nominal charge due to the applied bias. Thus U , N_0 and N are numbers and not matrices. The total charge N can be calculated from the NEGF density matrix using $N = \text{Trace}(\rho)$. U is the solution to the Poisson's equation, and may be written as the sum of two terms: (1) A Laplace (or homogeneous) solution U_L with zero charge on the molecule but with applied bias and (2) A particular (or inhomogeneous) solution U_P with zero bias but with charge present on the molecule. Thus $U = U_L + U_P$. U_L is easily written down in terms of the capacitive couplings

C_{MS} , C_{MD} and C_{MG} of the molecule (Fig. 4.4) with the source, drain and gate respectively:

$$U_L = \beta(-qV_{GS}) + \frac{(1-\beta)}{2}(-qV_{DS}) \quad (4.5)$$

where

$$\beta = \frac{C_{MG}}{C_{MS} + C_{MD} + C_{MG}} \quad (4.6)$$

is a parameter ($0 < \beta < 1$) and is a measure of how good the gate control is. Gate control is ideal when C_{MG} is very large as compared to C_{MS} and C_{MD} ⁴. In this case, $\beta = 1$ and the Laplace solution $U_L = -qV_{GS}$ is essentially tied to the gate. An estimate of gate control may be obtained from the numerical grid solution explained below by plotting β as a function of gate oxide thickness (Fig. 4.8).

The particular solution U_P may be written in terms of a charging energy U_0 as:

$$U_P = U_0(N - N_0) \quad (4.7)$$

The charging energy is treated as a parameter, and may be estimated as follows. The capacitance of a sphere of radius R is $4\pi\epsilon R$. If we distribute a charge of one electron on this sphere, the potential of the sphere is $q/4\pi\epsilon R$. For $R = 1$ nm the value of this potential is about 1.4 eV. Thus we use a charging energy $U_0 \sim 1$ eV. U_0 is the charging energy per electron per molecule and may also be estimated from the numerical grid solution by finding the average potential in the region occupied by the molecule and carrying one electronic charge distributed equally. This numerical procedure also yields $U_0 \sim 1$ eV and is used to estimate the charging energy while comparing the capacitance model with the numerical grid solution (see Fig. 4.9 and the related caption).

⁴We have assumed that $C_{MS} = C_{MD}$ in eq. 4.5, which is reasonable because the center of the molecule is equidistant from the source and drain contacts in our model (see Fig. 4.1). In general, if the source (drain) is closer to the molecule, then C_{MS} (C_{MD}) will be bigger [21]. With $C_{MS} = C_{MD}$, the molecular Laplace potential is $V_{DS}/2$ in the absence of a gate ($\beta = 0$), as is evident from eq. 4.5 (also see Fig. 4.6c,d and the related caption).

With the simple capacitance model just described, the Poisson's solution U is just a number. The self-consistent potential that adds to the p_z Hamiltonian (see Eq. 4.1) is then calculated as $U_{SC} = UI$, where I is the identity matrix of the same size as that of the Hamiltonian.

4.4.2 Numerical solution

We use a 2D real space grid to solve the discretized Poisson's equation for the geometry shown in Fig. 4.1a. The applied gate, source and drain voltages provide the boundary conditions. We use a dielectric constant of 3.9 for silicon dioxide and 2 for the self-assembled monolayer (SAM) [48].

The correct procedure to obtain the real space charge density $n(\vec{r})$ (see Eq. 4.4) from the p_z orbital space density matrix ρ is to make use of the following equation (Ψ_α is the α th p_z basis function):

$$n(\vec{r}) = \sum_{\alpha, \beta} \Psi_\alpha(\vec{r}) \Psi_\beta^*(\vec{r}) \rho_{\alpha\beta}.$$

However, we simplify the calculation of $n(\vec{r})$ by observing that a carbon or sulfur p_z orbital has a spread of about five to six Bohr radii (1 Bohr radius $a_B = 0.529 \text{ \AA}$). So for each atomic site α we distribute a charge equal to $\rho_{\alpha\alpha}$ equally in a cube with side $\sim 10a_B$ centered at site α .

The solution to Poisson's equation yields the real space potential distribution. However, the self-consistent potential U_{SC} that needs to be added to the Hamiltonian (Eq. 4.1) is in the p_z orbital space. We assume that U_{SC} is a diagonal matrix with each diagonal entry as the value of the real space solution U at the appropriate atomic position. For example, the diagonal entry in U_{SC} corresponding to the sulfur based p_z orbital would be equal to $U(\vec{r}_S)$ where \vec{r}_S is the position vector of the sulfur atom.

4.5 Results

The self-consistent procedure (Fig. 4.1c) is done with the two types of Poisson solutions discussed above. The simple capacitance model is fast while the 2D numerical solution is slow but more accurate. The capacitance model has two parameters, namely β which is a measure of the gate control, and U_0 which is the charging energy. These parameters can be extracted using the 2D numerical solution. We will first present results with the capacitance model by assuming ideal gate control, or $\beta = 1$. This ideal case is useful to explain the current saturation mechanism. We will then compare the results obtained from the capacitance model with those obtained from the numerical solution, and show that the two match reasonably well.

4.5.1 Ideal gate control, on state

Fig. 4.5 shows the molecular IV characteristic obtained by self-consistently solving the coupled NEGF - capacitance model Poisson's equations. We contrast the IV for ideal gate control ($\beta = 1$, Fig. 4.5a,b) with that for no gate control ($\beta = 0$, Fig. 4.5c,d). For each case, we have shown the IV for positive as well as negative drain voltage. We observe the following:

- With ideal gate control the IV is asymmetric with respect to the drain bias. For positive drain bias, we see very little gate modulation of the current. For negative drain bias we see current saturation and good gate modulation - the IV looks like that of a MOSFET.
- With no gate control the IV is symmetric with respect to the drain bias. There is no gate modulation.

These features of the IV characteristic may be understood as follows (Fig. 4.6). Let us first consider the ideal gate case. Since the gate is held at a fixed potential *with respect to the source*, the molecular DOS does not shift relative to the source

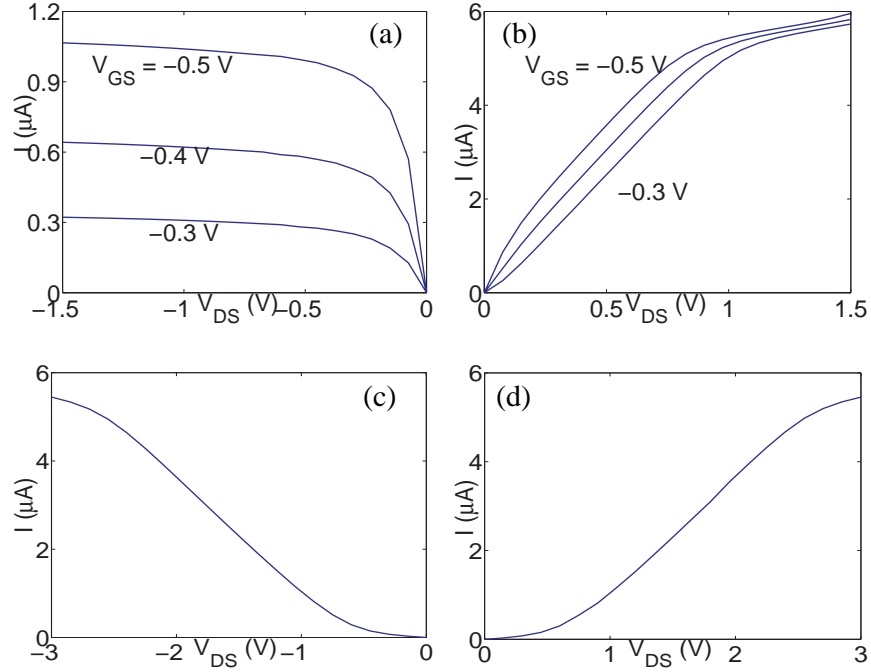


Fig. 4.5. Three terminal molecular drain current vs. drain to source voltage characteristic with (a) Ideal gate control ($\beta = 1$), negative drain bias, (b) Ideal gate control, positive drain bias, (c) No gate control ($\beta = 0$), negative drain bias and (d) no gate control, positive drain bias. With ideal gate control the IV is asymmetric with respect to drain bias. Good saturation and gate modulation is seen for negative drain bias but not for positive drain bias. With no gate control the IV is symmetric with respect to drain bias. For an explanation of the underlying mechanism for each of these IV curves, see Fig. 4.6.

Fermi level μ_1 as the drain bias is changed⁵. For negative drain bias (Fig. 4.6a), the drain Fermi level μ_2 moves up (towards the LUMO) with respect to the molecular DOS. Since the drain current depends on the DOS lying between the source and drain Fermi levels, the current saturates for increasing negative drain bias because the tail of the DOS dies out as the drain Fermi level moves towards the LUMO. If the gate bias is now made more negative, the molecular levels shift up relative to the source Fermi level, thereby bringing in more DOS in the energy range between μ_1 and μ_2

⁵This is true as long as the charging energy $U_0 \sim 1$ eV, which is typically the case. For high charging energies the particular solution U_P can dominate the Laplace solution U_L (see eqs. 4.5, 4.7 and related discussion), thereby reducing gate control.

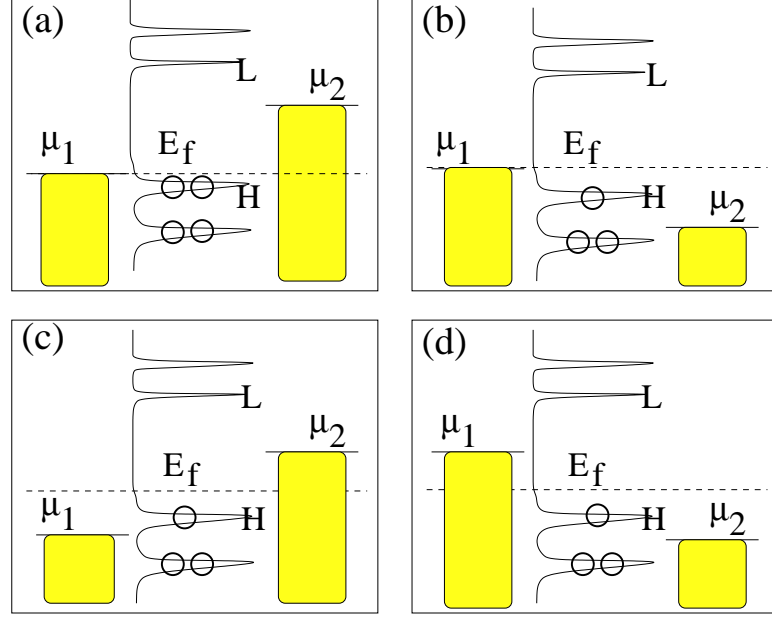


Fig. 4.6. Gate induced current saturation mechanism: Assuming that the gate is very close to the molecule (ideal gate control, $\beta = 1$), the gate holds the molecular DOS fixed relative to the source Fermi level μ_1 because the gate is held at a fixed potential with respect to the source. When a negative drain bias is applied (top left), the drain Fermi level μ_2 moves up relative to the molecular DOS. Since the DOS dies out in the gap, for sufficiently high drain bias, no more DOS comes in the μ_1 - μ_2 window and the current saturates. When a positive drain bias is applied (top right), μ_2 moves down relative to DOS and eventually crosses the HOMO. The IV is thus asymmetric. If the gate is far away (no gate control, $\beta = 0$), the DOS lies roughly halfway between the source and drain Fermi levels. In this case, for negative drain bias (bottom left), μ_1 crosses HOMO while for positive drain bias (bottom right) μ_2 crosses HOMO. No gate modulation is seen as expected, and the current is symmetric with respect to drain bias.

(referred to as the μ_1 - μ_2 window from now on) , and the current increases. Thus we get current saturation and gate modulation.

For positive drain bias (Fig. 4.6b), μ_2 moves down (towards the HOMO) with respect to the molecular DOS. The current increases with positive drain bias because more and more DOS is coming inside the μ_1 - μ_2 window as μ_2 moves towards the HOMO peak. Once μ_2 crosses the HOMO peak, the current levels off. This is the resonant tunneling mechanism. If the gate bias is now made more negative, no

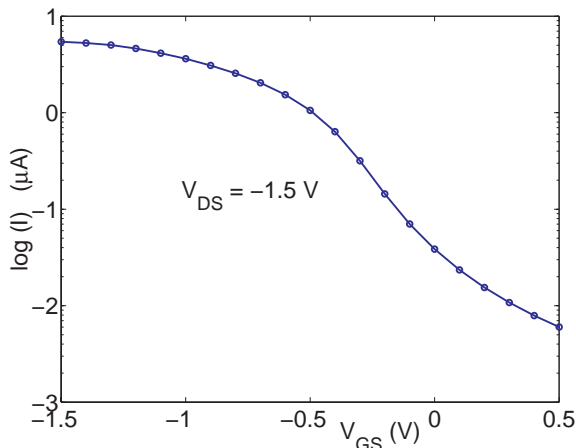


Fig. 4.7. Subthreshold IV characteristic assuming ideal gate control. The *temperature independent* subthreshold slope is ~ 300 mV/decade even with an ideal gate. This is because the DOS in the HOMO-LUMO gap dies out very slowly as a function of energy. This slow fall of the DOS may be attributed to the gold metal-induced gap states. Thus a molecular FET with a rigid molecule acting as the channel is a very poor switch.

appreciable change is made in the DOS inside the μ_1 - μ_2 window, and the maximum current remains almost independent of the gate bias.

Now let us contrast this with the case where no gate is present. In this case, due to the applied drain bias V_{DS} , the molecular DOS floats up by roughly $-qV_{DS}/2$ with respect to the source Fermi level. For either negative (Fig. 4.6c) or positive (Fig. 4.6d) drain bias, the current flow mechanism is resonant tunneling. Since the equilibrium Fermi energy lies closer to the HOMO, for negative drain bias μ_1 crosses HOMO while for positive drain bias μ_2 crosses HOMO [21, 13]. No gate modulation is seen as expected, and the IV is symmetric with respect to V_{DS} .

4.5.2 Ideal gate control, off state

Fig. 4.7 shows the log scale drain current as a function of gate bias at high drain bias. We note that despite assuming ideal gate control, the subthreshold slope of this molecular FET is about 300 mV/decade which is much worse than the ideal

room temperature $k_B T/q = 60$ mV/decade that a good MOSFET can come close to achieving. It is also worth noting here that our simulation is done at low temperature - the subthreshold slope of the molecular FET is *temperature independent* and only depends on the molecular DOS as explained below.

The poor subthreshold slope may be understood as follows. As the gate voltage is made more positive, the molecular DOS shifts down with respect to the μ_1 - μ_2 window. The HOMO peak thus moves away from the μ_1 - μ_2 window, and fewer states are available to carry the current. The rate at which the current decreases with increasing positive gate bias thus depends on the rate at which the tail of the DOS in the HOMO-LUMO gap dies away with increasing energy (Fig. 4.3). Typically we find that the tail of the DOS dies away at the rate of several hundred milli electronvolts of energy per decade, and this slow fall in the DOS determines the subthreshold slope. The slow fall in the molecular DOS may be attributed to the metal-induced gap (MIG) states - the gold source and drain contacts have a sizeable DOS near the Fermi energy, and are separated only by a few angstroms ⁶. Since the molecule is assumed to be rigid, the molecular DOS has no temperature dependence and hence neither does the subthreshold slope. Thus the molecular FET with a rigid molecule acting as the channel is a very poor switching device even with ideal gate control.

4.5.3 Estimate of gate control

The 2D numerical Poisson's solution may be used to estimate the gate control as follows. From Eq. 4.5 we see that

$$\beta = -\frac{1}{q} \left. \frac{\partial U_L}{\partial V_{GS}} \right|_{V_{DS}} \quad (4.8)$$

⁶For ballistic silicon MOSFETs, due to the band-limited nature of the doped silicon source/drain contacts, the MIG DOS is negligible. The subthreshold slope at a finite temperature is thus determined by the rate at which the difference in the source and drain *Fermi function tails* falls as a function of energy. This rate depends on the temperature, and the subthreshold slope is thus proportional to $k_B T/q$ (≈ 25 mV at room temperature) for ballistic Si MOSFETs [27]. Preliminary results for a molecular FET with doped silicon source and drain contacts do show a subthreshold slope proportional to $k_B T/q$; we are currently investigating this effect.

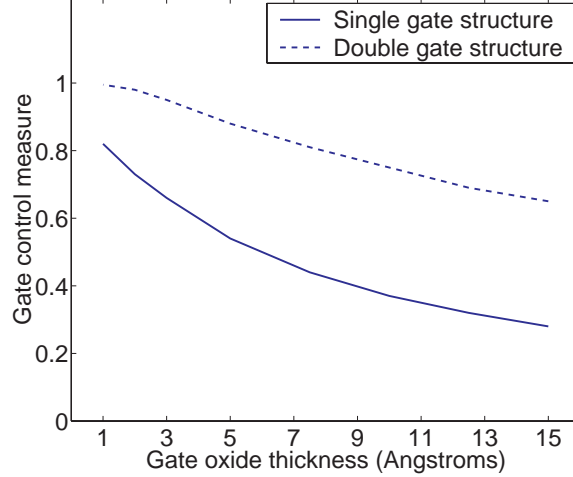


Fig. 4.8. Estimate of gate control using 2D numerical Laplace's solution with a single gated geometry (solid line) and a double gated geometry (dashed line). β (which is a measure of gate control, see Eqs. 4.5, 3.5, 4.8 and related discussion) is plotted as a function of the gate oxide thickness T_{ox} . The length of PDT molecule (equal to the channel length of the molecular FET) is about 1 nm. Thus in order to get good gate control ($\beta > 0.8$) the gate oxide thickness has to be about one tenth of the channel length, or about 1 Angstrom! We have used 3.9 and 2 as the dielectric constants for silicon dioxide and the self-assembled monolayer (SAM) respectively. For a double gated geometry, good gate control can be obtained with more realistic oxide thicknesses (of about 7 Angstroms), as expected simply because two gates can better control the channel than one.

Thus β may be estimated from the numerical solution by slightly changing V_{GS} (keeping V_{DS} constant) and calculating how much the Laplace's solution changes over the region occupied by the molecule. A plot of β calculated using this method as a function of the gate oxide thickness T_{ox} is shown in Fig. 4.8.

Knowing that the channel length (length of the PDT molecule) is about 1 nm, It is evident from Fig. 4.8 that in order to get good gate control ($\beta > 0.8$) the gate oxide thickness (T_{ox}) needs to be about one tenth of the channel length (L_{ch}), or about 1 Å! Thus we need $L_{ch}/T_{ox} \sim 10$ to get a good molecular FET. In a well-designed conventional bulk MOSFET, $L_{ch}/T_{ox} \sim 40$ [41]. This difference between a molecular FET and a conventional FET may be understood by noting that the

dielectric constant of the molecular environment ($\epsilon=2$) is about 6 times smaller than that of silicon ($\epsilon=11.7$)⁷.

Fig. 4.8 also shows β as a function of T_{ox} calculated using the 2D numerical Laplace's solution over a double gated molecular FET structure. In this case, we find that to get good gate control, we need $L_{ch}/T_{ox} \sim 1.6$. Thus a double gated structure is superior to a single gated one for a given L_{ch} and T_{ox} , as is also expected for conventional silicon MOSFETs. The reason for this is simply that two gates can better control the channel than one.

4.5.4 Comparison: capacitance model versus numerical solution

Fig. 4.9 compares the IV characteristic obtained by solving the self-consistent NEGF-Poisson's equations with the numerical Poisson's solution and the capacitance model. The parameters β and U_0 for the capacitance model were extracted from the numerical solution. We see a reasonable agreement between the two solutions despite the simplifications made in the capacitance model (the capacitance model assumes a flat potential profile in the region occupied by the molecule, which may not be true, especially at high bias). For $t_{ox} = 1.5$ nm (Fig. 4.9a) there is very little gate modulation and no saturation as expected. In this case $\beta = 0.28$ (Fig. 4.8) and the IV resembles that shown in Fig. 4.5c more than the one in Fig. 4.5a. Also seen in Fig. 4.9 are the results for $t_{ox} = 1$ Å. For this case $\beta = 0.82$ and we observe reasonable saturation and gate control. For realistic oxide thicknesses, however, we expect to observe an IV like the one shown in Fig. 4.9a. We have also calculated the IV characteristics with a double gated geometry (not shown here), and as expected from Fig. 4.8, saturating IVs can be obtained for more realistic oxide thicknesses (~ 7 Å).

⁷Mark Lundstrom, School of Electrical and Computer Engineering, Purdue University, West Lafayette, Indiana, private communication, 2002.

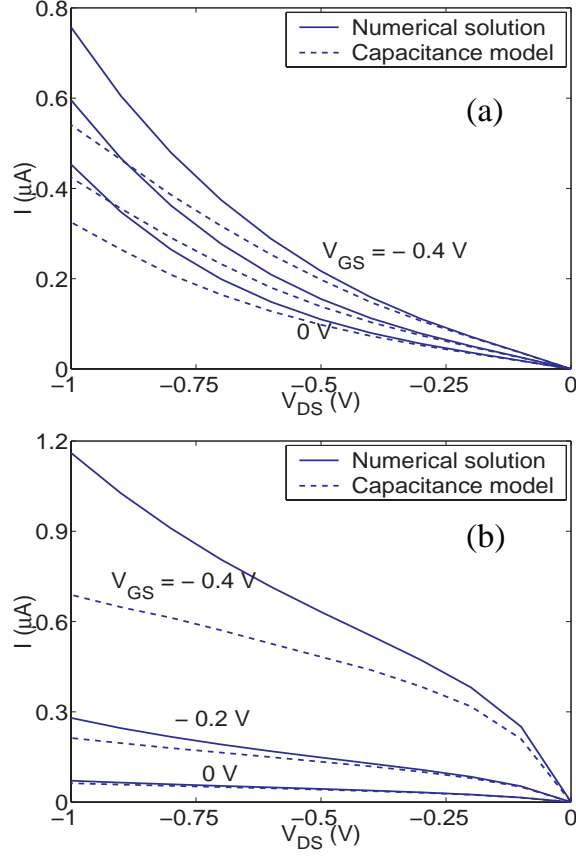


Fig. 4.9. Comparison of the numerical Poisson solution with the capacitance model. We see reasonable agreement in the two solutions despite the simplifications made in the capacitance model. (a) A realistic case with $t_{ox} = 1.5 \text{ nm}$ which yields $\beta = 0.28$ and $U_0 = 1.9 \text{ eV}$. No current saturation is seen, but some gate modulation is present. (b) An "absurd" case with $t_{ox} = 1 \text{ Angstrom}$ which yields $\beta = 0.82$ and $U_0 = 1 \text{ eV}$ (U_0 is less for this case because the gate is closer to the molecule; screening effect of the gate electrode is larger). The IV for this case looks like that for a MOSFET. Similar IV may be obtained with $t_{ox} = 1 \text{ nm}$, provided one uses a gate insulator with a dielectric constant about ten times that of silicon dioxide.

4.6 Summary

We have presented simulation results for a three terminal molecular device with a rigid molecule acting as the channel in a standard MOSFET configuration. The NEGF equations for quantum transport are self-consistently solved with the Poisson's equation. We conclude the following:

1. The current-voltage (IV) characteristics of molecular conductors are strongly influenced by the electrostatics, just like conventional semiconductors. With good gate control, the IV characteristics will saturate for one polarity of the drain bias and increase monotonically if the polarity is reversed. By contrast two-terminal symmetric molecules typically show symmetric IV characteristics.
2. The only advantage gained by using a molecular conductor for an FET channel is due to the reduced dielectric constant of the molecular environment. To get good gate control with a single gate the gate oxide thickness needs to be less than 10% of the channel (molecule) length, whereas in conventional MOSFETs the gate oxide thickness needs to be less than 3% of the channel length. With a double gated structure, the respective percentages are 60% and 20%.
3. Relatively poor subthreshold characteristics (a *temperature independent* subthreshold slope much larger than 60 mV/decade) are obtained even with good gate control, if metallic contacts (like gold) are used, because the metal-induced gap states in the channel preclude it from turning off abruptly. Preliminary results with a molecule coupled to doped silicon source and drain contacts, however, show a temperature dependent subthreshold slope ($\sim k_B T/q$). We believe this is due to the band-limited nature of the silicon contacts, and we are currently investigating this effect.
4. Overall this study suggests that superior saturation and subthreshold characteristics in a molecular FET can only arise from novel physics beyond that included in our model. Further work on molecular transistors should try to exploit the additional degrees of freedom afforded by the “soft” (as opposed to rigid) nature of molecular conductors [42].

5. MOLECULAR WIRES: AB INITIO ATOMIC ORBITAL MODEL

In the previous chapter we used a semi-empirical atomic orbital based model to treat a three-terminal molecular device. Although semi-empirical models are computationally efficient and easy to use, it is worthwhile to benchmark them against more rigorous ab initio models. To this end, in this chapter we will use an ab initio, atomic orbital based model to describe current flow in two-terminal molecular wires. The ab initio treatment of three-terminal molecular devices is a conceptually straightforward extension, and is left for future work.

5.1 Introduction

There is much current interest in molecular electronics due to the recent success in measuring the I-V characteristics of individual or small groups of molecules [2, 3, 4, 21, 22, 49, 50, 51]. The measured resistances exhibit a wide range of values. For example, n-alkane chains ($CH_3 - (CH_2)_{n-1}$) have large gaps (6 eV or greater) between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) and act like strong insulators [52, 53] while gold nanowires (or quantum point contacts) have zero gap (or a continuous density of states near the Fermi energy) and exhibit novel one-dimensional metallic conduction characteristics [54, 55]. Similar one-dimensional metallic conduction (but on a much larger length scale of the order of micrometers) has been observed in carbon nanotubes [56]. Then there are biological molecules like the DNA [57, 58] whose electrical conduction characteristics are currently the subject of much debate [59]. Interesting device functionalities such as switching have also been recently reported [4].

Understanding the correlation between the chemical and electronic properties of a wide class of molecules is a first step towards developing a ‘bottom-up’ molecule-based technology. Semi-empirical theories [43, 47, 60, 61, 62, 63, 64, 65, 66] have been used to qualitatively study electronic transport in molecules. Some first-principles methods have also been developed [38, 39, 40, 67, 68, 69]. Typically the first-principles methods are either computationally very expensive or ignore charging effects by employing non self-consistent calculations. Furthermore, the correct geometry and atomicity of the contacts is often ignored by employing a jellium-like model although surface effects like bonding and chemisorption are clearly very important. The purpose of this chapter is to describe a straightforward (computationally inexpensive) yet rigorous and self-consistent procedure for calculating transport characteristics while taking into account the effects mentioned above.

We partition the problem into a ‘device’ and a ‘contact’ subspace (see Figure 5.1) such that standard quantum chemistry techniques can be employed to analyze the electronic structure of a finite-sized device subspace, while incorporating the effects of the outside world (‘contacts’) through self-energy matrices (as discussed in Section 2.4). Such partitioning makes the molecular chemistry conceptually transparent as well as computationally tractable. The contact subspace is treated via a one-time calculation of the surface Green’s function of the contacts including their atomicity and crystalline symmetry. Different molecules coupled to the same contacts have different couplings but the contact surface Green’s function is independent of the molecule. Given the surface Green’s function and the contact-molecule coupling we can describe the chemisorption and bonding of the molecule with the infinite contacts through self-energy matrices of finite size (equal to that of the molecular subspace). The NEGF formalism (Chapter 2) has clear prescriptions to calculate the non-equilibrium density matrix from a knowledge of the molecular Fock matrix F ($F \equiv H + U_{SC}$, see Equation 2.35), the contact self-energy matrices $\Sigma_{1,2}$ and the con-

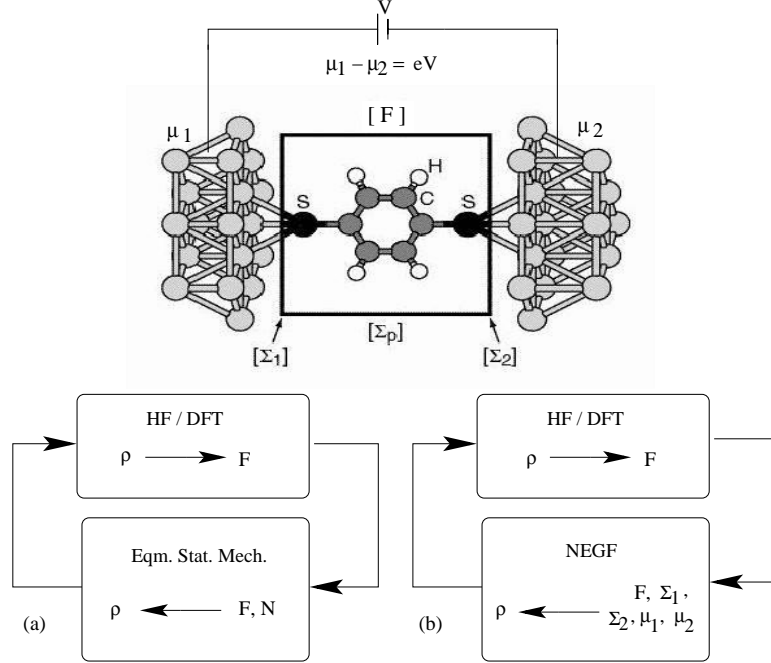


Fig. 5.1. Schematic view of a molecule (Phenyl Dithiol) connected to two semi-infinite gold (111) contacts. The self-energy matrices Σ_1 and Σ_2 *exactly* account for the contacts and are of the same size as the molecular Fock matrix F . The self-consistency scheme is shown for (a) isolated molecule in equilibrium and (b) contact-molecule-contact system under bias.

tact electrochemical potentials $\mu_{1,2}$. All quantities of interest like electron density and current may then be calculated from the self-consistently converged density matrix¹.

We illustrate our self-consistent method using a gold nanowire and a phenyl dithiol molecule sandwiched between gold contacts and study some previously addressed is-

¹For the most part of this chapter (one exception is made in Fig 5.12 a, bottom panel) we neglect non-coherent interactions of the molecule with its surroundings. For such a ‘ballistic’ molecule we may use the transmission formalism to obtain the current (see Eq. 2.27). We can modify the transmission formalism to phenomenologically model incoherent scattering via Büttiker probes [70] (see bottom panel in Fig 5.12 a). We associate a dephasing strength η and a single electrochemical potential μ_p with each Büttiker probe. μ_p is obtained iteratively by asserting that the *total* current injected by the Büttiker probes is zero (f_p is the Fermi function with electrochemical potential μ_p):

$$(2e/h) \left[\int dE T_{1p} (f_p - f_1) + \int dE T_{2p} (f_p - f_2) \right] = 0$$

This implies that we model incoherent *inelastic* scattering as opposed to incoherent *elastic* scattering which is commonly used [47, 71] to model phase breaking processes. This procedure is used for the bottom panel in Fig 5.12 a.

sues like (1) charge transfer and self-consistent band lineup (e.g. [72]), (2) I-V characteristics (e.g. [47, 60, 38]) and (3) charge density and voltage drop (e.g. [67, 63]). Metallic conduction with quantum unit conductance is observed in the gold nanowire. Upon reducing the coupling to contacts, the gold nanowire exhibits resonant tunneling type of conduction just as seen in a phenyl dithiol molecule. The introduction of a defect (a stretched bond) in the nanowire gives rise to a sharp voltage drop across the impurity as expected. The presence of the defect leads to negative-differential resistance (NDR) in the I-V characteristic of the wire. Periodic Friedel oscillations are observed in the charge density near the defect, the magnitude of these oscillations decreasing as expected upon the introduction of phase-breaking scattering.

According to our overall solution procedure (Section 1.2), we must first choose an appropriate basis set to describe the contact-molecule-contact system and then set up the molecular Fock matrix and the contact self-energy matrices in the chosen basis. The choice of basis and the calculation of the Fock matrix (which is the sum of the core Hamiltonian and the self-consistent potential) is described in Section 5.2. The calculation of the contact self-energy matrices is described in Section 5.3. Some results and the underlying physics is contained in Section 5.4. The chapter ends with a brief summary (Section 5.5).

5.2 Hamiltonian And Self-consistent Potential

The electronic structure of an isolated device is obtained in standard quantum chemistry through a self-consistent procedure [7, 8] shown schematically in Fig 5.1 a. The process consists of two steps that have to be iterated to obtain a self-consistent solution. Step 1: Calculate the Fock matrix F for a given density matrix ρ using a specific scheme such as Hartree-Fock (HF) or density functional theory (DFT) to obtain the self-consistent field, and Step 2: Calculate the density matrix from a given self-consistent Fock matrix F and total number of electrons N , based on the laws of equilibrium statistical mechanics.

The problem of calculating the current-voltage (I-V) characteristics is different from the above self-consistent procedure in the following ways : (i) We are dealing with an open system having a continuous density of states and variable (in principle fractional) number of electrons, rather than an isolated molecule with discrete levels and integer number of electrons; (ii) The molecule does not necessarily remain at equilibrium or even close to equilibrium - two volts applied across a short molecule is enough to drive it far from equilibrium; (iii) Surface effects like chemisorption and bonding with the contacts are expected to play a non-trivial role in transport. To calculate the density matrix (step 2 above) we therefore need a method based on non-equilibrium statistical mechanics applicable to such an open system with a continuous density of states. The NEGF formalism (Chapter 2) provides us with such a method and that is what we use in this chapter for step 2. Step 1, however, remains unchanged and we use exactly the same procedure as in standard quantum chemistry software. The overall procedure is shown schematically in Fig 5.1 b.

From a computational viewpoint, the most challenging part is the calculation of the Fock matrix F , which involves the core molecular Hamiltonian and the self-consistent potential. A number of researchers have developed their own schemes for performing such ab-initio computations [38, 67, 68]. We accomplish this part by exploiting the fast algorithms of Gaussian '98 [16], a commercially available quantum chemistry software ². Aside from the computational advantages of using an already well-established software, such a close coupling to the standard tools for analyzing molecules makes the chemistry of the system clearer. We describe the contacts and the molecule using the sophisticated LANL2DZ basis set [73, 74] which incorporates relativistic core pseudopotentials. The self-consistent potential is calculated using DFT with Becke-3 exchange [75] and Perdew-Wang 91 correlation [76]. Thus, equipped with a SUN workstation, we are able to perform a parameter-free analysis of conduction in molecular wires in a few hours.

²The modifications made to Gaussian '98 FORTRAN code are listed in Appendix D.

5.3 Self-energy

In Section 2.4 we discussed how the effect of large contacts can be incorporated into the device Green's function G using self-energy matrices $\Sigma_{1,2}$:

$$G = [ES - F - \Sigma_1 - \Sigma_2]^{-1}$$

where S is the overlap matrix and F is the device Fock matrix ($F = H + U_{SC}$). The self-energy functions are written in terms of the device-contact couplings $\tau_{1,2}$ and the contact surface Green's functions $g_{1,2}$ as

$$\Sigma_{1,2} = \tau_{1,2} g_{1,2} \tau_{1,2}^\dagger.$$

The geometry of the bonding surface determines the coupling matrix and the surface Green's function. For thiol bonds, for example, it is believed that a chemically bonded sulfur atom on a gold surface overlaps equally with three gold atoms that form an equilateral triangle as shown in Figure 5.1 [46, 77]. We use the LANL2DZ basis set [73, 74] to describe both the contacts and the molecule. LANL2DZ is a sophisticated basis set with relativistic core pseudopotentials and is observed to provide a good description of the contact surface density of states around the Fermi energy³ [78]. The coupling matrices $\tau_{1,2}$ are calculated by using Gaussian '98 to simulate an 'extended molecule' [72, 78] consisting of the molecule under consideration and an equilateral triangle of three gold atoms on either side of the molecule.

As explained in Section 3.2, the surface Green's function of a semi-infinite contact obeys the recursive equation [26] involving the on-site and coupling matrices α and β (see Fig. 5.2):

$$g^{-1} = \alpha - \beta g \beta^\dagger \tag{5.1}$$

³For a proper description of the bandstructure of gold away from the Fermi energy, we observe that it is necessary to include upto the 4th or 5th nearest neighboring interactions, because some of the basis functions in LANL2DZ are relatively delocalized.

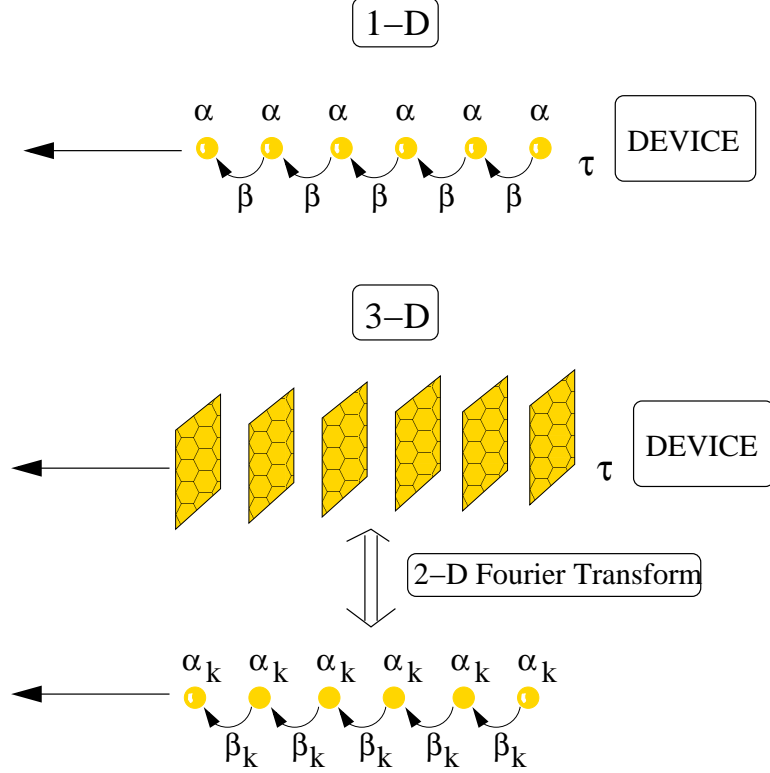


Fig. 5.2. Device coupled to a semi-infinite contact (top: 1-D contact, bottom: 3-D contact). The contact self-energy depends on the device-contact coupling τ and the contact surface Green's function g which is recursively solved for using the contact on-site and coupling matrices α and β (see text). In case of a 3-D contact, the in-plane periodicity is used to obtain an equivalent 1-D picture with k -dependent α_k and β_k .

For a semi-infinite 3-D contact of a specific material and orientation, we follow the procedure in Chapter 3 of [26] and go to the 2-D k -space representation for each cross-sectional plane of the lead, so that each k -point effectively acts as an independent 1-D problem for which the above recursive formula holds. The procedure is schematically illustrated in Fig. 5.2.

We use this procedure to calculate the surface Green's function of a gold (111) contact (commonly used in experiments) as follows. A cluster of gold atoms belonging to three consecutive (111) planes is shown in Fig 5.3. All the nearest neighbors (6 in-plane and 6 out-of-plane) of atom 1 are shown. We use Gaussian '98 to extract

in-plane and out-of-plane nearest-neighbor overlap matrix S and Fock matrix F components using such a gold cluster and define 2-D Fourier components in the (111) plane (see Fig. 5.2):

$$\begin{aligned} F_{a\vec{k}} &= \sum_n F_{mn} e^{-i\vec{k} \cdot (\vec{r}_m - \vec{r}_n)} \\ S_{a\vec{k}} &= \sum_n S_{mn} e^{-i\vec{k} \cdot (\vec{r}_m - \vec{r}_n)} \end{aligned} \quad (5.2)$$

where m is an arbitrary gold atom, and n involves a sum over m and all its nearest in-plane neighbors, with coordinate \vec{r}_n . The out-of-plane Fourier components $F_{b\vec{k}}$ and $S_{b\vec{k}}$ are also defined analogously.

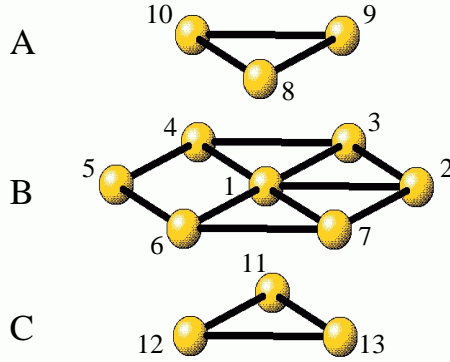


Fig. 5.3. Central 13 atom part of the 28 atom gold cluster in FCC (111) geometry used to calculate the coupling matrices S_{mn} and F_{mn} in Eq. 5.2. A reference atom (atom 1) and all its nearest neighbors are shown in this central 13 atom part. 28 atoms are used so as to reduce the ‘edge’ effects that tend to destroy the symmetries associated with the Fock matrices F_{mn} (see the discussion following Eq. 5.2 and Appendix A).

The F_{mn} and S_{mn} matrices must all obey the group theoretical symmetry of the FCC crystal. This symmetry is satisfied by the S_{mn} matrices since the overlap between two atoms depends only on those two atoms and not on the rest of the atoms in the gold cluster. The F_{mn} matrices, however, are obtained via a self-consistent

calculation (see Fig 5.1 a) and depend on the presence or absence of other atoms in the cluster. Ideally we need to simulate an infinite cluster (or crystal) in order to get the Fock matrices to obey the symmetry. For practical reasons we simulate a finite cluster consisting of 28 gold atoms (the central 13 atom part of this cluster is shown in Fig 5.3) and enforce the known symmetry rules for a gold (111) geometry on the resulting Fock matrices. A detailed description of this procedure of enforcing symmetry is given in Appendix A. With the correct symmetry imposed, the $F_{a\vec{k}}$ and $S_{a\vec{k}}$ matrices in Eq. 5.2 are Hermitian. The gold surface Green's function in \vec{k} space is then obtained by iteratively solving [26]

$$g_{\vec{k}}^{-1} = \alpha_{\vec{k}} - \beta_{\vec{k}} g_{\vec{k}} \beta_{\vec{k}}^{\dagger} \quad (5.3)$$

where

$$\alpha_{\vec{k}} = (E + i0^+) S_{a\vec{k}} - F_{a\vec{k}}$$

and

$$\beta_{\vec{k}} = (E + i0^+) S_{b\vec{k}} - F_{b\vec{k}}$$

The real-space gold (111) surface Green's function matrices are then obtained using

$$g_{mn} = \frac{1}{N} \sum_{\vec{k}} g_{\vec{k}} e^{i\vec{k} \cdot (\vec{r}_m - \vec{r}_n)} \quad (5.4)$$

where N is the number of unit cells in the (111) plane (or the number of k points). The surface Green's function matrices so obtained are independent of the molecule under consideration and depend only on the material and geometry of the contacts.

The energy dependent self-energy matrices discussed above exactly account for the valency and chemisorption of the molecule onto the contact surface. For gold (111) contacts we find that these issues are approximately taken care of even if we replace $\Sigma(E)$ with $\Sigma(E_f)$ (E_f is the gold Fermi energy) as is evident from Figs 5.4, 5.5, 5.7. This helps to reduce the time taken to compute the density matrix (see Eq. 2.25). We have developed a fast and elegant analytical method to evaluate the density matrix for an energy independent self-energy. This method is explained in Appendix B. Such a simplification may not be possible for platinum contacts having a significant

structure in the density of states near E_f . This makes the calculation of the density matrix computationally quite challenging because the correlation function may have sharp peaks in the range of integration. Integrating over a complex energy contour [72, 79] simplifies the computational complexity to a certain degree, but the energy range of integration has to be huge in order to take into account all the molecular levels (see Appendix B), and a faster scheme is desirable.

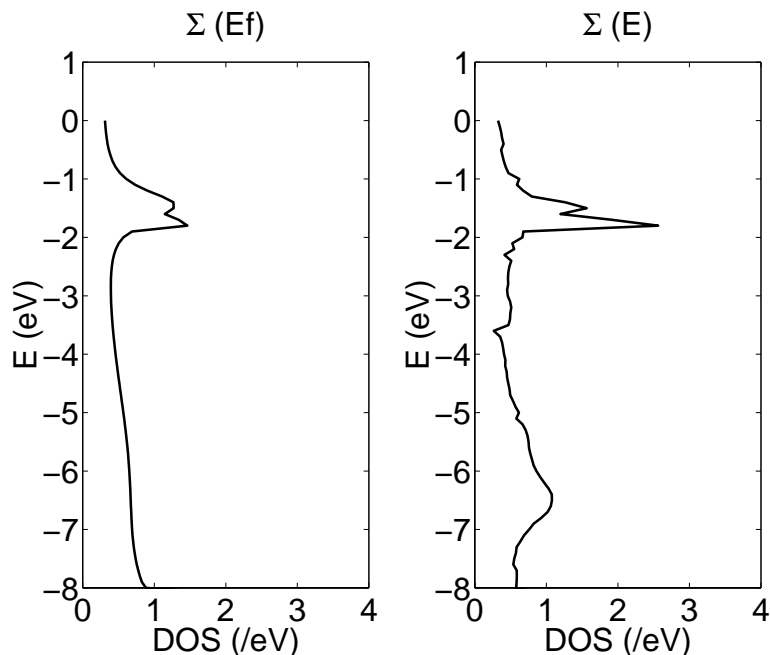


Fig. 5.4. Comparison of non-self-consistent evaluations of DOS for phenyl dithiol with and without a constant self-energy approximation. The good correspondence is a consequence of the relatively flat DOS of Au(111) near the Fermi energy.

We have thus managed to partition the system into a device and a contact subspace. The self-energy matrices replacing the contacts are non-Hermitian, their real parts representing the shift in the molecular energy levels due to coupling with the infinite contacts, and their imaginary parts representing the broadenings of these levels into a continuous density of states (the inverse of the non-Hermitian part of the self-energy represents the residence time of an electron in the molecule before it escapes into the contact). The action of partitioning away the contacts has thus lead

to an open device with a continuous density of states. The partitioning makes the problem computationally tractable, since the size of the self-energy matrices is the same as the device overlap and Fock matrices, even though they represent infinitely large contacts. In addition, the partitioning decomposes the problem into three different subspaces each involving three different areas of research: (i) the device Fock matrix F incorporates the quantum chemistry of the intrinsic molecule; (ii) the coupling matrix τ involves details of the bonding between the molecule and the contact (chemisorption, physisorption etc.) and (iii) the surface Green's function g involves the surface physics of the contact, which could in principle be extended to include additional effects such as surface states, band-bending, surface adsorption and surface reconstruction.

5.4 Results

5.4.1 Equilibrium

Restoration of bonding and valency via the self-energy. The partitioning of the contact-molecule-contact system (see Fig 5.1) is *exact* and the self-energy is expected to take care of the bonding and valency issues. For example, the energy levels of an isolated phenyl dithiol molecule are shown in Fig 5.5 (leftmost panel). If we look at the energy levels of the same molecule, but with the two end hydrogen atoms removed (Fig 5.5, second left panel), the levels look completely different due to unsatisfied valency of the sulfur atoms - sulfur atom needs to bond with the hydrogen atom in order to restore valency. It is well-known that the hydrogen atom desorbs when the molecule comes in contact with a gold substrate. If we calculate the energy levels of an 'extended molecule' (Fig 5.5, second right panel) we find that the original molecular levels (solid) are restored showing that the three gold atoms on each end effectively replace the end hydrogen atoms. In addition to the original molecular levels, we find that new gold-based levels are introduced (dashed). If we now look at the self-consistent DOS of a hydrogen-desorbed molecule connected to gold contacts (via a

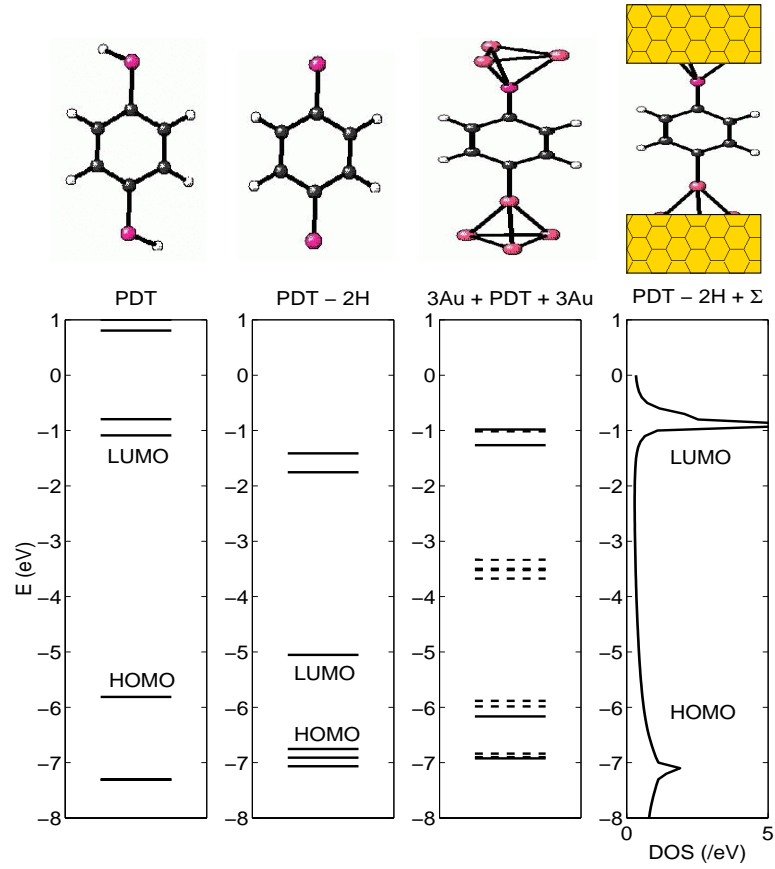


Fig. 5.5. An appropriate self-energy can take care of bonding even if the ‘molecule’ we define has an unsatisfied valence. The energy levels (B3PW91/ LANL2DZ) of phenyl dithiol (leftmost panel) differ considerably from those of the hydrogen-desorbed molecule (left center), but are restored on coupling to gold (111) clusters (right center). Hybridized electrons from the gold triangle replace those missing from the H atoms, restoring the valency of the sulfur atoms and also give rise to additional gold based levels, shown in dashed lines. Rightmost panel shows the self-consistently calculated density of states (DOS) of a hydrogen-desorbed phenyl dithiol (PDT-2H) chemisorbed on gold contacts (described by self-energy matrices). A sharp LUMO peak (PDT LUMO is localized on the phenyl ring) and a well-broadened HOMO peak (PDT HOMO is delocalized over the molecule) is seen in addition to a non-zero DOS in the gap attributed to metal induced gap states.

self-energy) as shown in the rightmost panel in Fig 5.5, we find a very sharp LUMO peak (as expected because the PDT LUMO is localized on the phenyl ring) and a well-broadened HOMO peak (as expected because the PDT HOMO is delocalized

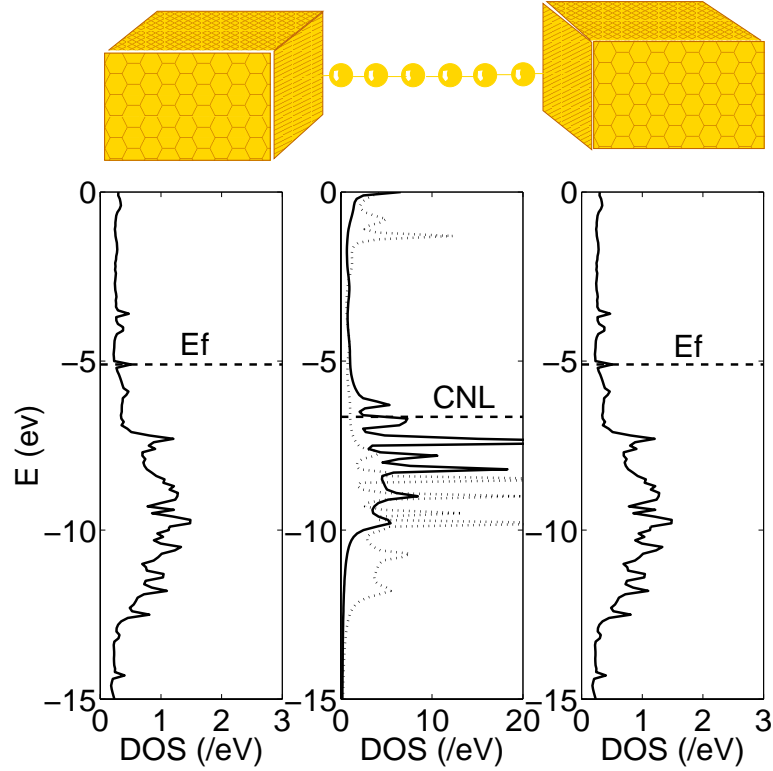


Fig. 5.6. Charge transfer and self-consistent band lineup upon coupling a gold nanowire to two gold contacts. The charge neutrality level (CNL) of the wire lies below the Fermi energy of gold, leading to electron flow from contacts to the wire. The DOS (dashed) of the charge-neutral wire redistributes and floats up (solid) self-consistently due to charging energies associated with the electron flow.

over the molecule), in addition to a non-zero DOS in the gap attributed to metal induced gap states (MIGS).

The above discussion on the adequacy of an appropriate self-energy in partitioning the system at any arbitrary point becomes an academic issue if we incorporate several metal layers into our device such that the molecule now has a satisfied valency. However, it is worth emphasizing that in principle we can draw a device-contact boundary at any point and take care of unsatisfied valency and charge neutrality through the self-energy matrices.

Charge transfer and band lineup. The charge transferred between a device and contacts leads to aligning of reference energy levels in the two subsystems. Such

band lineup issues are of primary importance in understanding semiconductor heterojunctions as well as plastic electronics. An analogous band-lineup diagram can be obtained on coupling a molecule to contacts. Fig 5.6 shows the surface DOS of the contacts and the DOS of the device (a gold nanowire in this case). The 6 atom gold chain has discrete energy levels that are broadened into a metallic band-like DOS on coupling to contacts. For the chain with a broadened density of states, one can define a charge-neutrality level (CNL) [80], such that filling up all the states below the CNL keeps the device charge-neutral. The Fermi energy of gold, approximately -5.1 eV (negative of the gold work function), is noticeably higher than the CNL of a 1-D gold nanowire. Owing to this difference electrons flow in from the contacts to the device. For a small molecule, the capacitive charging energy is large, so the original broadened levels (dashed) float up self-consistently (solid) due to charging till the charge transfer ceases, the band lineup is complete and the device and contacts are in chemical equilibrium.

5.4.2 Non-equilibrium

I-V characteristics. A gold nanowire connected to gold contacts has a continuous DOS near the equilibrium Fermi energy (see Fig 5.6, middle panel), and hence we expect it to exhibit metallic conduction as shown in Fig 5.7 a. The nanowire exhibits a conductance equal to the quantum unit conductance of $2e^2/h \approx 77\mu S$. This ohmic I-V is a testimony to the accuracy of our self-energy matrix. Only a correct self-energy matrix will get rid of spurious reflections at the contact and seamlessly couple the 1-D gold wire with the 3-D gold contact. Fig 5.7 b shows the I-V characteristic for the same gold nanowire, but with the coupling to the contacts reduced by a factor of 4. Due to the reduced coupling, the broadening of the levels is much less than the separation between the levels, and the wire now exhibits resonant conduction. There are marked plateaus in the I-V curve when a level lies between the contact Fermi levels. This is shown schematically in Fig 5.8. Fig 5.9 shows the energy levels

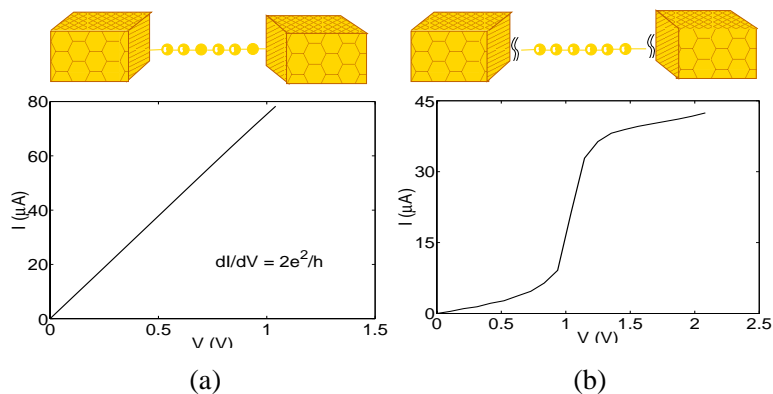


Fig. 5.7. (a) I-V of a Gold nanowire with six gold atoms forming a 1-D chain connected to two gold contacts. An energy-independent self-energy $\Sigma(E_f)$ is used to model the contacts. A quantum unit conductance of $2e^2/h$ signifies that $\Sigma(E_f)$ accounts for the perfect transmission by seamlessly joining the nanowire to the contacts. (b) I-V of the same gold nanowire but with reduced contact coupling which decreases level broadening and a resonant tunneling type conduction is seen. The I-V shows step-like behavior with the current jumping up when the contact Fermi levels μ_1 and/or μ_2 cross a molecular level (See Figs 5.8, 5.9).

of the weakly coupled gold nanowire as a function of applied bias. At equilibrium, the contact Fermi level lies close to the LUMO level. For a non-zero bias V , μ_1 and μ_2 separate by an amount eV ⁴. μ_2 moves closer to LUMO and wants to fill it up. Charging effects now come into the picture and the levels tend to float up and follow μ_2 until μ_1 comes close to crossing the HOMO level. At this point, μ_2 wants to put charge into the LUMO while μ_1 wants to empty the HOMO. It turns out that the number of electrons lost is more than the number gained, and now the levels tend to go down with μ_1 in order to prevent further loss of electrons. The current jumps up (see Fig 5.7 b) at that bias point where μ_2 crosses the LUMO level. Such resonant

⁴We choose to split μ_1 and μ_2 equally around the equilibrium Fermi energy E_f , or $\mu_1 = E_f - eV/2$ and $\mu_2 = E_f + eV/2$ (see Fig. 5.8). This choice is made in order to be consistent with the Gaussian '98 convention of applying the linear voltage drop (through the 'field' option) symmetrically across the molecule. In other words, for a molecule placed with its center (along the z -axis) at $z = 0$, Gaussian '98 applies the field such that one end of the molecule is at a voltage $+V/2$ and the other end is at $-V/2$. One may choose to split μ_1 and μ_2 in some other manner, and the results will not vary provided the field is applied consistently.

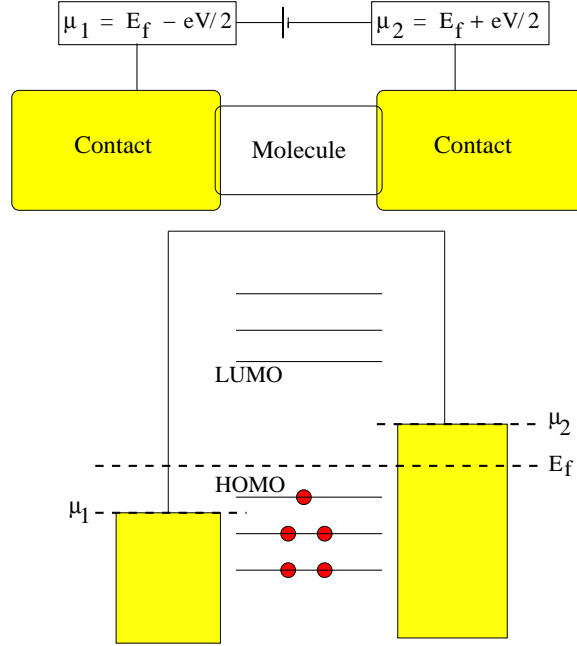


Fig. 5.8. At equilibrium ($V = 0$) $\mu_1 = \mu_2 = E_f$ and no current flows. For non-zero V the contact Fermi energies separate by an amount eV and a significant current flows only when a molecular level lies in between μ_1 and μ_2 (also see Fig 5.9).

conduction is also seen in a phenyl dithiol molecule (Fig 5.11) with pronounced peaks in the conductance dI/dV at resonance.

The I-V characteristic for a quantum point contact with a defect at the center exhibits a weak negative differential resistance (NDR) (Fig 5.10). We artificially stretch a bond in the middle of our 6 atom gold chain, which causes the left and right local density of states (LDOS) of the chain to separately be in equilibrium with the left and right contacts. Applying a bias then causes the two LDOS to sweep past each other within $\mu_{1,2}$. Owing to the presence of some sharp van-Hove singularities in the DOS and some smoothened out maxima, there is a progressive alignment and misalignment of peaks in the LDOS that leads to a weak NDR in the I-V. A similar mechanism for NDR has been observed in other calculations [68, 43].

Although most semi-empirical theories qualitatively match the shape of the resonant I-V characteristics for molecular conductors such as phenyl dithiol (PDT), quantitative agreements between experiment and theory have been largely unsatisfactory.

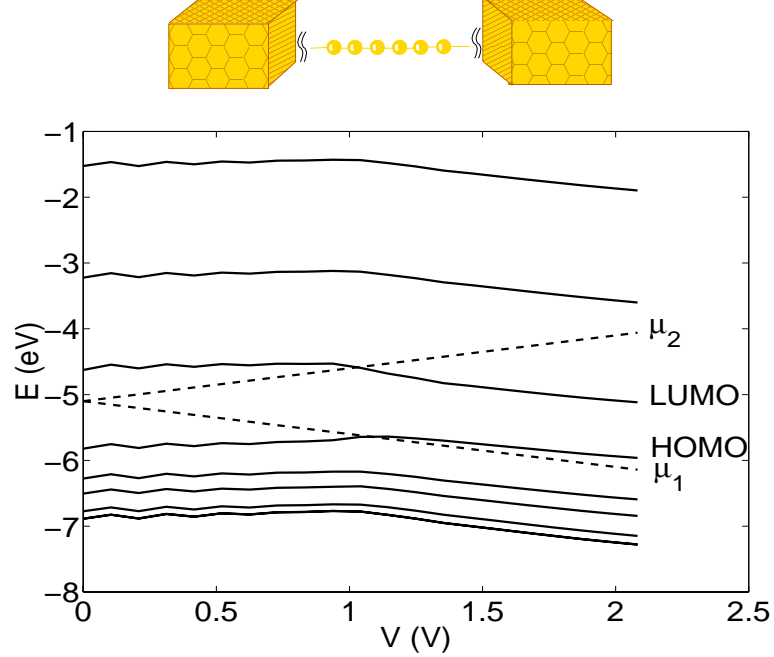


Fig. 5.9. Energy levels of a gold nanowire weakly coupled to contacts as a function of applied bias. The I-V for this case is shown in Fig 5.7 b. At zero bias the contact Fermi level is close to LUMO. At non-zero bias μ_2 wants to fill up the LUMO and add extra electrons in the device, so charging effects tend to float the levels up till μ_1 comes close to HOMO and wants to empty it. At this point, more electrons are lost than gained and charging effects make the levels float down parallel to μ_1 in order to minimize the loss of electrons.

The I-V for PDT can be quantified by the conductance gap and the current value at the onset of conduction. If we neglect the charging energy U_c the gap is roughly given by the proximity of the contact Fermi energy to the nearest molecular level, $E_{gap} = 2|E_F - E_{mol}|$ [21]; while the peak current at the onset of conduction is roughly given by $2e\Gamma_1\Gamma_2/(\Gamma_1 + \Gamma_2)\hbar$, evaluated at the energy E_{mol} . The current onset is smeared out over the tail of the molecular level, which in conjunction with charging effects leads to a rise in current at the molecular level crossing extended over a voltage width $\Delta V \approx (\Gamma_1 + \Gamma_2 + U_C)/e$. Ab-initio theories are indispensable for obtaining self-consistently the position of the Fermi energy relative to the molecular levels, as well as the broadenings of the levels. Both from ab-initio calculations for PDT, as well as ab-initio estimates based on the above arguments, the predicted conductance

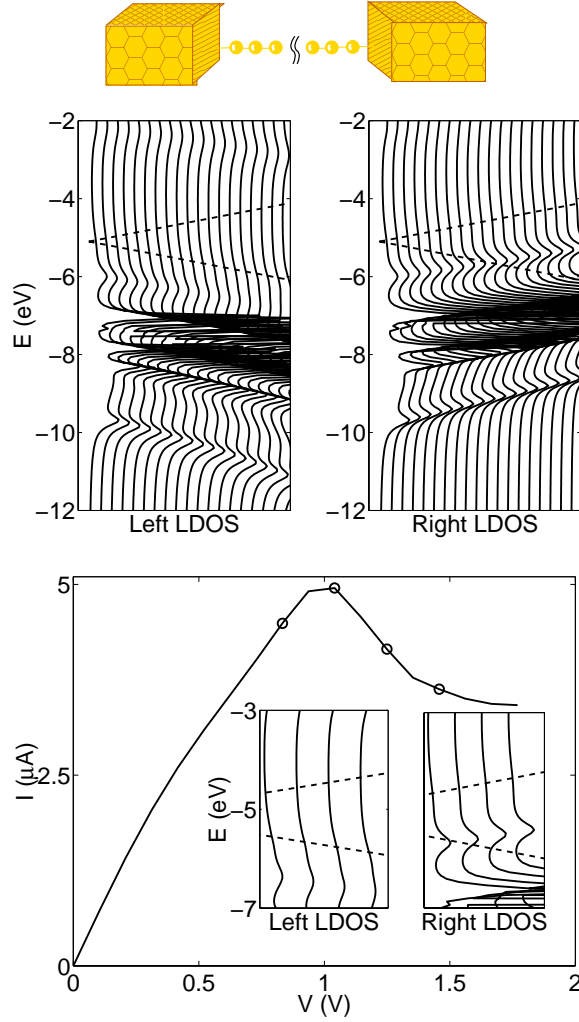


Fig. 5.10. Left and right LDOS for a wire with a defect in the center for 21 equidistant voltage values from 0 to 2 volts, laterally shifted equally for clarity. The electrons in the left and right segments of the wire are separately in equilibrium with the left and right contacts, and their respective LDOS follow the corresponding contact electrochemical potentials (dashed lines) because of charging effects similar to those seen in Fig 5.9. As the two LDOS slide past each other within the μ_1 - μ_2 window (shown magnified in the inset of the bottom figure at the four voltage points circled in the I-V graph) their peak values (tail of the peak for the left LDOS) come in and out of resonance, producing thereby a weak NDR in the I-V characteristic.

gap of PDT, set by the molecular HOMO level, comes out to about 4 Volts, while the maximum current turns out to be tens of microamperes. Experimentally both the conductance gap and the maximum current are much less than these theoretical

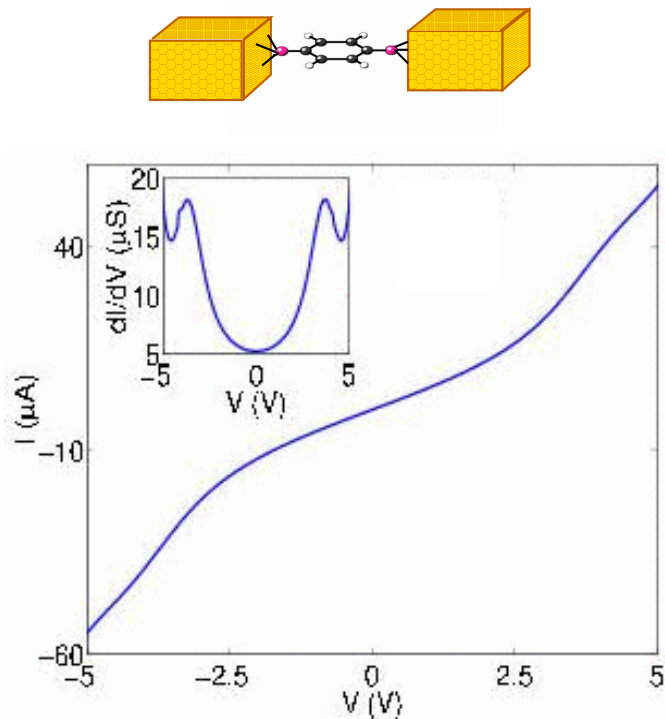


Fig. 5.11. I-V of phenyl dithiol connected to two gold contacts. The underlying mechanism is resonant tunneling - much like that in a gold nanowire weakly coupled to the contacts (see Fig 5.7 b). The gap depends on the proximity of the Fermi energy to the zero bias HOMO level if we neglect charging effects, while the maximum current at the onset of conduction is proportional to the parallel combination of broadenings at the molecular energy (see text). For phenyl dithiol, the predictions from the above estimates are about 4 volts and 40 μA respectively.

estimates. The decrease in current can be attributed either to weak coupling between molecular π orbitals and the s-orbitals of loose gold atoms at the contact surface [81, 38], or to tunneling between molecular units separately chemisorbed on two ends of a break-junction [37]. Thus a precise experimental knowledge of the contact conditions is essential prior to appropriate surface modeling. The conductance gap is as yet unexplained by the above attempts, and requires once again a precise knowledge of contact surface conditions which can vary the Fermi energy substantially enough to alter the conductance gap.

Voltage drop and electron density. An applied voltage is known to drop largely across the metal-molecule interface, leading to a weaker drop in the molecule. The precise nature of the potential profile is an important input to semi-empirical calculations of transport. Tian et al. [47] suggested using a flat potential profile inside the molecule, with a voltage division factor describing its position. Such a flat profile was obtained by Mujica et al. [63] by solving a 1-D Poisson equation, and experimentally measured for larger ($\sim \mu\text{m}$) wires by Seshadri and Frisbie [82]. However, in all these cases the geometry under consideration is a series of 2-D charge sheets with potential variations only along the wire axis. The 1-D Poisson equation allows variations only along one coordinate, while the measurements in [82] referred to a self-assembled monolayer (SAM) where once again transverse potential variations are screened out by the presence of neighboring molecules. In contrast, Lang and Avouris [67] obtained a significant potential drop in a carbon atomic wire, which is a consequence of fields penetrating from transverse directions, as correctly predicted for a break-junction geometry by a 3-D Poisson equation. Our particular geometry is suited to the break-junction, since the Hartree term in Gaussian '98 is calculated for a 3-D geometry.

Figure 5.12 a shows a gold nanowire weakly coupled to the contacts. In Fig 5.12 b, the coupling to contacts is strong, but a substitutional defect has been incorporated by artificially introducing a bond that is longer than the equilibrium gold bond length. The voltage drop across the device itself is smaller than the applied voltage bias owing to screening effects, incorporated self-consistently through the Hartree term of our Fock matrix. The plots on top show the net self-consistent voltage drop across the wire (solid lines). The electrostatic potential does not exhibit peaks at the atomic positions because the equilibrium potential profile has been subtracted for clarity. Shown for comparison is the linear voltage drop (dashed lines), the solution to Laplace's equation in the region. In (b), a large part of the voltage drop occurs at the defect as expected. In (a) one might ask why the voltage drops across a ballistic device. It is important to recognize the distinction between the electrostatic potential and the electrochemical

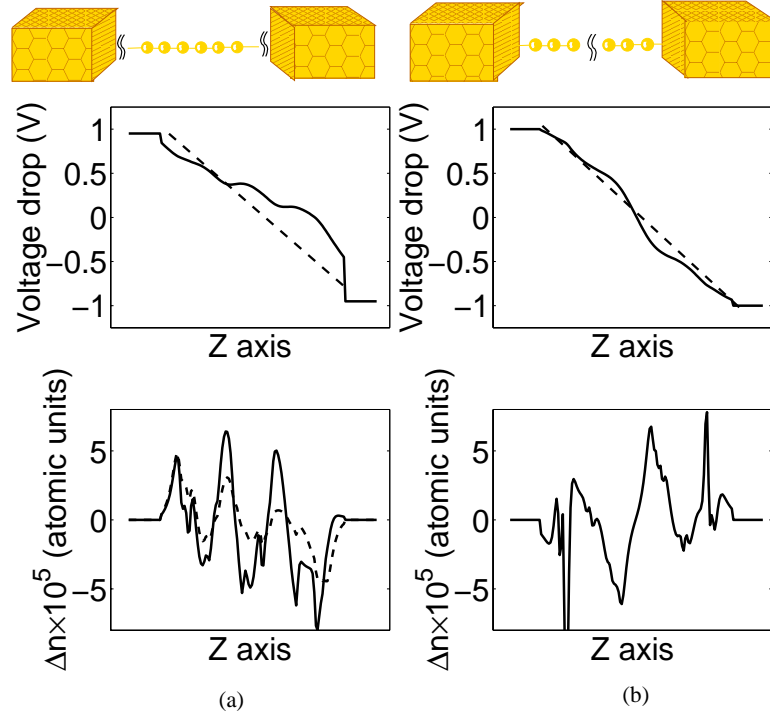


Fig. 5.12. Electrostatics of (a) a weakly contacted gold wire, and (b) a strongly contacted wire with a substitutional defect (a stretched bond at the center), in response to a 2 volt applied bias across the contacts. Top panel: schematic, middle panel: voltage drop along the wire (dashed lines show a linear drop for comparison), bottom panel: density of electrons along the wire. In all plots, the corresponding quantities at equilibrium have been subtracted out in order to remove peaks near the positions of the nuclei for clarity. There is a substantial voltage drop in the wires due to transverse fields penetrating the thin (single atom cross-section) conductors. The largest potential drop is at the barriers (wire-metal interfaces and defects), while the electronic charge piles up against the applied bias as expected. Superposed on this general polarization of charge are Friedel oscillations, which die out (dashed line in left bottom panel) on increasing incoherent scattering in the wire, incorporated through an additional phonon self-energy Σ_p introduced through a Büttiker probe.

potential at this point. In a ballistic device, it is the electrochemical potential that remains constant due to absence of scattering. The electrostatic potential, however, obeys Poisson's equation and may or may not vary depending on the charge density inside the device. For example, a vacuum tube is a ballistic device with a linear voltage (electrostatic potential) drop due to the absence of any charge. The potential is not

well screened out due to the thin (one atom) cross-section of the wire, which allows transverse fields to penetrate. In contrast, in a self-assembled monolayer (SAM), the potential variation is expected to resemble that predicted in [47]. However, this requires us to disable the Hartree term in Gaussian '98 and replace it with our own evaluation.

The transferred number of electrons (Fig 5.12, bottom panels) Δn along the wire obtained from our self-consistent solution shows how the voltage drops develop. In both plots, there are sizeable oscillations in the charge density, which we identify as Friedel oscillations. The equilibrium electron density is $n = 1/d$ for the wire (d is the spacing between atoms in the wire), leading to a 1-D Fermi wave-vector $k_F = n\pi/2 = \pi/2d$. The corresponding Friedel oscillation wavelength is $2\pi/(2k_F) = 2d$, about twice an interatomic spacing, predicting about three oscillation cycles in a six atom device, as we indeed observed. Furthermore, the amplitudes of the oscillations decrease on incorporating an incoherent scattering process in the wire through a Büttiker probe [70] (which shows up as an additional self-energy matrix Σ_p representing coupling of the wire with a phonon-bath, for example). Superposed on the Friedel oscillations is the general trend of the charge distribution, to flow against the applied field (i.e., piling to the left). In (a) this piled up electronic charge screens the applied field and reduces the slope of the voltage drop in the wire. In (b), the charge piles up on each side of the barrier, leading to a charge resistivity dipole in the center. This dipole has a polar field that is now in the direction of the applied field, thus increasing the voltage drop across the barrier .

5.5 Summary

In this chapter we have described a straightforward but rigorous procedure for calculating the I-V characteristics of molecular wires. The Fock matrix is obtained using a Gaussian basis set (LANL2DZ) and the self-consistent potential is obtained using density functional theory with the B3PW91 approximation. This approach is

identical to that used in standard quantum chemistry programs like Gaussian '98. The difference lies in our use of the NEGF formalism to calculate the density matrix from the Fock matrix. This allows us to handle open systems far from equilibrium which are very different from the isolated molecules commonly modeled in quantum chemistry. The close coupling with standard quantum chemistry programs not only makes the procedure simpler to implement but also makes the relation between the I-V characteristics and the chemistry of the molecule more obvious. Partitioning the contact-molecule-contact system into a molecular subspace and a contact subspace allows us to focus on the quantum chemistry of the molecule while exactly taking into account the bonding of the molecule with the contacts, contact surface physics and atomicity etc. We use our method to interpolate between two extreme examples of transport through a molecular wire connected to gold contacts: band conduction in a metallic (gold) nanowire, and resonant conduction through broadened, quasidiscrete levels of a phenyl dithiol molecule. We examine quantities of interest like the I-V characteristic, voltage drop and charge density.

6. MOLECULAR WIRES II: THE “EXTENDED” MOLECULE

In the previous chapter we presented an ab initio treatment of two-terminal molecular wires. The contact-molecule-contact system was partitioned into a molecular subsystem and two contact subsystems. Self-energy functions were used to take into account the effect of contacts on the molecular subsystem. In this chapter we extend that treatment with one significant change made to the molecular subsystem - a few contact gold atoms are now treated as part of the molecule (see Figure 6.1). We will use the term “extended molecule” to denote this extended molecular subsystem. The rest of the contacts are still treated using self-energy functions. In fact, the overall procedure illustrated in Figure 5.1 remains the same, the only change is that the extended molecular Fock matrix is now very big in size (about 700 basis orbitals per extended molecule, or a Fock matrix of size 700×700 , as compared to a size of about 100×100 for non-extended molecules used in the previous chapter).

6.1 Advantages Of Using An Extended Molecule

It is desirable to include a few contact metal (gold, in our case) atoms as part of the device for a number of reasons:

- The molecule may affect a few nearby atoms on the surface of the metallic contact. These surface effects will be automatically accounted for in the self-consistent calculation if a few surface metal atoms are included as part of the device.
- The charges in the molecule are imaged on the metallic contact. These image charges need to be taken into account in order to accurately calculate the self-

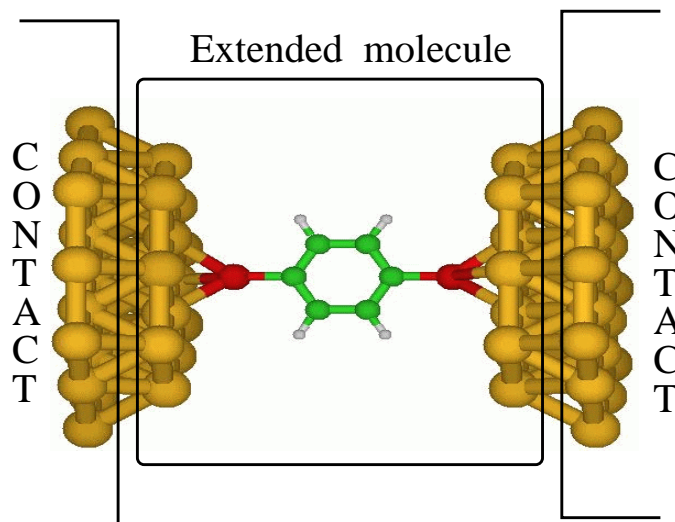


Fig. 6.1. Schematic picture of a Phenyl Dithiol molecule coupled to gold contacts.

The term “extended molecule” is used to denote the molecular subsystem that includes a few contact atoms. For the results shown in this chapter we include 12 gold atoms on each side of the molecule. The rest of the contacts are treated using self-energy matrices. The contact surface Green’s function is evaluated on the 19 gold atoms adjacent to the 12 “molecular” gold atoms as shown. The coupling between the 12 and 19 gold atoms is used as the molecule-contact coupling, and is evaluated once using unmodified Gaussian ’98 and stored for use in subsequent calculations. As an example we have shown above the Phenyl Dithiol extended molecule, with the following color scheme: gold for gold atoms, red for sulfur atoms, green for carbon atoms and white for hydrogen atoms. We will use this color scheme for the rest of this chapter.

consistent potential. It is reasonable to expect that the image charges will reside on a few surface metal atoms that are close to the molecule, and hence the inclusion of these surface atoms in the device should account for the image charge effect on the self-consistent potential.

- Inclusion of contact atoms in the molecule is expected to self-consistently develop the correct amount of electrostatic potential drop across the molecule under applied bias, given two different electrochemical potentials in the two contacts [83]. This eliminates the need to artificially impose fixed boundary

values for the electrostatic potential through externally applied fields etc.¹, as has already been observed in the case of dual-gate nanotransistors (see reference [27] and Section 3.3).

- Density functional theory is traditionally used for a finite system with an integer number of electrons, or periodic systems. Extending DFT to a non-neutral open molecular subsystem with fractional number of electrons is a topic of intense research [84, 85]. However an extended molecule which includes enough metal atoms is effectively charge neutral² and allows the standard DFT formalism to go through.
- Finally, the atoms near the two ends of the molecule will have slightly erroneous charge densities because the device and contact basis functions are not orthogonal to each other and the partitioning of charge leads to ambiguities at the interfaces (see the discussion on Mulliken/Löwdin partitioning in [7] and the discussion on non-orthogonal basis sets in Section 2.5). So it is desirable to ‘pad’ the molecule with a few metal atoms on the two ends so as to allow an accurate calculation of the molecular charge.

6.2 Locating The Fermi Energy

When a molecule is coupled to contacts, the discrete levels of the isolated molecule broaden into a continuous density of states (DOS). The broadening is related to the lifetime of an electron inside the molecule - electrons can flow in or out of the molecule because of the contact coupling. The contact-molecule-contact system eventually

¹All the results presented in this chapter, however, are obtained by applying an external field along the molecule through the Gaussian ’98 “field” option. The main reason for doing this is that we are constrained, due to computational limitations, to using only twelve gold atoms in the top layer of the contact as part of the extended molecule. Further investigation is necessary to establish how many layers of the contacts need to be included before eliminating the need to apply an external field.

²With 12 gold atoms on each side of the molecule, we still obtain a fractional charge in the extended molecule, but as the computational resources improve, we expect to observe a reduction in the fractional charge with enough number of gold atoms included.

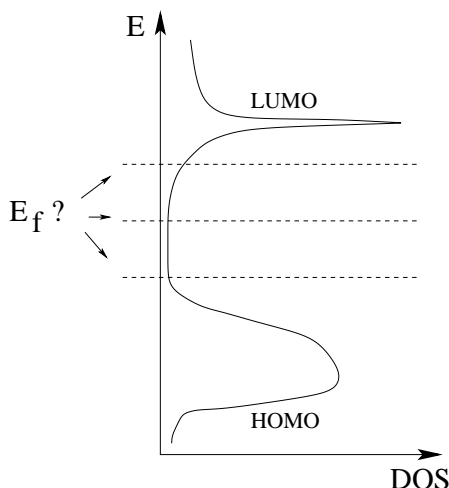


Fig. 6.2. When a molecule is coupled to contacts, there is some charge transfer and the system relaxes to a common Fermi energy E_f at equilibrium. The discrete levels of the molecule broaden into a continuous density of states. When a bias is applied across the molecule, the broadened level closest to the equilibrium Fermi energy carries most of the current, and the conductance gap depends on how far this level is from E_f (see Section 2.1). If the molecule is used in a three-terminal configuration, the location of E_f relative to the HOMO or LUMO band decides if it will function as a p-type or n-type transistor (see Chapter 4).

settles down in a state of equilibrium, and a common Fermi energy dictates the occupancy or otherwise of electronic states in the contacts and the molecule. One important question is where does the equilibrium Fermi energy E_f lie relative to the molecular density of states (see Figure 6.2). The answer to this question is not clear yet, the position of E_f , for a given molecule, seems to sensitively depend on the contact model one uses. For example, in the case of the Phenyl Dithiol (PDT) molecule (Figure 6.1), a jellium model [38] for the contacts predicts that E_f is closer to the LUMO level whereas an extended Hückel theory based model [47] predicts that E_f is closer to the HOMO level (see Figure 6.2). Our ab initio model [17] seems to suggest that E_f lies close to the HOMO level of the PDT molecule, and that is what we used in Chapter 4 (see Figures 4.2, 4.3 and the related discussion).

One would expect an ab initio model to correctly answer the question of locating E_f for a given contact-molecule-contact system. However, considering that E_f

depends very sensitively on the contact model and geometry, and considering the approximations inherent in our ab initio model ³, we believe that we are not yet ready to answer this question in a convincing manner. However, what we can answer is how we fix the location of E_f given our contact model, and that is what we will do in this section.

6.2.1 Choice of the “window”

As discussed in Appendix B, we use an analytical method to evaluate the density matrix, given an energy independent self-energy $\Sigma(E_f)$. We believe that using the $\Sigma(E_f)$ approximation for gold contacts is justified because of two reasons: (1) For bulk gold it is known that around the Fermi energy the local density of states is approximately independent of energy [45], and (2) The transport properties of the molecule are determined by the molecular levels near E_f (labeled as “upper” levels in Figure 6.3) and the core levels contribute to the total number of electrons by remaining full in the bias range of interest.

In view of this, We split the energy range of integration (see Equation 2.25) in two parts and denote the boundary by E_b . In the energy range between E_b to E_f (labeled as the “window” in Figure 6.3) we replace $\Sigma(E)$ by $\Sigma(E_f)$, while in the energy range from $-\infty$ to E_b we replace $\Sigma(E)$ by $-i\eta$, a constant infinitesimal broadening. The location of E_b , or the width of the window, is dictated by two conflicting requirements: (1) E_b should be close to E_f to improve the validity of the $\Sigma(E_f)$ approximation, and at the same time, (2) E_b should lie in an energy range where the density of states is small, for reasons of iterative convergence.

In order to reasonably satisfy the above two conflicting requirements, we have found the following choice of the width of the window to be useful in general. E_b

³The approximations in our ab initio model include: (1) Imposing the gold FCC (111) symmetry (Appendix A), (2) Using an energy independent self-energy $\Sigma(E_f)$ to describe gold contacts (Appendix B), and (3) Including, due to computational constraints, the effect of two nearest neighbors in the calculation of the surface Green’s function, when an accurate calculation needs about five nearest neighbors (Appendix C).

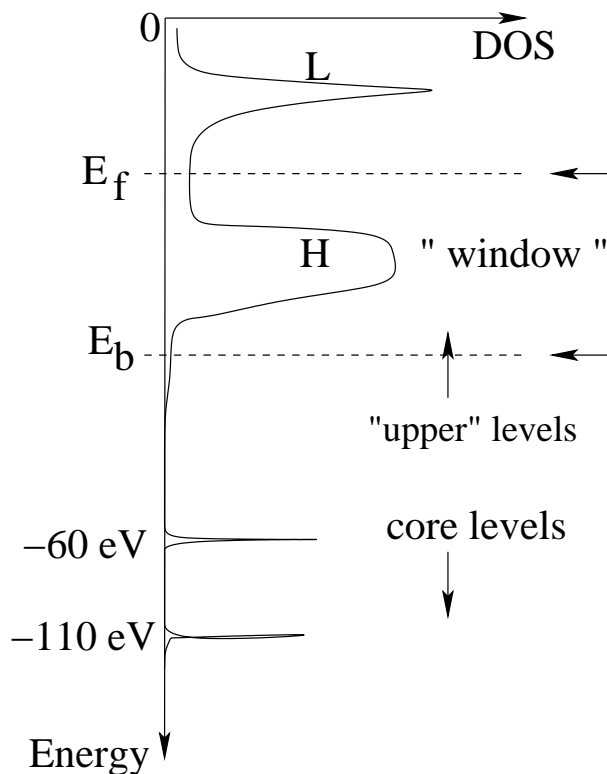


Fig. 6.3. In our analytical (but approximate) method of calculating the density matrix (see appendix B), we replace the energy dependent self-energy $\Sigma(E)$ with an energy independent $\Sigma(E_f)$ for energies lying in the “window” between E_f and E_b . E_b is fixed between one to two electron-volts below the lowest “upper” level. Since the location of E_b affects the location of E_f relative to the molecular density of states, an independent test for our choice of window necessary. The core levels are expected not to influence the transport properties of the molecule, and are accounted for by using a small (one micro-electron-volt) broadening. Typically the lowest upper level is found near -20 eV, while the core levels are found at -60 eV and -110 eV. The HOMO band of levels is denoted by H while the LUMO band is denoted by L .

is chosen such that it lies between one to two electron-volts below the lowest upper level of the extended molecule (see Figure 6.3). Thus having chosen E_b , and using $E_f = -5.1$ eV, we can proceed with the self-consistent equilibrium calculation (see Figure 5.1) and let the molecular density of states float up or down relative to E_f . At the end of this self-consistent calculation, the location of E_f relative to the molecular DOS is known.

6.2.2 Rigorous approach

In this section we propose a rigorous scheme to locate the equilibrium Fermi energy of the contact-molecule-contact system and compare it with the results obtained using the choice of window discussed in the previous section. We illustrate the rigorous approach using the PDT molecule as an example. The PDT molecule is chosen because the HOMO and LUMO wavefunctions of PDT are known to retain their characteristic shapes even after forming a bond with the gold contact. This property of the wavefunctions to retain characteristic shapes after forming bonds with gold contacts is central to our proposed scheme. The step by step procedure is as follows:

1. First, we identify the HOMO (or, equivalently, the LUMO) wavefunction of the isolated molecule shown in Figure 6.4a. This wavefunction is labeled Ψ_a^H .
2. Then we identify, by visual inspection, that wavefunction Ψ_b^H of the isolated extended molecule (Figure 6.4b) which closely resembles Ψ_a^H , as shown in Figure 6.5. The energy eigenvalue of Ψ_b^H is labeled E_b^H . Note that E_b^H may or may not be equal to the HOMO energy eigenvalue of the isolated extended molecule.
3. We then take the Fermi energy E_f to be the mean value of the HOMO and LUMO energies of the isolated extended molecule. For a sufficiently large gold cluster on each side of the molecule, we expect to pin down E_f to a desired accuracy. Due to computational constraints, we are now limited to having about twelve gold atoms on each side in the extended molecule, but in principle we could increase this number. Thus we now know E_f relative to E_b^H . Let $\Delta_1 = E_f - E_b^H$.
4. Figure 6.4c shows the extended molecule coupled to contacts. In equilibrium, the two situations depicted in Figures 6.4b and 6.4c are *exactly* equivalent, provided that the isolated extended molecule includes a sufficiently big number of gold atoms on each side. Now we can set some guess value for E_b (which is the lower boundary of the window, the upper boundary E_f is known from

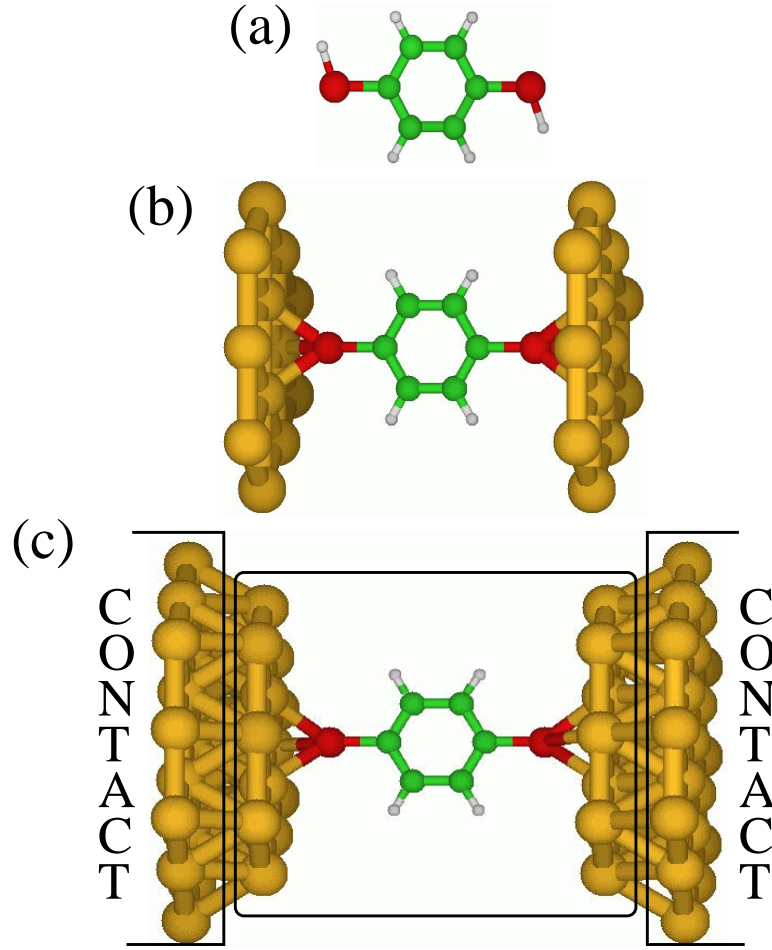


Fig. 6.4. Independent scheme to test the choice of window width. (a) Isolated molecule, (b) Isolated extended molecule and (c) Extended molecule coupled to contacts. For a sufficiently large cluster of contact atoms in the extended molecule, the two cases (b) and (c) are *exactly* equivalent. Calculation of energy levels in case (b) does not involve any E_b , while that in case (c) does. Knowing the HOMO wavefunction Ψ_a^H of the isolated molecule (a), wavefunctions Ψ_b^H and Ψ_c^H that retain the characteristics of Ψ_a^H can be identified. E_f lies between the HOMO and LUMO levels of the isolated extended molecule (b). Knowing the relative distance between E_f and the energy E_b^H of Ψ_b^H , the relative distance between E_f and the energy E_c^H of Ψ_c^H can be obtained for any given E_b . The two distances should be equal for a sufficiently large extended molecule.

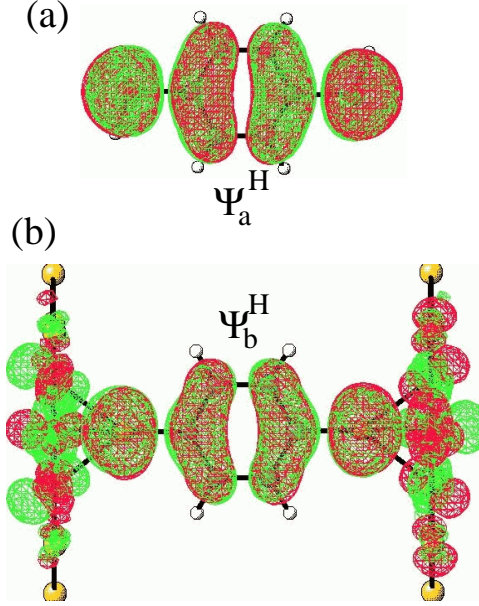


Fig. 6.5. (a) Real space contour plot of the HOMO wavefunction Ψ_a^H of the isolated molecule shown in Figure 6.4a. (b) Real space contour plot of that wavefunction Ψ_b^H of the isolated extended molecule (Figure 6.4b) which most resembles Ψ_a^H . Ψ_b^H is identified by visual inspection of wavefunctions of the isolated extended molecule.

step 3 above) and do an open system equilibrium calculation (Figure 5.1) to obtain the self-consistently converged Fock matrix F for the extended molecule coupled to contacts.

5. We then identify that wavefunction Ψ_c^H of the extended molecule coupled to contacts (represented by $F + \Sigma_1(E_f) + \Sigma_2(E_f)$) which most closely resembles Ψ_b^H . Since the size of all wavevectors in the cases shown in Figures 6.4b and 6.4c are equal, we can obtain Ψ_c^H by maximizing the inner product of Ψ_b^H with all the wavevectors of the extended molecule coupled to contacts, as shown in Figure 6.6. The real part of the energy eigenvalue (which is complex in general, because $\Sigma_{1,2}$ are not Hermitian) corresponding to Ψ_c^H is labeled E_c^H . Let $\Delta_2 = E_f - E_c^H$.

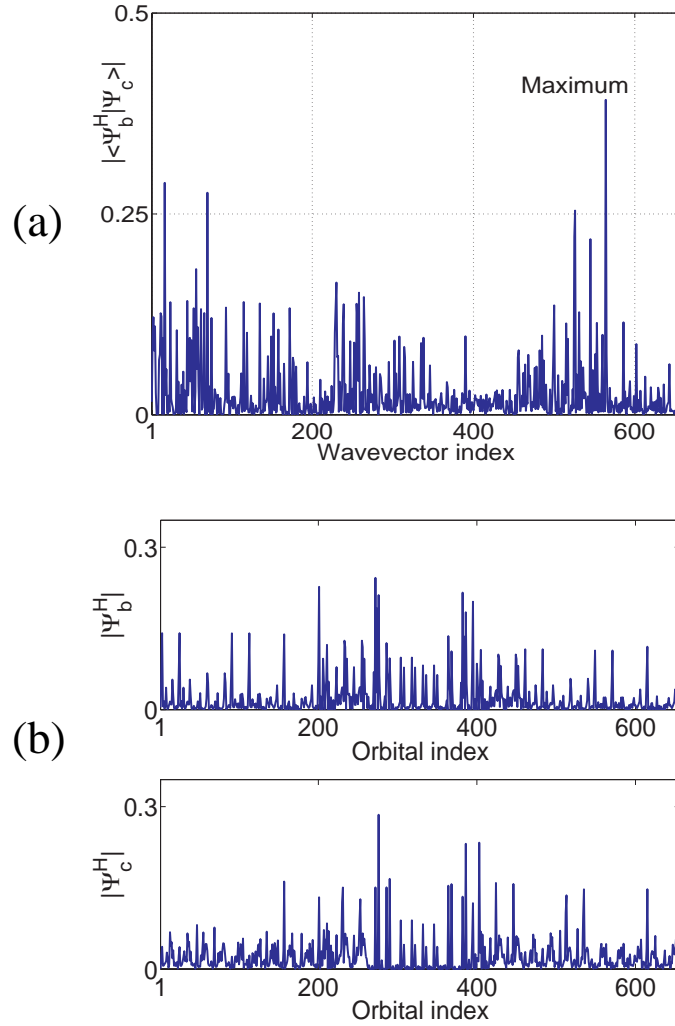


Fig. 6.6. Identifying that wavefunction of the extended molecule coupled to contacts which most resembles the HOMO wavefunction of the isolated molecule. (a) Absolute value of the inner product of Ψ_b^H (see Figure 6.5) with all wavefunctions (eigenvectors) Ψ_c of the matrix $F + \Sigma_1 + \Sigma_2$ (this matrix represents the extended molecule coupled to contacts). Of all the wavevectors Ψ_c , one wavevector Ψ_c^H maximizes the inner product, implying that Ψ_c^H most resembles Ψ_b^H (as shown in (b) above), which in turn is similar to the HOMO wavefunction Ψ_a^H of the isolated molecule (see Figure 6.5). This inner product method of identification is unambiguous because the eigenvectors Ψ_b^H and Ψ_c^H have the same number of components, unlike Ψ_a^H , which is defined on a smaller Hilbert space. The inner product method is necessary because visual identification in real space (like the one shown in Figure 6.5) of wavefunctions of an open system is currently not feasible.

6. If Δ_1 and Δ_2 are equal (to an acceptable accuracy), we accept the value of E_b used in step 4 above. Otherwise, we choose a different E_b and repeat steps 4 and 5 above till Δ_1 and Δ_2 are acceptably close.
7. Finally, we compare this chosen value of E_b with that dictated by our choice of window discussed in the previous section.

We implemented the above procedure for a PDT molecule and for a quantum point contact (QPC) (Figure 6.7). Table 6.1 summarizes the results. The second column in Table 6.1 shows the values of $\Delta_1 = E_f - E_b^H$, and the third column shows the corresponding values of $\Delta_2 = E_f - E_c^H$. The fourth column shows the guess E_b values (see step 4 above) used to obtain the second and third columns - these guess E_b values were obtained from the choice of window discussed in the previous section. We see that Δ_1 and Δ_2 end up within 0.5 eV of each other, giving us confidence in the choice of the window width. The results presented in Table 6.1 were obtained using twelve gold atoms on each side in the extended molecule. An accuracy better than 0.5 eV is expected with bigger gold clusters.

For the case of the PDT molecule, Δ_1 seems to depend on whether we use twelve or thirteen gold atom clusters in the isolated extended molecule. This discrepancy may be because when three gold atoms form a bond with the sulfur, the remaining ten gold atoms form a closed shell system (in the LANL2DZ basis, gold atom has nineteen valence electrons) when thirteen atoms are present in the cluster. When twelve atoms are present in the cluster, the remaining nine gold atoms do not form a closed shell system, which may lead to some spurious level shifts. This effect is not expected in the case of the QPC, because of the absence of a gold-sulfur bond. However, this odd-even discrepancy needs further investigation before any concrete conclusions may be drawn. In any case, with a sufficiently large number of gold atoms in the cluster, we expect these odd-even level shifts to be practically negligible.

Table 6.1

Results of test for the choice of window to determine the location of E_b . *PDT* denotes Phenyl Dithiol (Figure 6.1) while *QPC* denotes quantum point contact, which is a linear chain of six gold atoms (Figure 6.7). Au_n denotes that n gold atoms were present on each side in the extended molecule. E_b was adjusted till $E_f - E_b^H$ and $E_f - E_c^H$ came within about 0.5 eV of each other, more accuracy needs a bigger gold cluster on each side in the extended molecule. The results of the calculation with the isolated extended *PDT* molecule (Figure 6.1) depend on whether we use $n = 12$ (row (1) below) or $n = 13$ (row (2) below) gold atoms. We expect this odd-even discrepancy to become smaller and smaller as the number n of gold atoms increases. The final adjusted E_b does lie slightly below the lowest upper levels (Figure 6.3), thus giving us confidence in our choice of the window width.

	$E_f - E_b^H$ (eV)	$E_f - E_c^H$ (eV)	E_b (eV)
(1) $Au_{12} - PDT - Au_{12}$	-0.20	0.25	-25
(2) $Au_{13} - PDT - Au_{13}$	0.40	0.25	-25
(3) $Au_{12} - QPC - Au_{12}$	0.42	0.26	-15

6.3 I-V Characteristics

We can now proceed with the calculation of I-V characteristics with extended molecules. Figure 5.11 shows the I-V and G-V characteristic of the non-extended PDT molecule. We see a significant conductance gap because E_f is about 2 eV above the HOMO level in that case. However, from Table 6.1 we notice that when an extended molecule is used, E_f is much closer to the HOMO level, which will lead to an almost ohmic I-V with no conductance gap. Such an ohmic I-V for an extended PDT molecule has indeed been calculated by another group [86]. This difference in the I-V characteristic with and without an extended molecule only underscores the extreme sensitivity of the location of E_f to the contact model and geometry one uses. Another issue with the PDT molecule that has recently been brought to our attention ⁴ is the poor quality of self-assembled monolayer (SAM) formed with PDT on a gold substrate, which further complicates finding the location of E_f . For the

⁴Laurie Calvet, Department of Chemistry, Harvard University, private communication, 2002.

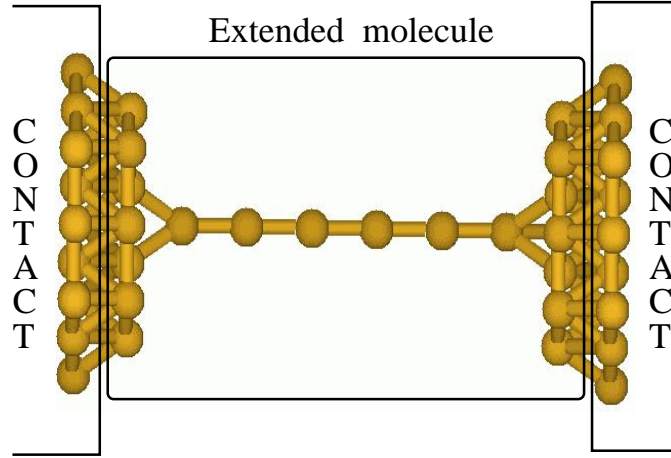


Fig. 6.7. Schematic picture of a quantum point contact (a chain of six gold atoms) with twelve contact atoms included in the extended molecule.

above reasons, instead of PDT, we focus on another molecule (Xylyl Dithiol, XYL for short) which is known to form good quality SAMs ⁵.

6.3.1 Quantum point contact

As was done in the previous chapter, we first test our self-energy matrices to see if we get perfect quantum unit conductance of $2e^2/h$ with the QPC extended molecule. Then we just replace the chain of gold atoms in the QPC with the XYL molecule and calculate the I-V characteristic of XYL.

Figure 6.7 shows a schematic picture of the QPC extended molecule. The transmission through the chain of gold atoms is one near the Fermi energy (Figure 6.8) leading to an ohmic I-V with quantum unit conductance (Figure 6.9).

⁵Unlike PDT, XYL does not have π -conjugation due to the presence of a $-CH_2$ group between the phenyl ring and the thiol ($-SH$) group. Also, the wavefunctions of XYL do not retain their characteristics upon bonding with the gold contact. Hence we are unable to apply the rigorous procedure to locate E_f (Subsection 6.2.2). However, since the choice of window (Subsection 6.2.1) has been tested with PDT and QPC, we can apply it to XYL.

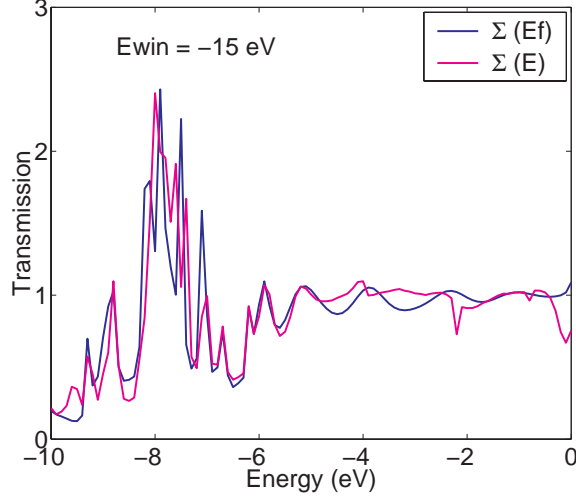


Fig. 6.8. Transmission of a quantum point contact as a function of energy. Blue line is a self-consistent calculation with $\Sigma(E_f)$, while the magenta line is a non self-consistent calculation (Fock matrix obtained from the $\Sigma(E_f)$ self-consistent calculation) using $\Sigma(E)$. The Fermi energy E_f is at -5.1 eV. The transmission is close to one over a range of energies around E_f , giving rise to a quantized conductance of $2e^2/h$ as expected (Figure 6.9). The close agreement with $\Sigma(E)$ calculation reemphasizes the validity of the $\Sigma(E_f)$ approximation (also see Figure 5.4).

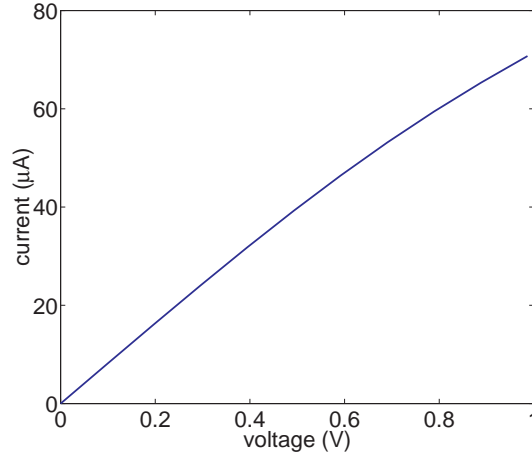


Fig. 6.9. I-V characteristic of a quantum point contact (compare with Figure 5.7a). The close agreement with the perfect quantum conductance $2e^2/h = 80 \mu A/V$ provides support for the accuracy of the calculation of the self-energy matrices $\Sigma_{1,2}(E_f)$.

6.3.2 Xylyl Dithiol: symmetric coupling

Now we can replace the chain of gold atoms with the XYL molecule, as shown in Figure 6.10. The I-V characteristic is shown in Figure 6.11, and the calculated and experimental G-V is shown in Figure 6.12. The shape of the calculated G-V nicely matches the shape of the experimental G-V, however, the overall magnitude of the calculated G-V was scaled down by a factor of one thousand. This orders of magnitude discrepancy in the theoretical and experimental conductance values is well known in the literature [37, 38, 47] but no satisfactory explanation exists yet, at least for STM experiments.

We can list three possible scenarios in which the magnitude of the calculated current will be close to the experimentally measured values:

1. The distance between the sulfur atom and the gold contact is larger than the usually accepted distance of about 2 Å. We have done some calculations with larger distances, but we find that the current decreases very slowly with increasing distance (about two orders of magnitude decrease with five to ten angstroms increase in distance, as is also noted in [87]). Furthermore, if the contact is so far away as to not broaden some molecular levels, then we get into the coulomb blockade regime and consequently, convergence problems.
2. One contact is farther away than the other, as is typically the case in STM experiments. In this case, it is more difficult to get into the coulomb blockade regime because of the close proximity with one contact. But the current in this case is not symmetric with respect to the polarity of the applied voltage, contrary to the mostly symmetric I-Vs measured in STM experiments [47]. This scenario is discussed in more detail in the following subsection.
3. Contrary to the conventionally accepted picture of the sulfur atom sitting in the middle of a triangle formed by three gold atoms (Figure 6.10), if the sulfur atom sits directly on top of a gold atom, then the sulfur p_z orbital has zero overlap

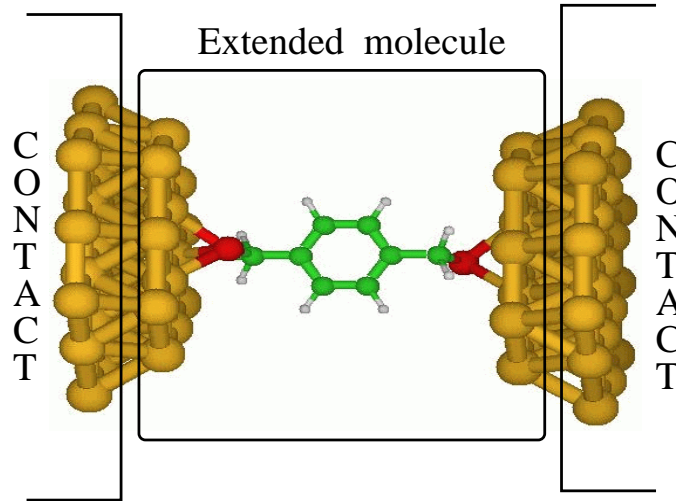


Fig. 6.10. Schematic picture of the Xylyl Dithiol extended molecule. For an explanation of the color scheme see the caption to Figure 6.1.

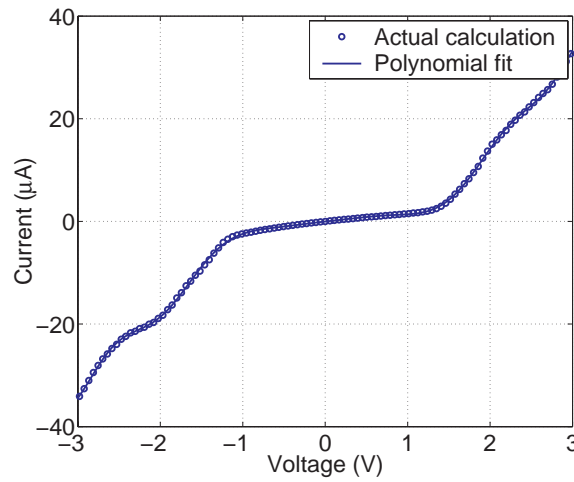


Fig. 6.11. I-V characteristic of Xylyl Dithiol. Small asymmetry in the current is due to small numerical asymmetries in the computation of the self-energy matrices Σ_1 and Σ_2 . The polynomial fit smoothens the data thereby reducing noise in the calculation of the derivative dI/dV (see Figure 6.12).

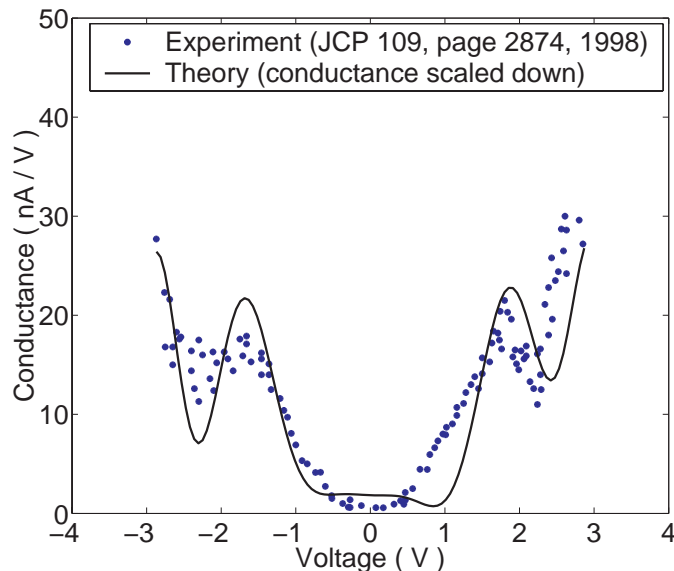


Fig. 6.12. Conductance of Xylyl Dithiol as a function of applied voltage. The conductance is calculated by numerically differentiating the polynomial fit to the I-V (see Figure 6.11). Symbols represent experimental data from [47]. To obtain a fit to the experiment, all calculated conductance values were scaled down by one thousand. This orders of magnitude discrepancy in the theoretical and experimental conductance values is well known in the literature [37, 38, 47] but no satisfactory explanation exists yet, at least for STM experiments.

with the gold s orbital which will reduce the current [38, 81]. This scenario is the most attractive because it is expected to lead to small magnitudes of current in conjunction with symmetric I-Vs. We do not explore this scenario here, but recommend it for further investigation.

6.3.3 Xylyl Dithiol: asymmetric coupling

The second scenario listed above is discussed in more detail in this subsection. The strong coupling of the extended molecule with the left contact (see Figure 6.13) is modeled using the self-energy $\Sigma_1(E_f)$, while the weak coupling with the right contact is modeled using a small (50 meV) broadening on the sulfur and hydrogen

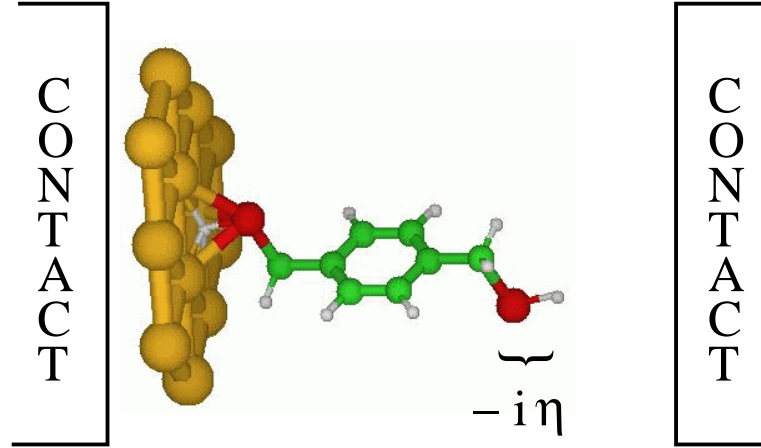


Fig. 6.13. Schematic picture of a Xylyl Dithiol molecule asymmetrically coupled to two contacts. The strongly coupled contact is modeled using the self-energy matrix $\Sigma_1(E_f)$, while the weakly coupled contact is modeled using an artificial small (50 meV) broadening on the sulfur and hydrogen orbitals on the right side.

atom orbitals on the right side. The I-V and G-V characteristics of XYL with this asymmetric coupling to contacts are shown in Figure 6.14.

The reason behind the asymmetry of the I-V and G-V characteristics with respect to the applied voltage polarity is as follows (see Figure 6.15). The Fermi energy at equilibrium lies near the HOMO level. When a positive voltage is applied to the strongly coupled contact, (the substrate) the electrons in the HOMO level try to escape into the substrate. Since the HOMO level is normally full, charging effects kick in and to recover the lost charge, the HOMO moves down and tries to follow the electrochemical potential of the substrate. Due to this charging effect, the current does not increase with positive voltage as fast as it does with negative voltage. This is because when a negative voltage is applied to the substrate, electrons in the HOMO level try to escape into the weakly coupled contact. But due to the weak coupling, the electrons cannot escape easily and the HOMO level remains filled. Due to this, there is no charging effect as before, and the current rises quickly as the HOMO level gets caught between the left and right contact electrochemical potentials. A more

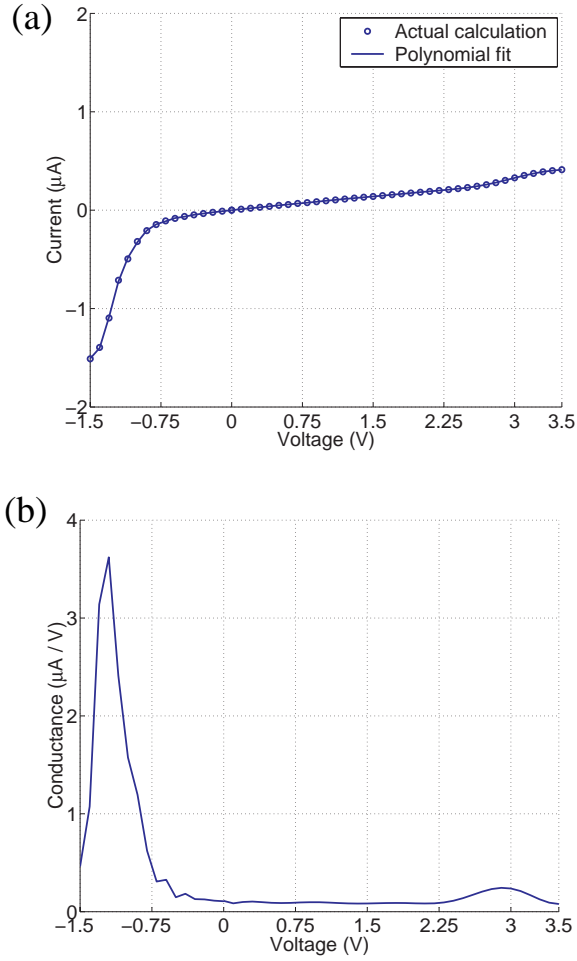


Fig. 6.14. I-V and G-V characteristics of a Xylyl Dithiol molecule asymmetrically coupled to two contacts. (a) I-V characteristic, and (b) G-V characteristic obtained by numerically differentiating the polynomial fit to the I-V. When a positive voltage is applied to the strongly coupled contact, current is suppressed due to charging effects [88].

detailed discussion about such I-V asymmetry arising due to charging effects can be found in [88].

6.4 Summary And Outlook

In this chapter, we have extended the ab initio treatment of molecular wires presented in the previous chapter with one significant change made to the molecular

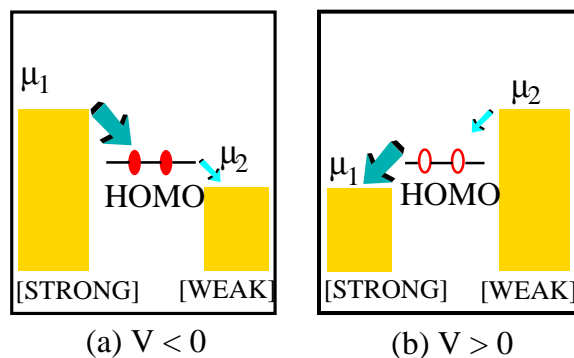


Fig. 6.15. Origin of I-V asymmetry lies in charging effect. One side (substrate) is strongly coupled to the molecule, while the other is weakly coupled. Although the same HOMO level is crossed by the contact electrochemical potentials μ_1 and μ_2 , for positive substrate voltage (b), the HOMO level is emptied by the strongly coupled contact, which positively charges the molecule and shifts the energy levels down. Such a shift, not present for negative substrate voltage (a), postpones the onset of conduction and effectively stretches out the voltage axis in the direction of positive substrate voltage. See [88] for a more detailed discussion on I-V asymmetry due to charging effects.

subsystem - a few contact gold atoms are now treated as part of the molecule. The overall procedure illustrated in Figure 5.1 remains the same, the only change is that the extended molecular Fock matrix is now very big in size (about 700 basis orbitals per extended molecule, or a Fock matrix of size 700×700 , as compared to a size of about 100×100 for non-extended molecules used in the previous chapter). The results presented in this chapter were generated using the parallelized version of Gaussian '98 [16] running on a Linux Beowulf compute cluster with two hundred nodes.

In this and the previous chapter we have presented a rigorous model to study current flow through individual molecules coupled to gold contacts. Extension of this model to three-terminal molecular devices is conceptually straightforward, and is left to future work. Other extensions of this model may include relaxing the approximations made while modeling the gold contacts (which include: (1) Imposing the gold FCC (111) symmetry (Appendix A), (2) Using an energy independent self-energy $\Sigma(E_f)$ to describe gold contacts (Appendix B), and (3) Including, due to computa-

tional constraints, the effect of two nearest neighbors in the calculation of the surface Green's function, when an accurate calculation needs about five nearest neighbors (Appendix C).), especially the approximation of an energy independent self-energy function. If a calculation using the full energy dependent self-energy is done, it will allow us to locate the position of the Fermi energy relative to the molecular density of states more reliably than the current calculation (Section 6.2). Furthermore, and perhaps more importantly, an energy independent self-energy will also allow us to treat contact materials other than gold, like silicon. Treatment of molecules coupled to silicon contacts is important because of the tremendous infrastructure already available to a technology that involves silicon - molecular devices that complement silicon technology are more attractive than those that try to replace silicon technology [89, 90].

Another possible area of future work may involve modeling the conformational degree of freedom of molecules [42]. In this dissertation we have assumed that molecules are "rigid", that they do not deform under the influence of external fields. However, the "soft" nature of molecules is believed to play a role in some experiments [4] involving switching and negative differential resistance, and we need to be able to model such conformational effects.

Overall, in this dissertation we have put forth a rigorous yet practical approach to treat nanoscale devices of the future. The basic procedure (Section 1.2) can be used with semi-empirical (and fast) models as well as *ab initio* (but slow) ones. Our approach is based on the non-equilibrium Green's function formalism which is rapidly gaining acceptance as the method of choice to treat quantum transport in nanostructures. We have demonstrated the power of this approach by applying it with equal ease to diverse systems ranging from silicon transistors to organic molecules.

LIST OF REFERENCES

- [1] D. Frank, R. Dennard, E. Nowak, P. Solomon, Y. Taur, and H.-S. Wong, "Device scaling limits of Si MOSFETs and their application dependencies," *Proc. of the IEEE*, vol. 89, pp. 259–288, 2001.
- [2] M. Reed, "Molecular-scale electronics," *Proc. of the IEEE*, vol. 87, pp. 652–658, 1999.
- [3] C. Collier, E. Wong, M. Belohradsky, F. Raymo, J. Stoddart, P. Kuekes, R. Williams, and J. Heath, "Electronically configurable molecular-based logic gates," *Science*, vol. 285, pp. 391–394, 1999.
- [4] J. Chen, M. Reed, A. Rawlett, and J. Tour, "Large on-off ratios and negative differential resistance in a molecular electronic device," *Science*, vol. 286, pp. 1550–1552, 1999.
- [5] G. Brown and B. Lindsay, "The numerical solution of Poisson's equation for two-dimensional semiconductor devices," *Solid State Electron.*, vol. 19, pp. 991–992, 1976.
- [6] R. E. Bank and D. J. Rose, "Global approximate Newton methods," *Numerische Mathematik*, vol. 37, pp. 279–295, 1981.
- [7] A. Szabo and N. Ostlund, *Modern Quantum Chemistry*. Mineola, New York: Dover Publications Inc., 1996.
- [8] F. Jensen, *Introduction to Computational Chemistry*. New York: Wiley, 1999.
- [9] G. Mahan, "Quantum transport equation for electric and magnetic fields," *Phys. Reports*, vol. 145, pp. 251–318, 1987.
- [10] P. Danielewicz, "Quantum theory of nonequilibrium processes .1., " *Annals of Phys.*, vol. 152, pp. 239–304, 1984.
- [11] S. Datta, *Electronic Transport in Mesoscopic Systems*. Cambridge, England: Cambridge University Press, 1995.
- [12] S. Datta, "Nanoscale device modeling: the Green's function method," *Superlattices and Microstructures*, vol. 28, pp. 253–278, 2000.
- [13] M. Paulsson, F. Zahid, and S. Datta, "Resistance of a molecule," in *Handbook of Nanoscience, Engineering and Technology* (W.A. Goddard III *et al.*, ed.), Boca Raton, Florida: CRC Press, 2002.
- [14] P. Damle, A. Ghosh, and S. Datta, "Nanoscale device modeling," in *Molecular Nanoelectronics* (M. Reed and T. Lee, eds.), Stevenson Ranch, California: American Scientific Publishers, 2003.

- [15] P. Damle, M. Paulsson, T. Rakshit, and S. Datta, "Current-voltage characteristics of molecular conductors: two versus three terminal," *IEEE Trans. on Nanotechnology*, vol. 1, pp. 145–153, 2002.
- [16] Gaussian 98, Revision A.11.3, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, N. Rega, P. Salvador, J. J. Dannenberg, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian, Inc., Pittsburgh PA, 2002 .
- [17] P. Damle, A. Ghosh, and S. Datta, "Unified description of molecular conduction: from molecules to metallic wires," *Phys. Rev. B (Rapid Comm.)*, vol. 64, p. 201403, 2001.
- [18] P. Damle, A. Ghosh, and S. Datta, "First-principles analysis of molecular conduction using quantum chemistry software," *Chem. Phys.*, vol. 281, pp. 171–187, 2002.
- [19] C. Caroli, R. Combescot, P. Nozieres, and D. Saint-James, "Direct calculation of the tunnelling current," *J. Phys. C*, vol. 4, pp. 916–929, 1971.
- [20] S. Datta, *Quantum Phenomena*. Boston, Massachusetts: Modular Series on Solid State Devices, Vol. 8, Addison-Wesley, 1989.
- [21] S. Datta, W. Tian, S. Hong, R. Reifenberger, J. I. Henderson, and C. Kubiak, "Current-voltage characteristics of self-assembled monolayers by scanning tunneling microscopy," *Phys. Rev. Lett.*, vol. 79, pp. 2530–2533, 1997.
- [22] M. Reed, C. Zhou, C. Muller, T. Burgin, and J. Tour, "Conductance of a molecular junction," *Science*, vol. 278, pp. 252–254, 1997.
- [23] R. Haydock, V. Heine, and M. Kelly, "Electronic structure based on the local atomic environment for tight-binding bands," *J. Phys. C*, vol. 5, pp. 2845–2858, 1972.
- [24] E. Mele and J. Joannopoulos, "Electronic states at unrelaxed and relaxed GaAs (110) surfaces," *Phys. Rev. B*, vol. 17, pp. 1816–1827, 1978.
- [25] M. L. Sancho, J. L. Sancho, and J. Rubio, "Quick iterative scheme for the calculation of transfer matrices: application to Mo(100)," *J. Phys. F*, vol. 14, pp. 1205–1215, 1984.
- [26] M. Samanta, "Electronic conduction through organic molecules," Master's thesis, Purdue University, 1995.
- [27] Z. Ren, R. Venugopal, S. Datta, M. Lundstrom, D. Jovanovic, and J. Fossum, "The ballistic nanotransistor: A simulation study," *IEDM Tech. Dig.*, pp. 715–718, 2000.

- [28] Y. Naveh and K. Likharev, "Modeling of 10-nm-scale ballistic MOSFETs," *IEEE Electron Device Lett.*, vol. 21, pp. 242–244, 2000.
- [29] Z. Ren, R. Venugopal, S. Datta, and M. Lundstrom, "Examination of design and manufacturing issues in a 10 nm double gate MOSFET using nonequilibrium Green's function simulation," *IEDM Tech. Dig.*, pp. 107–110, 2001.
- [30] R. Venugopal, M. Paulsson, S. Goasguen, S. Datta, and M. Lundstrom, "A simple quantum mechanical treatment of scattering in nanoscale transistors," *J. Appl. Phys.*, 2003. Accepted for publication.
- [31] R. Pierret, *Advanced Semiconductor Fundamentals*. Boston, Massachusetts: Modular Series on Solid State Devices, Vol. 6, Addison-Wesley, 1989.
- [32] M. Lundstrom, "Elementary scattering theory of the Si MOSFET," *IEEE Electron Device Lett.*, vol. 18, pp. 361–363, 1997.
- [33] M. Lundstrom and Z. Ren, "Essential physics of carrier transport in nanoscale MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, pp. 133–141, 2002.
- [34] J. Reichert, R. Ochs, D. Beckmann, H. Weber, M. Mayor, and H. Löhneysen, "Driving current through single organic molecules," *Phys. Rev. Lett.*, vol. 88, p. 176804, 2002.
- [35] M. D. Ventra, S. Pantelides, and N. Lang, "The benzene molecule as a molecular resonant-tunneling transistor," *Appl. Phys. Lett.*, vol. 76, pp. 3448–3450, 2000.
- [36] E. Emberly and G. Kirczenow, "Multiterminal molecular wire systems: A self-consistent theory and computer simulations of charging and transport," *Phys. Rev. B*, vol. 62, pp. 10451–10458, 2000.
- [37] E. Emberly and G. Kirczenow, "Models of electron transport through organic molecular monolayers self-assembled on nanoscale metallic contacts," *Phys. Rev. B*, vol. 64, p. 235412, 2001.
- [38] M. D. Ventra, S. Pantelides, and N. Lang, "First-principles calculation of transport properties of a molecular device," *Phys. Rev. Lett.*, vol. 84, pp. 979–982, 2000.
- [39] M. Brandbyge, J. Mozos, P. Ordejón, J. Taylor, and K. Stokbro, "Density-functional method for nonequilibrium electron transport," *Phys. Rev. B*, vol. 65, p. 165401, 2002.
- [40] J. Palacios, A. P. Jiménez, E. Louis, and J. Vergés, "Fullerene-based molecular nanobridges: A first-principles study," *Phys. Rev. B*, vol. 64, p. 115411, 2001.
- [41] Y. Taur and T. Ning, *Fundamentals of VLSI Devices*. Cambridge, England: Cambridge University Press, 1998.
- [42] A. Ghosh, T. Rakshit, and S. Datta, "Gating of a molecular transistor: electrostatic and conformational." Online archive - <http://xxx.lanl.gov/abs/cond-mat/0212166>, 2002.
- [43] M. Paulsson and S. Stafström, "Self-consistent-field study of conduction through conjugated molecules," *Phys. Rev. B*, vol. 64, p. 035416, 2001.

- [44] R. Saito, G. Dresselhaus, and M. Dresselhaus, *Physical Properties of Carbon Nanotubes*. London, England: Imperial College Press, 1998.
- [45] D. Papaconstantopoulos, *Handbook of the Band Structure of Elemental Solids*. New York: Plenum Press, 1986.
- [46] N. B. Larsen, H. Biebuyck, E. Delamarche, and B. Michel, "Order in microcontact printed self-assembled monolayers," *J. Am. Chem. Soc.*, vol. 119, pp. 3017–3026, 1997.
- [47] W. Tian, S. Datta, S. Hong, R. Reifengerger, J. I. Henderson, and C. Kubiak, "Conductance spectra of molecular wires," *J. Chem. Phys.*, vol. 109, pp. 2874–2882, 1998.
- [48] M. Rampi, O. Schueller, and G. Whitesides, "Alkanethiol self-assembled monolayers as the dielectric of capacitors with nanoscale thickness," *Appl. Phys. Lett.*, vol. 72, pp. 1781–1783, 1998.
- [49] A. Yazdani, D. Eigler, and N. Lang, "Off-resonance conduction through atomic wires," *Science*, vol. 272, pp. 1921–1924, 1996.
- [50] C. Zhou, M. Deshpande, M. Reed, L. Jones, and J. Tour, "Nanoscale metal/self-assembled monolayer/metal heterostructures," *Appl. Phys. Lett.*, vol. 71, pp. 611–613, 1997.
- [51] R. Andres, T. Bein, M. Dorogi, S. Feng, J. Henderson, C. Kubiak, W. Mahoney, R. Osifchin, and R. G. Reifengerger, "'Coulomb staircase' at room temperature in a self-assembled molecular nanostructure," *Science*, vol. 272, pp. 1323–1325, 1996.
- [52] U. Durig, O. Zuger, B. Michel, L. Haussling, and H. Ringsdorf, "Electronic and mechanical characterization of self-assembled alkanethiol monolayers by scanning tunneling microscopy combined with interaction-force-gradient sensing," *Phys. Rev. B*, vol. 48, pp. 1711–1717, 1993.
- [53] C. Boudas, J. Davidovits, F. Rondelez, and D. Vuillaume, "Suppression of charge carrier tunneling through organic self-assembled monolayers," *Phys. Rev. Lett.*, vol. 76, pp. 4797–4800, 1996.
- [54] H. Ohnishi, Y. Kondo, and K. Takayanagi, "Quantized conductance through individual rows of suspended gold atoms," *Nature*, vol. 395, pp. 780–783, 1998.
- [55] A. Yanson, G. Bollinger, H. van den Brom, N. Agrait, and J. van Ruitenbeek, "Formation and manipulation of a metallic wire of single gold atoms," *Nature*, vol. 395, pp. 783–785, 1998.
- [56] C. Dekker, "Carbon nanotubes as molecular quantum wires," *Phys. Today*, vol. 52, pp. 22–28, 1999.
- [57] H. Fink and C. Schonenberger, "Electrical conduction through DNA molecules," *Nature*, vol. 398, pp. 407–410, 1999.
- [58] D. Porath, A. Bezryadin, S. de Vries, and C. Dekker, "Direct measurement of electrical transport through DNA molecules," *Nature*, vol. 403, pp. 635–638, 2000.

- [59] M. Ratner, "Photochemistry: Electronic motion in DNA," *Nature*, vol. 397, pp. 480–481, 1999.
- [60] E. Emberly and G. Kirczenow, "Theoretical study of electrical conduction through a molecule connected to metallic nanocontacts," *Phys. Rev. B*, vol. 58, pp. 10911–10920, 1998.
- [61] L. E. Hall, J. R. Reimers, N. S. Hush, and K. Silverbrook, "Formalism, analytical model, and a priori Green's-function-based calculations of the current-voltage characteristics of molecular wires," *J. Chem. Phys.*, vol. 112, pp. 1510–1521, 2000.
- [62] S. Yaliraki, A. Roitberg, C. Gonzalez, V. Mujica, and M. Ratner, "The injecting energy at molecule/metal interfaces: Implications for conductance of molecular junctions from an ab initio molecular description," *J. Chem. Phys.*, vol. 111, pp. 6997–7002, 1999.
- [63] V. Mujica, A. Roitberg, and M. Ratner, "Molecular wire conductance: Electrostatic potential spatial profile," *J. Chem. Phys.*, vol. 112, pp. 6834–6839, 2000.
- [64] C. Tsu and R. Marcus, "A sequential formula for electronic coupling in long range bridge-assisted electron transfer: Formulation of theory and application to alkanethiol monolayers," *J. Chem. Phys.*, vol. 106, pp. 584–598, 1997.
- [65] M. Magoga and C. Joachim, "Conductance and transparency of long molecular wires," *Phys. Rev. B*, vol. 56, pp. 4722–4729, 1997.
- [66] A. Onipko, "Analytical model of molecular wire performance: A comparison of π and σ electron systems," *Phys. Rev. B*, vol. 59, pp. 9995–10006, 1999.
- [67] N. Lang and P. Avouris, "Carbon-atom wires: Charge-transfer doping, voltage drop, and the effect of distortions," *Phys. Rev. Lett.*, vol. 84, pp. 358–361, 2000.
- [68] B. Larade, J. Taylor, H. Mehrez, and H. Guo, "Conductance, i-v curves, and negative differential resistance of carbon atomic wires," *Phys. Rev. B*, vol. 64, p. 075420, 2001.
- [69] J. Seminario, A. Zacarias, and J. Tour, "Molecular current-voltage characteristics," *J. Phys. Chem. A*, vol. 103, pp. 7883–7887, 1999.
- [70] M. Büttiker, "Coherent and sequential tunneling in series barriers," *IBM J. Res. Dev.*, vol. 32, pp. 63–75, 1988.
- [71] X.-Q. Li and Y. Yan, "Electrical transport through individual DNA molecules," *Appl. Phys. Lett.*, vol. 79, pp. 2190–2192, 2001.
- [72] Y. Xue, S. Datta, and M. Ratner, "Charge transfer and "band lineup" in molecular electronic devices: A chemical and numerical interpretation," *J. Chem. Phys.*, vol. 115, pp. 4292–4299, 2001.
- [73] P. Hay and W. Wadt, "Ab initio effective core potentials for molecular calculations. Potentials for the transition metal atoms Sc to Hg," *J. Chem. Phys.*, vol. 82, pp. 270–283, 1985.

- [74] P. Hay and W. Wadt, "Ab initio effective core potentials for molecular calculations. Potentials for main group elements Na to Bi," *J. Chem. Phys.*, vol. 82, pp. 284–298, 1985.
- [75] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," *J. Chem. Phys.*, vol. 98, pp. 5648–5652, 1993.
- [76] J. P. Perdew, "Unified theory of exchange and correlation beyond the local density approximation," in *Electronic structure of solids* (P. Ziesche and H. Eschrig, eds.), pp. 11–20, Berlin, Germany: Akademik Verlag, 1991.
- [77] N. Camillone, C. E. D. Chidsey, G. Y. Liu, and G. Scoles, "Superlattice structure at the surface of a monolayer of octadecanethiol self-assembled on Au(111)," *J. Chem. Phys.*, vol. 98, pp. 3503–3511, 1993.
- [78] Y. Xue, *Physics of molecular electronic devices: electronic structures and transport properties*. PhD thesis, Purdue University, 2000.
- [79] R. Zeller, J. Deutz, and P. Dederichs, "Application of complex energy integration to selfconsistent electronic structure calculations," *Solid State Comm.*, vol. 44, pp. 993–997, 1982.
- [80] J. Tersoff, "Theory of semiconductor heterojunctions: The role of quantum dipoles," *Phys. Rev. B*, vol. 30, pp. 4874–4877, 1984.
- [81] S. Datta, D. Janes, R. Andres, C. Kubiak, and R. Reifengerger, "Molecular ribbons," *Semicond. Sci. Tech.*, vol. 13, pp. 1347–1353, 1998.
- [82] K. Seshadri and C. Frisbie, "Potentiometry of an operating organic semiconductor field-effect transistor," *Appl. Phys. Lett.*, vol. 78, pp. 993–995, 2001.
- [83] E. Louis, J. Vergés, J. Palacios, A. P. Jiménez, and E. SanFabian, "Implementing the Keldysh formalism into the ab initio Gaussian Embedded Cluster Method for the calculation of quantum transport." Online archive - <http://xxx.lanl.gov/abs/cond-mat/0212115>, 2002.
- [84] W. Yang, Y. Zhang, and P. Ayers, "Degenerate ground states and a fractional number of electrons in density and reduced density matrix functional theory," *Phys. Rev. Lett.*, vol. 84, pp. 5172–5175, 2000.
- [85] V. Russier, "Extensions of the density-functional theory (DFT) to systems with a fractional number of electrons: Consequences for the meaning of the DFT energy levels," *Phys. Rev. B*, vol. 45, pp. 8894–8901, 1992.
- [86] K. Stokbro, J. Taylor, M. Brandbyge, J.-L. Mozos, and P. Ordejón, "Theoretical study of the nonlinear conductance of Di-thiol benzene coupled to Au(111) surfaces via thiol and thiolate bonds," *Computational Materials Science*, 2003. In press.
- [87] T. Lee, *Electronic properties of Au nanoclusters/semiconductor structures with low resistance interfaces*. PhD thesis, Purdue University, 2000.
- [88] A. Ghosh, F. Zahid, P. Damle, and S. Datta, "Insights into molecular conduction from I-V asymmetry." Online archive - <http://xxx.lanl.gov/abs/cond-mat/0202519>, 2002.

- [89] M. Lundstrom, “Moore’s law forever?,” *Science*, vol. 299, pp. 210–211, 2003.
- [90] M. Hersam, N. Guisinger, and J. Lyding, “Silicon-based molecular nanotechnology,” *Nanotechnology*, vol. 11, pp. 70–76, 2000.

APPENDIX A: SELF-ENERGY - IMPOSING THE GOLD FCC SYMMETRY

In order to calculate the gold (111) surface Green's function, we use Gaussian '98 to simulate a 28 atom gold cluster with FCC (111) geometry and then extract the in-plane and out-of-plane overlap and Fock matrix components S_{mn} and F_{mn} (see the discussion following Eq. 5.2 in Section 5.3). The cluster, however, does not have the full group theoretical symmetry of the FCC (111) crystal due to edge-effects. This symmetry needs to be imposed on the extracted Fock matrix elements in order to be consistent with the assumptions of two-dimensional periodicity that go into our \vec{k} -space formalism. The overlap matrix elements automatically satisfy these symmetry requirements, since they involve products of two on-site localized orbitals.

A typical cluster is shown in Fig 5.3, with atoms numbered from 1 to 13. An FCC (111) surface consists of repetitions of such clusters in the sequence ABCABCABC ..., B being the plane numbered 1 to 7, and A and C being the planes with the rotated triangles above and below it. Each plane is parallel to the x-y plane, atom 1 is at the origin and atom 2 lies on the x-axis. This gives us a z -axis in the (111) direction, along which the molecule is assumed to align after contacting with the Au surface. The Fock matrix for the cluster looks as follows:

$$F = \begin{bmatrix} F_{11} & F_{12} & \cdots & F_{1,13} \\ F_{21} & F_{22} & \cdots & \\ \cdots & & & \\ \cdots & & & \\ \cdots & & & \end{bmatrix} \quad (\text{A.1})$$

The LANL2DZ basis for gold consists of three sets of s, three sets of p and 2 sets of d orbitals, giving us $3 + 3 \times 3 + 2 \times 5 = 22$ orbitals on each gold atom. Each of the above

matrix components F_{11} , F_{12} etc. is therefore 22×22 in size. Using Gaussian '98, we can extract all these matrix components separately. For a Au(111) surface, however, the elements should be interrelated by the point group symmetries of the FCC crystal. Starting with two independent matrix components, say, an on-site element F_{11} that is characteristic of the gold atom, and a nearest neighbor element F_{12} that depends on the lattice parameter, we construct the rest of the elements of F using the symmetry operations.

The FCC crystal has the following symmetry: on rotating the cluster in Fig. 5.3 by 60 degrees about the z -axis followed by a reflection in the x - y plane, we get back the same cluster. This means there is a unitary matrix M that takes F_{12} into F_{13} and so on. The matrix M can be block-decomposed into contributions from each of the s , p and d orbital subspaces. Since s orbitals are spherically symmetric, the corresponding block for the 3 s orbitals is the 3×3 identity matrix. For p orbitals, the M_p matrix consists of the following reflection (R) and rotation (U) matrices:

$$\begin{aligned}
 M_p &= R_p U_p \\
 R_p &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \\
 U_p &= \begin{pmatrix} \cos \pi/3 & -\sin \pi/3 & 0 \\ \sin \pi/3 & \cos \pi/3 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{A.2}
 \end{aligned}$$

A similar matrix can be constructed for the d -orbital subspace. At the end of the process we thus get a 22×22 M matrix describing rotation-reflection of the entire LANL2DZ basis set.

The nearest neighbor in-plane couplings of atom 1 can now be generated by repeated applications of the rotation-reflection transformation:

$$F_{17} = M^\dagger F_{12} M$$

$$\begin{aligned}
F_{16} &= M^\dagger F_{17} M \\
F_{15} &= M^\dagger F_{16} M \\
F_{14} &= M^\dagger F_{15} M \\
F_{13} &= M^\dagger F_{14} M \\
F_{12} &= M^\dagger F_{13} M
\end{aligned} \tag{A.3}$$

A simple out-of-plane rotation by 120 degrees can generate F_{18} out of F_{12} . Once again, we explicitly write down the contribution for the p orbitals. The d orbitals need to be considered analogously while the s orbitals remain unchanged.

$$\begin{aligned}
F_{18} &= (M')^\dagger F_{12} M' \\
M'_p &= \begin{pmatrix} -1/2 & -1/2\sqrt{3} & \sqrt{2/3} \\ 1/2\sqrt{3} & 5/6 & \sqrt{2/3} \\ -\sqrt{2/3} & \sqrt{2}/3 & -1/3 \end{pmatrix}
\end{aligned} \tag{A.4}$$

All the other matrix components can be similarly obtained, using translation and rotation-reflection invariance. For example:

$$\begin{aligned}
F_{1,12} &= M^\dagger F_{18} M \\
F_{1,10} &= M^\dagger F_{1,12} M \\
F_{22} &= F_{11} \\
F_{23} &= F_{14}
\end{aligned} \tag{A.5}$$

and so on.

The crystalline symmetries also constrain the structures of the two fundamental matrix components F_{11} and F_{12} out of which all other components are constructed. F_{11} is a diagonal block of F and is Hermitian. F_{12} satisfies the following condition:

$$\left(M^\dagger\right)^3 F_{12} M^3 = F_{12}^\dagger \tag{A.6}$$

Three operations of the rotation-reflection transformation (equivalently, a coordinate inversion) on F_{12} generates the matrix F_{15} , which equals F_{51}^\dagger by bond-inversion sym-

metry. Translational symmetry implies $F_{51} = F_{12}$, which leads to the above equation. The form of F_{12} consistent with the above condition is given by:

$$F_{12} = \begin{pmatrix} a & b & c \\ -b^\dagger & e & d \\ c^\dagger & -d^\dagger & f \end{pmatrix} \quad (\text{A.7})$$

where a , e and f are hermitian matrices. The F_{12} matrix extracted from Gaussian will have this structure for a large enough cluster, else we need to impose this condition as well.

With a properly symmetrized set F_{mn} and S_{mn} , the matrices $F_{a\vec{k}}$ and $S_{a\vec{k}}$ in Eq. 5.2 are Hermitian as expected, and we may proceed to calculate the gold (111) surface Green's function as outlined in Section 5.3. Ignoring the symmetries, however, can lead to non-physical results, including negative density of states or unsymmetric Green's function matrices.

APPENDIX B: DENSITY MATRIX - ANALYTICAL INTEGRATION

We have developed a fast and elegant algorithm to evaluate the density matrix given an energy independent self-energy. For bulk gold it is known that around the Fermi energy the local DOS is approximately a constant [45]. It is reasonable to expect that all the transport properties of the molecule are determined by the molecular levels near E_f and the deep levels contribute to the total number of electrons by remaining full in the bias range of interest. In view of this, We split the energy range of integration in two parts and denote the boundary by E_b . E_b is chosen such that it is a few electron-volts below the Fermi energy E_f and the molecular density of states at E_b is negligible. We replace $\Sigma(E)$ by $\Sigma(E_f)$ in the energy range between E_b to E_f .¹ For the energy range from $-\infty$ to E_b we replace $\Sigma(E)$ by $-i\eta$, a constant infinitesimal broadening. The only energy dependence in the integrand in Eq. 2.25 now appears through the Green's function G and we can perform an analytical integration in an appropriate eigenspace, as explained below.

To evaluate the density matrix, we now need to solve integrals of the type (see equation 2.25)

$$2\pi\rho = \int_{-\infty}^{\infty} dE f(E) G(E) \Gamma G^\dagger(E) \quad (\text{B.1})$$

where Γ is the anti-hermitian part (see Eq. 2.22) of the energy independent self-energy and $f(E)$ is the Fermi function with some electrochemical potential μ . In this paper we assume $T = 0$ K which modifies the integral in the above equation as

$$2\pi\rho = \int_{E_{min}}^{\mu} dE G(E) \Gamma G^\dagger(E) \quad (\text{B.2})$$

¹If the DOS is finite at E_b , there will be a dependence of our results on E_b , leading to a corresponding ambiguity regarding the precise location of E_f at equilibrium. This is an issue that deserves more attention, since the precise location of E_f is altered significantly even if the total number of electrons, integrated over the entire energy range, is miscalculated by a small fraction. As such, the constant self-energy approximation used here could also influence E_f in a similar way.

where it is understood that E_{min} tends to $-\infty$. Using Eq. 2.40 for $G(E)$ we may rewrite the above equation as

$$2\pi\rho = S^{-\frac{1}{2}} \left[\int_{E_{min}}^{\mu} dE \overline{G}(E) \overline{\Gamma} \overline{G}^{\dagger}(E) \right] S^{-\frac{1}{2}} \quad (\text{B.3})$$

where S is the overlap matrix $\overline{\Gamma}$ and $\overline{G}(E)$ are defined as

$$\overline{\Gamma} = S^{-\frac{1}{2}} \Gamma S^{-\frac{1}{2}}$$

and

$$\overline{G}(E) = (EI - \overline{F})^{-1}$$

with I being the identity matrix and

$$\overline{F} = S^{-\frac{1}{2}} (F + \Sigma_1 + \Sigma_2) S^{-\frac{1}{2}}$$

In the above equation, $\Sigma_{1,2}$ are the energy independent self-energy matrices as discussed at the beginning of this appendix and F is the molecular Fock matrix.

To solve the integral in Eq. B.3 analytically, we work in the eigenspace of the energy independent non-Hermitian matrix \overline{F} given by the above equation. Since \overline{F} is non-Hermitian, we need its eigenkets $\{|m\rangle\}$ and their dual kets $\{|\tilde{m}\rangle\}$ in order to form a complete basis set. The corresponding eigenvalues are then $\{\epsilon_m\}$ and $\{\epsilon_m^*\}$. Expanding in this mixed basis set, we get

$$\begin{aligned} 2\pi\rho &= S^{-\frac{1}{2}} \left[\int_{E_{min}}^{\mu} dE \overline{G}(E) \left(\sum_p |p\rangle \langle \tilde{p}| \right) \overline{\Gamma} \overline{G}^{\dagger}(E) \left(\sum_q |\tilde{q}\rangle \langle q| \right) \right] S^{-\frac{1}{2}} \\ &= S^{-\frac{1}{2}} \left[\sum_{pq} \overline{\Gamma}_{\tilde{p}\tilde{q}} |p\rangle \langle q| \int_{E_{min}}^{\mu} \frac{dE}{(E - \epsilon_p)(E - \epsilon_q^*)} \right] S^{-\frac{1}{2}} \end{aligned} \quad (\text{B.4})$$

where we used $\overline{G}(E)|m\rangle = |m\rangle/(E - \epsilon_m)$ and $\overline{G}^{\dagger}(E)|\tilde{m}\rangle = |\tilde{m}\rangle/(E - \epsilon_m^*)$.

It is now straightforward to do the above integral analytically and obtain the density matrix. The current I given by Eq. 2.27 may also be evaluated analytically using a similar procedure. This analytical procedure allows us to do the integration extremely fast and avoids errors arising out of a numerical integration over a grid.

APPENDIX C: GOLD SURFACE DOS WITH MORE NEAR NEIGHBORS

The k-space calculation of the gold surface Green's function involves a summation of nearest neighbor couplings (see Equation 5.2 and the related discussion). Usually it is enough to include only one or two nearest neighbors in such calculations because of the localized nature of the basis functions. The real space discrete lattice used in Chapter 3, the semi-empirical p_z orbital basis used in Chapter 4 and the extended Hückel basis functions that are extensively used to model molecules are examples of such localized basis functions. However, the LANL2DZ basis functions used to describe gold contact atoms in Chapters 5 and 6 are relatively delocalized, so much so that an estimated five to six near neighbors should be accounted for in order to accurately obtain the surface Green's function ¹. Due to computational constraints, the surface Green's function used in Chapters 5 and 6 was calculated by considering two nearest neighbor couplings. In this appendix, we test this approximation by comparing the surface density of states (DOS) obtained from a calculation with two near neighbors to that obtained from a calculation with more near neighbors (five in plane and three out of plane).

Conceptually the simplest way to account for more near neighbors is to increase the size of the unit cell. A 258 atom gold cluster used to obtain the couplings between three-atom unit cells is shown in Figure C.1. The central part of this big cluster is best visualized by looking at the one-atom unit cell gold cluster shown in Figure 5.3, but replacing each atom by a three-atom unit similar to the one shown in Figure C.1.

¹This estimate was obtained by observing that the fifth near neighbor overlap is very small as compared to the nearest neighbor overlap. We can also support this estimate by calculating the eigenvalues of the overlap matrix of a large cluster of gold atoms. If the overlap between less than the fifth near neighbors is set to zero, some eigenvalues (previously small non-zero) become negative, which is contrary to the fact that the eigenvalues of an overlap matrix are always positive [7].

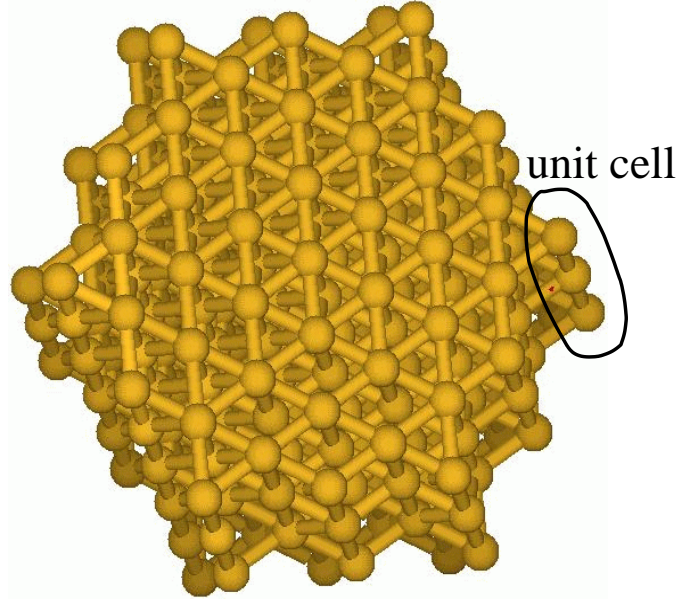


Fig. C.1. A 258 gold atom cluster used to obtain the couplings between nearest neighbor three-atom unit cells. Using this cluster ensures that we account for third nearest neighbor atom (out of plane) and fifth nearest neighbor atom (in plane) coupling. Ideally, we need about five to six near neighbors for an accurate calculation of the surface Green's function with the LANL2DZ basis set.

The 258 atom cluster is still too big for us to be able to obtain the Fock matrix couplings self-consistently using Gaussian '98. However, it is relatively easy to just calculate the overlap matrix S of this big cluster and then use it to obtain the Fock matrix F from the well-known Hückel principle:

$$F_{ij} = 1.75 S_{ij} \frac{(F_{ii} + F_{jj})}{2}$$

where the diagonal entries F_{ii} of the Fock matrix (which correspond to the on-site energies of every orbital in the basis) are known from our previous self-consistent calculation with two near neighbors.

Using these Fock matrix and overlap matrix couplings we calculated the surface Green's function using the procedure described in Section 5.3. The results are shown in Figure C.2. We observe that the two near neighbor calculation (dashed line) is qualitatively similar to the bigger calculation (solid line). The small quantitative difference is expected because of the delocalized nature of the LANL2DZ basis functions.

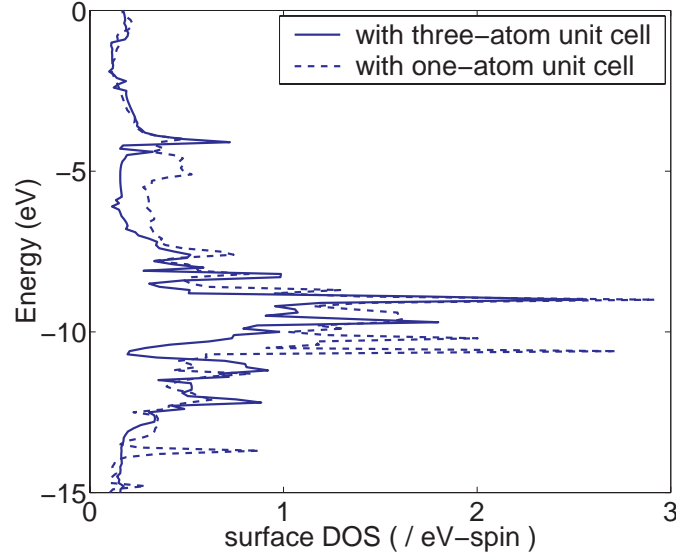


Fig. C.2. Gold surface density of states (LANL2DZ basis). The three-atom unit cell calculation accounts for many more near neighbor atoms as compared to the one-atom unit cell calculation, and the results for the two cases are qualitatively similar, but differ quantitatively. This quantitative difference is expected in light of the fact that the LANL2DZ basis functions are relatively delocalized.

To calculate the surface Green's function (and thereby the contact self-energy) more accurately (as compared to the two near neighbor calculation used in Chapters 5 and 6) we either need to do a self-consistent calculation with the 258 atom cluster shown above to obtain the Fock matrix couplings, or we need to use another relatively localized set of basis functions to describe the gold contacts.

APPENDIX D: MODIFICATIONS TO GAUSSIAN '98

In this appendix we list the modifications we made to the standard quantum chemistry software Gaussian '98 (revision A.11.3) [16] in order to obtain the results shown in Chapters 5 and 6. Gaussian '98 is written in FORTRAN. We have modified only three links (or subroutines) in Gaussian '98, namely, link 302 (Section D.1), link 401 (Section D.2) and link 502 (Section D.3)¹. We invoke a MATLAB compiled binary code from inside link 502 for the calculation of the density matrix, once Gaussian '98 is done calculating the Fock matrix (see Figure 5.1 and the discussion in Section 5.2 for details of the calculation procedure). Section D.4 contains the MATLAB script.

D.1 Link 302

Link 302 (FORTRAN subroutine 1302.F) is used to calculate the overlap matrix and the core Hamiltonian. For gold atoms, the calculation of the core Hamiltonian takes a long time because the effective core potential (ECP) integrals are computationally intensive. We have modified 1302.F so that these ECP integrals are calculated only once, and then read in for all subsequent runs with the same molecule. A patch of code in the unmodified version is as follows:

```
If(IDoECP.ne.1.and.ITqry(512).gt.0) then
    If(DoSpar) Call GauErr('No sparse with ECPs yet.')
    Call PcckI(0,Junk,IPrAll,MaxAtm,NOpAll,NAtoms,NOpAll)
    Call DecECP(100,ICntrl,AccECP)
    Call ECPInt(IOut,IPrint,.True.,.True.,ICntrl,NAtoms,NBasis,1,
```

¹In addition to these links, we have also commented out line number 15973 in the utility file `utilam.F`. This line, if left uncommented, returns an error if a non-integer number of electrons in the molecule is detected. For the open systems that we work on, however, fractional number of electrons can be present in the molecule.

```
$    NAtoms,1,NOp,AccECP,V(jNEqSh),SymAll,V(jNEqB),IPrAll,C,XX,
$    V(jS),XX,XX,XX,V(jStart),V(jStart),LenV)
    Call UnPckI(0,Junk,IPrAll,MaxAtm,NOpAll,NAtoms,NOpAll)
```

The modified version is:

```
    If(IDoECP.ne.1.and.ITqry(512).gt.0) then
        If(DoSpar) Call GauErr('No sparse with ECPs yet.')
        Call PcckI(0,Junk,IPrAll,MaxAtm,NOpAll,NAtoms,NOpAll)
        Call DecECP(100,ICntrl,AccECP)

    open(unit=10,file='ECP_done',status='unknown')
    read(10,*)ecppd
    close(10)
    If(ecppd.eq.0) then
        Call ECPInt(IOut,IPrint,.True.,.True.,ICntrl,NAtoms,NBasis,1,
$    NAtoms,1,NOp,AccECP,V(jNEqSh),SymAll,V(jNEqB),IPrAll,C,XX,
$    V(jS),XX,XX,XX,V(jStart),V(jStart),LenV)
        open(unit=10,file='ECPInt',status='unknown')
        do jpd=1,NTT
            write(10,'(E28.15e3)')V(jS+jpd-1)
        enddo
        close(10)
        write(6,*)'Written ECP integral'
    else
        open(unit=10,file='ECPInt',status='unknown')
        read(10,'(E28.15e3)')(V(jS+jpd-1),jpd=1,NTT)
        close(10)
        write(6,*)'Read ECP integral'
    endif
```

```
Call UnPckI(0,Junk,IPrAll,MaxAtm,NOpAll,NAtoms,NOpAll)
```

D.2 Link 401

Link 401 (FORTRAN subroutine 1401.F) is used to generate or read in an initial guess density matrix. A patch of code in the unmodified version is as follows:

```

      If(ICmp.eq.0) then
        Call FormP2(.True.,NBasis,NEX,V(IA),V(IB))
        If(IUHF.eq.0.and.IROHF.eq.0)
$          Call AScale(NTT,Two,V(IB),V(IB))
        Call ConDDF(NumD,NTT)
        write(6,*)' NumD is : ',NumD
        Call FileIO(1,-NumD,NTT,V(IB),0)
      else
        PFact = GFloat(2-IUHF)
        Call CmpDen(In,IOut,IPrint,ICmp,IMix,IOCMat,IODMat,V(IA),
$          V(IB),V(IAA),NBasis,NBasis,PFact,2*I,NEX)
        endIf
        Call TStamp(1,'Form P')
      endIf

```

The modified version is:

```

      If(ICmp.eq.0) then
        Call FormP2(.True.,NBasis,NEX,V(IA),V(IB))
        If(IUHF.eq.0.and.IROHF.eq.0)
$          Call AScale(NTT,Two,V(IB),V(IB))
        Call ConDDF(NumD,NTT)
        write(6,*)' NumD is : ',NumD
        Call FileIO(1,-NumD,NTT,V(IB),0)

```

```

      If(Iflagpd.eq.1) then
        open(unit=10,file='in',status='unknown')
        read(10,'(E28.15e3)') (V(IB+jpd),jpd=0,NTT-1)
        close(10)
        Call AScale(NTT,Two,V(IB),V(IB))
        Call FileIO(1,-NumD,NTT,V(IB),0)
        write(6,*)'NEGF Density guess received'
      endIf

    else
      PFact = GFloat(2-IUHF)
      Call CmpDen(In,IOut,IPrint,ICmp,IMix,IOCMat,IODMat,V(IA),
$      V(IB),V(IAA),NBasis,NBasis,PFact,2*I,NEX)
      endIf
      Call TStamp(1,'Form P')
    endIf

```

D.3 Link 502

Link 502 (FORTRAN subroutine 1502.F) is used to calculate the self-consistent potential ². A patch of code in the unmodified version is as follows:

```

      else if(ROHF) then
        Call DgFock(IOut,IPrint,ROHF,ZD0,RComp,IV1,IDiagX,ISyOcc,
$      IWfnSy,DoDiag,NBlks,NBasis,NOrbs,NAE,IRwTM,IRwF1,IRwCA,IOCAS,
$      0,IRwEig,0,IRwPA,AccDag,NBFU,NOccSy,.False.,JJA,CutOff,

```

²Here we only list the changes made to that part of the unmodified code which does a restricted (spin degenerate) calculation. Virtually identical changes were also made to the part that does an unrestricted (spin non-degenerate) calculation, but are not reproduced here. In this dissertation we have only shown results from restricted calculations.

```

$   'Alpha',JCycle,IPFON,.True.,TempSt,ErrMax,Temp,Beta,EF,F,D,V,
$   Scr2,LScr2-1,Scr2(1,2))
    If(.not.RHF) then
        Call FormPSA(IOut,.True.,IPFON,NBasis,NOrbs,NBE,Temp,Beta,
$       ErrMax,EF,Scr2,V,D,Scr2(1,2))
        If(IPrint.ge.3) Call LTOutS(IOut,'Beta density at cycle',
$       JCycle,NBasis,D,0)
    endIf
    Call FileIO(1,-IRwEig,NOrbs,Scr2,NOrbs)
    Call FileIO(1,-IRwCB,NBNO,V,0)
    Call FileIO(1,-IRwPB,NTT,D,0)

```

The modified version is:

```

    else if(ROHF) then
        Call DgFock(IOut,IPrint,ROHF,ZDO,RComp,IV1,IDiagX,ISyOcc,
$       IWfnSy,DoDiag,NBlks,NBasis,NOrbs,NAE,IRwTM,IRwF1,IRwCA,IOCAS,
$       0,IRwEig,0,IRwPA,AccDag,NBFU,NOccSy,.False.,JJA,CutOff,
$       'Alpha',JCycle,IPFON,.True.,TempSt,ErrMax,Temp,Beta,EF,F,D,V,
$       Scr2,LScr2-1,Scr2(1,2))
        Call FileIO(2,-IRwF1,NTT,FApd,0)
        open(unit=10,file='out',status='unknown')
        do jpd=1,NTT
            write(10,'(E28.15e3)'),FApd(jpd)
        enddo
        close(10)
C       Having written out the Fock matrix,
C       go to MATLAB to get the density matrix
        MLABpd=system("./sc")
        open(unit=10,file='in',status='unknown')
        read(10,'(E28.15e3)') (DApd(jpd),jpd=1,NTT)

```

```

close(10)

Call FileIO(1,-IRwPA,NTT,DAPd,0)

If(.not.RHF) then
    Call FormPSA(IOut,.True.,IPFON,NBasis,NOrbs,NBE,Temp,Beta,
$      ErrMax,EF,Scr2,V,D,Scr2(1,2))
    If(IPrint.ge.3) Call LTOutS(IOut,'Beta density at cycle',
$      JCycle,NBasis,D,0)
    endIf

    Call FileIO(1,-IRwEig,NOrbs,Scr2,NOrbs)
    Call FileIO(1,-IRwCB,NBNO,V,0)
    Call FileIO(1,-IRwPB,NTT,DAPd,0)

```

D.4 MATLAB Script

The core part of our MATLAB script is reproduced here. The full script contains some additional helper functions, which are mainly used to ensure smooth operation of the overall calculation. Interested readers can find all the MATLAB functions (including those given below) at <http://shay.ecn.purdue.edu/~damle>. The function `sc.m` reproduced below computes the density matrix using the NEGF equations (see Chapter 2 and Appendix B). Note that this function is converted to a binary executable named `sc` using the MATLAB compiler before being called through link 502 in Gaussian '98.

```

function []=sc()
disp('entering link matlab');
% load data file containing the variables
% S, X (= S^-0.5), nsto, hartree_unit (= 27.211 eV),
% sig12, gam1, gam2, sig12w, gam1w, gam2w, mu1, mu2,
% Ewin, damping, Niter, Fbak, P.
load sc_data

```



```
%load out, the file containing the Fock matrix
iout=fopen('out','r');
out=fscanf(iout,'%E');
fclose(iout);
% construct the full Fock matrix
F=zeros(nsto);
for ii=1:nsto
    for jj=1:ii
        k=jj + ii*(ii-1)/2;
        F(ii,jj)=out(k);
    end
end;
for ii=1:nsto-1
    for jj=ii+1:nsto
        F(ii,jj)=F(jj,ii);
    end
end
% Mixing for iterative convergence
if Niter==1
    Fbak=F;
else
    F = Fbak + damping * (F - Fbak);
    Fbak=F;
end
% Integrate from Ewin to mu1 / mu2
Heff = F*hartree_unit + sig12;
Heff_o=X'*Heff*X;
gam1_o=X'*gam1*X;
gam2_o=X'*gam2*X;
```

```
[V,D]=eig(Heff_o); D=diag(D);
% error check
if sum(imag(D)>0)
    fprintf(1,'%d imag elements on diagonal of D
            are positive',sum(imag(D)>0));
end
Vt = V';
Vc = eye(nsto)/Vt;
P1 = density(V,D,Vt,Vc,gam1_o,Ewin,mu1);
P2 = density(V,D,Vt,Vc,gam2_o,Ewin,mu2);
Eminf=-1e100; % minus infinity
% Integrate from minus infinity to Ewin
Heffw=F*hartree_unit+sig12w;
Heffw_o=X'*Heffw*X;
gam1w_o=X'*gam1w*X;
gam2w_o=X'*gam2w*X;
[Vw,Dw]=eig(Heffw_o); Dw=diag(Dw);
% error check
if sum(imag(Dw)>0)
    fprintf(1,'%d imag elements on diagonal of Dw
            are positive',sum(imag(Dw)>0));
end
Vtw = Vw';
Vcw = eye(nsto)/Vtw;
Pw = density(Vw,Dw,Vtw,Vcw,gam1w_o+gam2w_o,Eminf,Ewin);
fprintf(1,'trace of P1, P2, Pw: %6.2f      %6.2f
          %6.2f\n',trace(P1),trace(P2),trace(Pw));
P = P1 + P2 + Pw;
Ne=2*trace(P);
```

```
P=X*P*X';
% Use only the real part of P for self-consistency
P=real(P);
% error check
if sum(diag(P)<0)
    fprintf(1,'%d elements on diagonal of P
              are negative',sum(diag(P)<0));
end
% Construct a column vector from lower triangular part of
% P in order for Gaussian '98 to read it
Pout=zeros(nsto*(nsto+1)/2,1);
for ii=1:nsto
    for jj=1:ii
        k=jj+ii*(ii-1)/2;
        Pout(k)=P(ii,jj);
    end
end
iin=fopen('in','w');
count=fprintf(iin,'%1.15E\n',Pout);
fclose(iin);
Niter=Niter+1;
fprintf(1,'Ne=%d \n',Ne);
save sc_data Fbak P Niter -append
disp('leaving link matlab');
% =====
% density matrix: analytical integration (see Appendix B)
function [den]=density(V,D,Vt,Vc,gam,Emin,mus)
Nd=size(gam,1);
repdum=ones(1,Nd);
```

```
DD = D*repdum;
logmat = (log(1-mus./D))*repdum;
logmat2 = (log(1-Emin./D))*repdum;
invmat = 1./(2*pi*(DD-DD'));
pref2 = logmat-logmat';
pref3 = logmat2-logmat2';
prefactor=invmat.*(pref2 - pref3);
Gammam = Vc'*gam*Vc;
prefactor=prefactor.*Gammam;
den = V*prefactor*Vt;
```

VITA

Prashant Subhash Damle was born on May 25th, 1976 in Pune, India. He received the B.Tech. degree in Electrical Engineering from the Indian Institute of Technology (IIT), Bombay, India, in 1997, and the M.S. and Ph.D. degrees in Electrical Engineering from Purdue University, West Lafayette, Indiana, in 1999 and 2003 respectively. Currently Prashant is employed by Intel Corporation, Santa Clara, California as a device engineer.