11

Modeling Electronics at the Nanoscale

Narayan R. Aluru
University of Illinois

Jean-Pierre Leburton
University of Illinois

William McMahon,
University of Illinois

Umberto Ravaioli
University of Illinois

Slava V. Rotkin
University of Illinois

Martin Staedele
Infineon Technologies

Trudy van der Straaten
University of Illinois

Blair R. Tuttle
Penn State University

Karl Hess
University of Illinois

CONTENTS

11.1 Introduction
11.2 Nanostructure Studies of the Si-SiO_2 Interface
Si-H Bonds at the Si-SiO_2 Interface • Reliability Considerations at the Nanoscale • Tunneling in Ultra Thin Oxides
11.3 Modeling of Quantum Dots and Artificial Atoms
The Many-Body Hamiltonian of Artificial Atoms • Quantum Modeling of Artificial Molecules and Exchange Engineering
11.4 Carbon Nanotubes and Nanotechnology
Operation of Nanoelectromechanical Switches • Nanotube Mechanics • Electrostatics • Analytical Consideration for the Pull-In • Outlook
11.5 Simulation of Ionic Channels
Hierarchical Approach to Modeling Ion Channels • Drift-Diffusion Models • Monte Carlo Simulations
11.6 Conclusions
Acknowledgments
References

11.1 Introduction

Nanostructure research is defined by a scale — the nanometer length scale. Simulation of nanostructures, however, must be multiscale in its very nature. It is not the nanostructures themselves that open the horizon to new opportunities and applications in all walks of life; it is the integration of nanostructures into large systems that offers the possibility to perform complex electrical, mechanical, optical, and chemical tasks.

Conventional electronics approaches to nanometer dimensions and simulation techniques must increasingly use atomistic methods to compute, for example, tunneling and size quantization effects as well as the features of the electronic structure of the solids that define the nanometer-sized device. The atomistic properties need then to be linked to macroscopic electromagnetic fields and to the equations of Maxwell and, ultimately, to systems performance and reliability. The transition from the quantum and atomistic scale to the classical macroscopic scale is of great importance for the accuracy of the simulation. It can be described by the Landauer–Buettiker formalism, by Bardeen's transfer Hamiltonian method, or by more demanding methods such as the Schroedinger Equation Monte Carlo approach.[1] To encompass all of these scales and transitions, a hierarchy of methods (sets of equations) that supply each other with parameters is needed even for conventional silicon technology. Similar hierarchical

© 2003 by CRC Press LLC

approaches will be needed for future devices and their integration in electronics as well as electromechanics. One can already anticipate the demand for simulation methods that merge electronics, mechanics, and optics as well as the highly developed methods of chemistry.

A theoretical tool of ever-increasing use and usefulness is density functional theory (DFT). DFT describes, for example, the electrical and optical properties of a quantum dot (the prototype for future electron devices) and is also widely used in chemistry. The simulation methods become altogether more fundamental and powerful as simulation of nanostructure technology, both present and more futuristic, progresses. For example, the same simulation methods that have been developed in the last decade for silicon technology can also be applied to some biological systems, the carbon-based devices of nature, and the newly emerging field of carbon nanotubes. In turn, the methods developed in biochemistry become increasingly useful to answer questions in electronics and electromechanics at the nanoscale.

It is currently not possible to give an overview of all these opportunities in the limited space of this chapter. We present therefore only four vignettes that demonstrate the wide range of knowledge that is needed in nanostructure simulations and what can be anticipated in the future for simulations ranging from silicon-based electronics and nanoelectromechanics to biological systems such as protein-based ion channels.

11.2 Nanostructure Studies of the Si-SiO_2 Interface

In this section we discuss modeling of the Si-SiO_2 interface, mostly in the context of Metal-Oxide-Semiconductor Field-Effect Transistors (MOSFETs). New insights are gained by explicitly calculating material properties using nanostructure and atomic-level techniques. This section thus offers an example of how nanostructure simulation is already necessary for conventional silicon technology as encountered in the highly integrated chips of today.

11.2.1 Si-H Bonds at the Si-SiO_2 Interface

Hydrogen has long been used in the processing of MOSFETs in order to passivate electrically active defects that occur, for instance, at the Si-SiO_2 interface. The Si-H binding energy was commonly assumed to be the threshold energy for H-related degradation in MOSFETs.[2] We have used density functional calculations to investigate the energetics of the hydrogen dissociation process itself.[3–8] These calculations show that there are several mechanisms by which hydrogen can desorb through processes that involve much lower energies than the Si-H binding energy of ~3.6 eV. These results explain continued hot-electron degradation in MOSFETs even as operating voltages have been scaled to below 3.6 eV.[5] Moreover, a distribution of dissociation energies due to disorder at the interface is expected. Such a distribution indicates that the probability of degradation will increase dramatically as MOSFETs are scaled to sub-100 nanometer channel lengths.[9,10]

11.2.1.1 Density Functional Calculations

Density functional theory (DFT) has become the leading theoretical tool for understanding nanoscale phenomena in physics and chemistry. This is because DFT allows an accurate determination of electronic structure and also efficiently scales with the number of atoms in a calculation. We have performed a comprehensive DFT study of the mechanisms of Si-H bond breaking at the Si-SiO_2 interface. We have used several atomic models of the interface including the cluster model shown in Figure 11.1. These studies demonstrate how DFT can be used to model electronics on the nanoscale.

Our main results for Si-H at the Si-SiO_2 interface are as follows. The energy needed to dissociate an isolated silicon–hydrogen bond (placing the hydrogen in a vacuum state at infinity) is found to be ~3.6 eV. For an Si-H bond at the Si-SiO_2 interface, if the dissociated hydrogen atom enters bulk SiO_2, then the dissociation or dissociation energy is also 3.6 eV because atomic hydrogen interacts only weakly with the rather open, insulating oxide. However, the Si-H dissociation energy can be significantly reduced for Si-H bonds at the Si-SiO_2 interface because hydrogen can desorb by first entering bulk silicon. The energy

© 2003 by CRC Press LLC

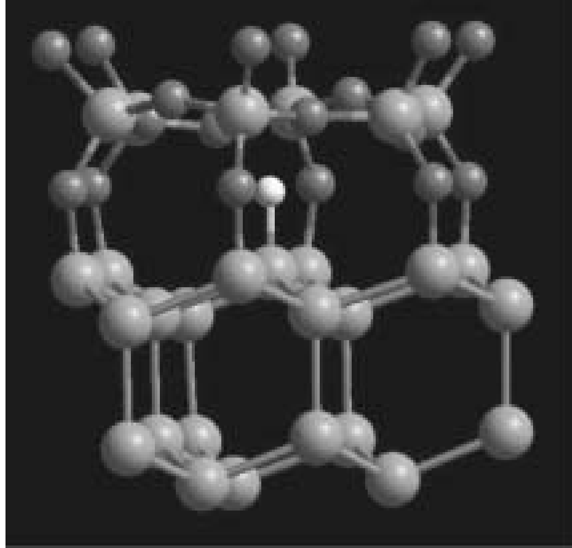


FIGURE 11.1 Atomic ball-and-stick model of an isolated Si-H bond at the Si-SiO_2 interface. Smaller balls represent oxygen atoms, and larger balls represent silicon.

needed to place a neutral hydrogen atom, arising from the silicon dangling bond site, into bulk silicon far from any defects is ~2.5 eV. As hydrogen diffuses to a surface or interface, it can passivate other defects or combine with another hydrogen atom to form H_2. At a surface or an open interface such as the Si-SiO_2 interface, H_2 molecules can easily diffuse away, leaving behind the silicon dangling bonds. Experimentally, the thermally activated dissociation of hydrogen from the (111)Si-SiO_2 interface is measured at 2.56 eV. This is consistent with our calculated mechanism with H entering bulk silicon before leaving the system as H_2.

In addition to the above considerations, the threshold energy for hot-electron degradation can be greatly reduced if dissociation occurs by multiple vibrational excitations. For low voltages, Si-H dissociation involving multiple vibrational excitations by the transport electrons becomes relatively more likely. Because hydrogen is very light, the hydrogen in an Si-H bond is a quantum oscillator. Hot-electrons can excite the hydrogen quantum oscillator from the ground state into an excited state. Because the Si-H

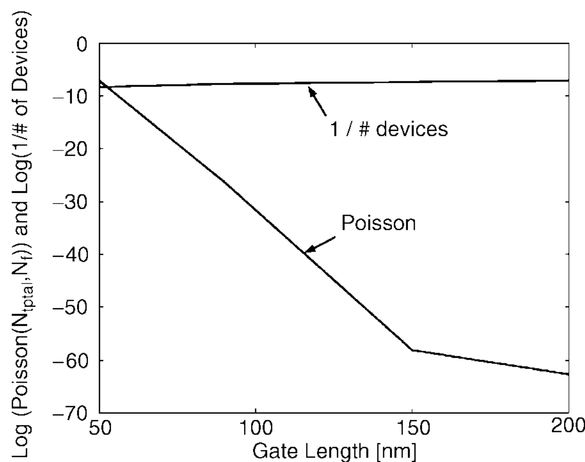


FIGURE 11.2 Example of failure function for interface trap generation.

© 2003 by CRC Press LLC

vibrational modes are well above the silicon phonon modes, the excited state will be long-lived, allowing for multiple vibrational excitation. In this case, the Si-H dissociation can take place at channel electron energies lower than 2.5 eV and perhaps as low as 0.1 eV, the vibrational energy of the Si-H bending mode.[11]

11.2.2 Reliability Considerations at the Nanoscale

11.2.2.1 Increasing Effect of Defect Precursor Distribution at the Nanoscale

For micron-sized devices, many Si-H bonds at the Si-SiO_2 interface must be broken before the device has significantly degraded. For nanoscale devices, a much smaller number of defects (possibly on the order of 10s or lower) could cause a device to fail. Because of the smaller number of defects required, there is an increasing probability with decreasing device size of having a significant percentage of defect precursors with lifetimes in the short-lifetime tail of the Si-H dissociation energy distribution mentioned above. This results in an increasing number of short-time failures for smaller devices. In order to quantify this result, the shape of the distribution of dissociation energies must be known. Fortunately, this shape can be determined from the time dependence of trap generation under hot-electron stress, and from this the effect of deviations from this distribution on the reliability can be calculated. To understand how the reliability can be understood, we specifically look at the example of interface trap generation in nMOSFETs.

11.2.2.2 Hot-Electron Interface Trap Generation for Submicron nMOSFETs

A sublinear power law of defect generation with time is observed for the generation of interface traps at the Si-SiO_2 interface. Because the hydrogen is relatively diffusely spread throughout the interface (with only around one silicon–hydrogen bond for every hundred lattice spacings), it is clear that any process which breaks these bonds will be first order in the number of Si-H bonds. That is, the rate equation for this process can be written

$$\frac{dN(E_b)}{dt} = -\frac{N(E_b)}{\tau(E_b)}$$

where $N(E_b)$ is the number of silicon–hydrogen bonds (which must be a function of E_b, the bond energy), and τ is some lifetime that, for a hot-electron-driven process, would involve an integration of the electron distribution with the cross-section for defect creation, also a function of the bond energy. In order to get the true number of defects as a function of time for an average device, the solution to this rate equation must be integrated over the distribution of bond energies. This gives

$$N_{tot}(t) = \int_0^\infty f(E_b) N_0 \exp\left(-\frac{t}{\tau(E_b)}\right) dE_b$$

This integral is what produces the sublinearity of the time dependence of the generation of interface traps. The importance of this integral lies in the fact that it relates the sublinearity of the time dependence of the generation of interface traps with the average distribution of defect energies, which can be related to the distribution of defect generation lifetimes. This distribution can be used to determine the failure function for the failure mode involving this type of defect. One can extract the defect activation energy distribution from this integral once one knows the sublinear power law for defect generation with time.[12]

11.2.2.3 Reliability from Defect Precursor Distribution

Again utilizing the assumption of independent defects, the failure function of a device (defined as the probability of having a sufficient number of defects that will fail before some time *t*) will be a binomial or, approximately, a Poisson distribution. One of the characteristics of this failure function is an exponential increase in the probability of failure as the number of defects required to cause failure gets small. This is demonstrated in Figure 11.2, where we compare the reciprocal of the number of devices on a chip (which gives an idea of how much the reliability of a single device on that chip must increase) with the

© 2003 by CRC Press LLC

variation in the Poisson distribution with the number of interface traps required for the failure of a device. This is done for four gate lengths, with the gate lengths and number of devices at a given gate length taken from the semiconductor roadmap. The number of defects required for device failure comes from ISE-TCAD simulations. Notice the number of failures increases exponentially as the gate length is reduced below 100 nanometers.

Using knowledge of the type of defect involved in a degradation process, one can analytically derive the expected failure function for that type of degradation. This is not restricted to interface trap generation by Si-H dissociation, as the assumptions that went into the model are very few: first-order kinetics and a distribution of dissociation energies.

11.2.3 Tunneling in Ultra Thin Oxides

The thickness of gate oxides in MOSFETs is approaching 1–2 nm, i.e., only a few Si-O bond lengths. Consequently, gate leakage currents have become a major design consideration. For such ultra thin oxides, it is increasingly important to understand the *influence* of microscopic structure and composition of the oxide and its interface with silicon *on* the magnitude of oxide transmission probabilities and tunneling currents.

To fully explore the microscopic nature of gate leakage currents, an atomic-orbital formalism for calculating the transmission probabilities for electrons incident on microscopic models of Si-SiO_2-Si heterojunction barriers was implemented.[13–15] Subsequently the magnitude of leakage currents in *real* MOSFETs was calculated by incorporating the incident electron density from device simulations.[16,17] Such an approach allows one to examine the influence of atomic structure on tunneling. Significant results include assessing the validity of the bulk band structure picture of tunneling, determining the energy dependence of tunneling effective mass, and quantifying the nature of resonant tunneling through defects. Below, we will briefly discuss the most important details and results.

The microscopic supercell models of Si[100]-SiO_2-Si[100] heterojunctions that have been used were constructed by sandwiching unit cells of (initially) tridymite or *beta*-quartz polytype of SiO_2 between two Si[100] surfaces. The models are periodic in the plane perpendicular to the interface with periodic lengths of 0.5–1.5 nm. As more detail is desired, e.g., to examine the effects of interfacial morphology, the lateral periodic length scale can be increased with added computational costs. As an example, Figure 11.3 shows a ball-and-stick skeleton of a tridymite-based cell.

Reflection and transmission coefficients of the supercells described above were calculated using a transfer-matrix-type scheme embedded in a tight-binding framework. We solve the Schrodinger equation with open boundary conditions for the whole junction at a fixed energy E (measured relative to the silicon conduction band minimum on the channel side of the oxide) and in-plane momentum $k_{||}$ (that is a good quantum number due to the lateral periodicity) in a *layer-orbital basis*. An empirical sp^3 tight-binding basis with second-nearest neighbor interactions for both silicon and the oxide were used. The tight-binding parameters were chosen to yield experimental bulk band gaps and to reproduce density functional calculations of the effective masses of the lowest conduction bands. An electron state propagating toward the oxide from the channel side of the junction, characterized by E, $k_{||}$ and its wavevector component normal to the interface ($k_{perp,in}$), is scattered into sets of reflected and transmitted states (characterized by wavevector components $k_{perp,out}$). From the scattering wavefunctions, transmission amplitudes and dimensionless transmission coefficients are obtained.

The present microscopic models allow one to predict the intrinsic decay properties of the wavefunctions into the gate oxide. Because of the local nature of bonding in the oxide, a bulk picture of tunneling persists qualitatively even for the thinnest oxide barriers. We have analyzed the complex bands of the present bulk oxide models and find that (1) only one single complex band is relevant for electron tunneling; (2) several different bands are involved in hole tunneling; and (3) all complex oxide bands are highly nonparabolic. Because of the mismatch in the Brillouin zones for the oxide on top of the silicon, the bulk silicon $k_{||}$ is not conserved and different states have differing decay constants. The energy dependence of the integrated transmission is shown for oxide thicknesses between 0.7 and 4.6 nm in Figure 11.4, which also includes effective–mass-based results with a constant (EM) and the

© 2003 by CRC Press LLC

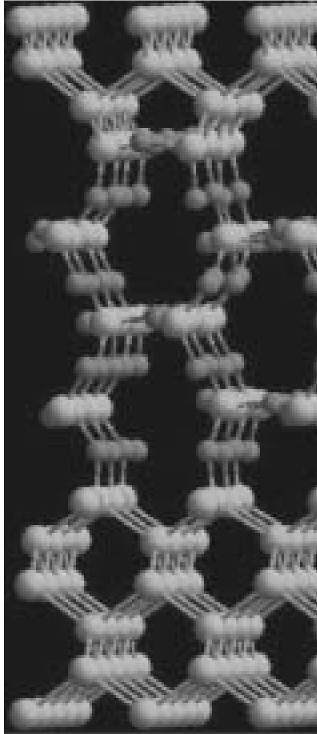


FIGURE 11.3 Ball-and-stick model of an Si[100]-SiO_2-Si[100] model heterojunction based on the 1.3 nm thin gate oxide based on the tridymite polytype of SiO_2. (Dark = oxygen and light = silicon.)

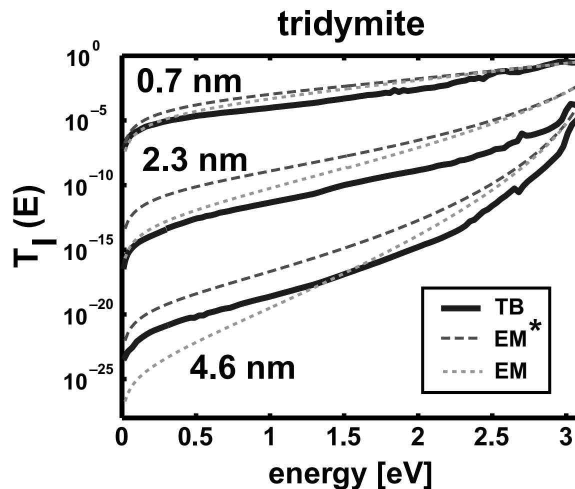


FIGURE 11.4 Integrated transmission (T_I) vs. the energy of the incident electron for tridymite-based oxides with thicknesses of 0.7, 2.3, and 4.6 nm. Results are for calculations with our atomic-level tight-binding method (TB, solid line), effective mass approaches with constant (EM, dotted line), and energy dependent (EM*, dashed line) effective masses.

energy-dependent (EM*) electron mass, which was fitted to our tight-binding complex band structures. The parabolic effective mass approximation overestimates the transmission for oxides thinner than ~1 nm. As oxide thicknesses increase, the tight-binding transmission is underestimated at low energies

© 2003 by CRC Press LLC

and overestimated at higher energies. The higher slope of the transmission obtained in the parabolic effective mass approximation is consistent with the findings for the tunneling masses and explains previous errors in oxide thicknesses derived from tunneling experiments and a constant parabolic effective mass model.[13] Using the correct tight-binding dispersion of the imaginary bands in an effective mass calculation (i.e., the EM* results in Figure 11.4) leads to qualitatively correct slopes for transmission; however, the absolute values are typically overestimated by one to two orders of magnitude. A possible reason for much of this discrepancy may be that the effective-mass-based transmission calculation underestimates the full band structure mismatch of silicon and its oxide.

The transmission coefficients were combined with electron densities and the corresponding distribution functions at the Si-SiO_2 interface of prototypical MOSFETs with channels of 50 nm and 90 nm. These quantities were obtained from full-band Monte Carlo simulations and were used to calculate the absolute magnitudes for gate leakage currents which, for oxide thicknesses smaller than ~4 nm, are dominated by tunneling of cold electrons in the source and drain contacts for defect-free oxides. As a consequence, the tunneling current densities (integrated over the entire gate length) decrease upon applying a drain-source voltage. The elastic gate leakage currents were recalculated including oxygen vacancies for a given energy E_{vac} in the oxide band gap from 0 to 3 eV above the silicon conduction band edge. The leakage currents at an arbitrary vacancy density were calculated using an interpolation formula.[16,17] Interestingly, we find that for all possible combinations of vacancy energy and density, the gate currents are still dominated by cold electrons originating in the contact regions. We have calculated the direct gate current densities from the source contact for the 50 nm transistor with a 1.3 nm oxide and the 90 nm transistor with a 2.9 nm oxide for defect densities in the range of 10^{10}–10^{13} cm^{-2} and a homogeneous as well as various Gaussian distributions of E_{vac} in energy space. The magnitude of the defect-induced current increase is very sensitive to the density and the energy distribution of the defects. For defect densities greater than 10^{12} cm^{-2}, the enhancement can be as high as 2–3 orders of magnitude. Also, the resonant effects are somewhat less pronounced for the thinner oxide.

We regard this work as the first steps toward the full understanding of oxide tunneling from a microscopic point of view. The theoretical approach presented here[13,16,17] could certainly be applied to other systems; and there are other methods to calculate electron transport at the atomic scale, which are of general interest for those interested in modeling nanoelectronic devices.[18–20]

11.3 Modeling of Quantum Dots and Artificial Atoms

The quantum dot is, in a way, the prototype of any future device that is designed to occupy a minimum of space. It is important in this context that quantum dots can be arranged and interconnected in three dimensions, at least in principle. In the last 10 years, the physics of quantum dots has experienced considerable development because of the manifestation of the discreteness of the electron charge in single-electron charging devices, as well as the analogy between three-dimensionally quantum confined systems and atoms.[21] Early studies were motivated by the observation of single-electron charging in granular metallic islands containing a "small" number of conduction electrons (N~100–1000) surrounded by an insulator characterized by a small capacitance C.[22,23] In metallic dots however, quantum confinement is relatively weak; and the large effective mass of conduction electrons makes the energy spectrum a quasi-continuum with negligible separation between electron states even at low temperature, $\Delta E << k_B T$. Hence, the addition of an electron to the island requires the charging energy $e^2/2C$ from a supply voltage source to overcome the electrostatic repulsion or Coulomb blockade from the electrons present in the dot, with negligible influence of the energy quantization in the system.[24]

Advances in patterning and nanofabrication techniques have made possible the realization of semiconductor quantum dots with precise geometries and characteristic sizes comparable to the de Broglie wavelength of charge carriers.[25] These quantum dots are realized in various configurations by combining heterostructures and electrostatic confinement resulting from biased metal electrodes patterned on the semiconductor surfaces. In three-dimensional confined III-V compound semiconductors, the small effective mass of conduction electrons results in an energy spectrum of discrete bound states with energy

© 2003 by CRC Press LLC

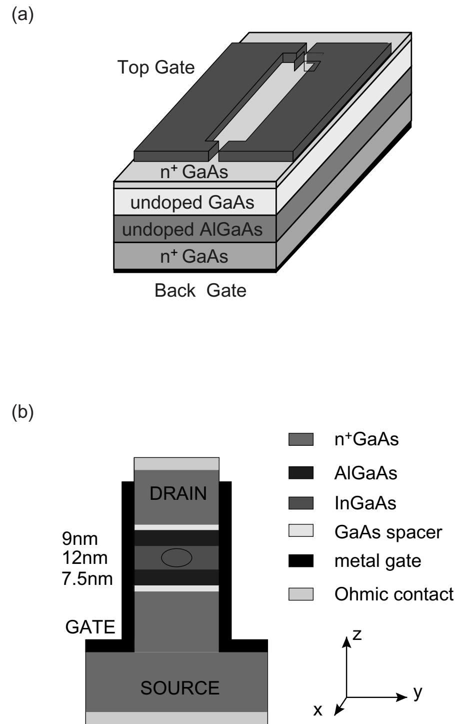


FIGURE 11.5 (a) Schematic representation of a planar quantum dot structure with layered materials; the dark areas represent the confining metallic gates at the surface. (b) Vertical quantum dot structure with different constituting materials; the vertical dark areas on the side represent the controlling metallic gate.

separation comparable to, or even larger than, the charging energy $e^2/2C$. The ability to vary the electrostatic potential over large voltage ranges allows for fine-tuning of the quantum dot charge of just a few electrons (N~1–10).[26] Early experiments on single-electron charging were made with layered AlGaAs/GaAs structures by patterning several Schottky metal gates on top of a two-dimensional electron gas to achieve lateral confinement. A back gate controls the number of electrons in the two-dimensional gas and the dot (Figure 11.5a).[25] In these planar structures, the current flows parallel to the layers, and the tunneling barriers between the dot and the two-dimensional gas are electrostatically modulated by the top gates. In vertical quantum dots, the electrons are sandwiched vertically between two tunneling hetero-barriers, while the lateral confinement results from a vertical Schottky achieved by deep mesa etching of the multilayer structure (Figure 11.5b).[27,28] In this case current flows perpendicularly to the two-dimensional gas between the two hetero-barriers, which are usually high and thin because they are made of different semiconductor materials, e.g., InGaAs and AlGaAs. In general, planar dots have a poor control of the exact number of electrons, while vertical dots lack the barrier tenability of lateral structures.

In semiconductor quantum dots, discrete energy levels with Coulomb interaction among electrons for achieving the lowest many-body state of the system is reminiscent of atomic structures. In cylindrical quantum dots, shell structures in the energy spectrum and Hund's rule for spin alignment with partial shell filling of electrons have recently been observed.[28] One of the peculiarities of these nanostructures is the ability to control not only the shape of the dot but also the number of electrons through gate electrodes.[29] Hence, *artificial atoms* can be designed to depart strongly from the three-dimensional spherical symmetry of the central Coulomb potential and its nucleus charge. In this context, the physics of a few electrons in quantum dots offers new opportunities to investigate fundamental concepts such as the interaction between charge carriers in arbitrary three-dimensional confining potentials and their elementary excitations. Moreover, because Hund's rule is the manifestation of spin effects with shell

© 2003 by CRC Press LLC

filling in quantum dots, the electron spin can, in principle, also be controlled by the electric field of a transistor gate.[30] The idea of controlling spin polarization, independent of the number of electrons in quantum dots, has practical consequences because it provides the physical ingredients for processing quantum information and making quantum computation possible.[31] In addition, spin degrees of freedom can be utilized for storing information in new forms of memory devices. Aside from the investigation of basic quantum phenomena, artificial atoms are also promising for applications in high-functionality nanoscale electronic and photonic devices such as ultra-small memories or high-performance lasers.[22,32,33]

11.3.1 The Many-Body Hamiltonian of Artificial Atoms

The electronic spectrum of N-electron quantum dots are computed by considering the many-body Hamiltonian:

$$\hat{H} = \sum_i \hat{H}_{0i} + \sum_{i,j} \hat{H}_{ij} \tag{11.1}$$

where H_{0i} is the single-particle Hamiltonian of the *i*th electron and:

$$H_{ij} = \frac{e^2}{\varepsilon\left|\vec{r}_i - \vec{r}_j\right|} \tag{11.2}$$

is the interaction Hamiltonian describing the Coulomb interaction between carriers. Here ε is the dielectric constant of the material. In the second term of Equation (11.1), the sum is carried out for *i/j*, avoiding the interaction of carriers with themselves. Quite generally, the Hamiltonian Equation (11.1) is used for solving the Schroedinger equation for the many-particle energies and wavefunctions,

$$E = E_N(1, 2, 3, \ldots, N)$$

$$\Psi = \Psi_N(\vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots, \vec{r}_N)$$

which, given the two-body interaction Equation (11.2), can only be solved exactly for $N = 2$. In this section we will describe a natural approach toward the solution of this problem for a general number N of electrons by considering successive approximations.

11.3.1.1 Single-Particle Hamiltonian and Shell Structures

We start by considering a system of independent and three-dimensional confined electrons in the conduction band. By neglecting the interaction H_{ij}, the Hamiltonian Equation (11.1) is reduced to a sum of single particle Hamiltonians, each of the same form:

$$\hat{H}_{0i} = \hat{H}_i = \frac{\hat{p}_{xi}^2 + p_{yi}^2 + p_{zi}^2}{2m^*} + \hat{V}(\vec{r}_i)$$

Here we assume the electrons can be described with an effective mass m^*; p_{xi}, p_{yi}, and p_{zi} are the components of the *i*th electron momentum, and $V(r_i)$ is the external potential that contains several contributions according to the confinement achieved in the quantum dot. We will assume that the quantum dot is realized by confinement of the electrons in a heterostructure quantum well along the y-direction and electrostatic confinement in the x–z plane (Figure 11.6a). The latter confinement results usually from dopant atoms in neighboring semiconductor layers and from the fringing field of the metal electrodes on the semiconductor surface. This configuration is most commonly achieved in planar quantum dots and vertical quantum dots and results in a first approximation in a two-dimensional parabolic potential in the x–z plane (Figure 11.6b). Confinement at the heterostructure along the y-direction is generally strong (~10 nm) with energy separation of the order of 50–100 meV, while the x–z

© 2003 by CRC Press LLC

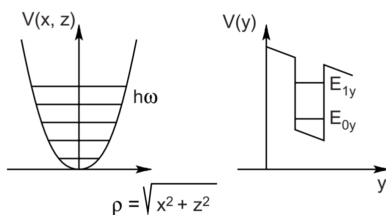


FIGURE 11.6 Schematic representation of (a) a two-dimensional parabolic potential with cylindrical symmetry in the x–z plane showing equally spaced energy levels, and (b) the square potential with the first two quantized levels in the y direction with E_{2y}–E_{1y}>>hω.

planar confinement is much weaker with energy separation of the order of 1 meV over larger distance (>~100 nm). In that case, the external potential is separable in a first approximation,

$$\hat{V}(\vec{r}) = \hat{V}_1(x, z) + \hat{V}_2(y),$$

which results in the energy spectrum $E_{\nu, nx, nz} = E_\nu + E_{n_x, n_z}$ with corresponding wavefunctions $\Psi_\nu(y)\Psi_{n_x, n_z}(x, z)$, where $E_\nu(E_{n_x, n_z})$ is the spectrum resulting from the y-potential (x–z potential). Hence, each value of the ν-quantum number gives a series of x–z energy levels. At low temperature, given the large separation between the E_ν energy states, only the first levels of the lowest series ν = 0 are occupied by electrons. If one further assumes that the $\hat{V}_1(x, z)$ potential is cylindrically symmetric, the ν = 0 energy spectrum is written as:[34]

$$E_{0, n_x, n_z} = E_{0, m, l} = E_0 + m\hbar\omega$$

where ω is frequency of the cylindrical parabolic potential. Here each *m*-level is 2*m*-times degenerate, with the factor 2 accounting for the spin degeneracy. The number m (= 1,2,3…) is the radial quantum number, and the number *l* (= 0,±1,±2,...) is the angular momentum quantum number. Hence the two-dimensional cylindrical parabolic potential results in two-dimensional s,p,d,f,...-like orbitals supporting 2, 4, 6, 8… electrons, which give rise to shell structures filled with 2, 6, 12, 20, … particles, thereby creating a sequence of numbers that can be regarded as the two-dimensional analogues of *magic numbers* in atomic physics.[28,35]

In the absence of cylindrical or square symmetry, the parabolic potential is characterized by two different frequencies, ω_x and ω_z, which lift the azimuthal degeneracy on the *l*-number of the two-dimensional artificial atoms. Therefore, electronic states are spin-degenerate only and determine a sequence of shell filling numbers 2, 4, 6, 8, … of period or increment 2. Only when the ratio ω_x/ω_z is commensurable does the sequence of filling numbers deviate from the period 2 and provide a new sequence of numbers for particular combinations of the n_x and n_z quantum numbers in the case of accidental degeneracy.[36]

Another important class of three-dimensional confined systems includes quantum dots obtained by self-assembled or self-organized Stranski–Krastanov (SK) epitaxial growth of lattice-mismatched semiconductors, which results in the formation of strained-induced nanoscale islands of materials. InAs and InGaAs islands on GaAs have been obtained with this technique in well-controlled size and density.[33,37,38] For these materials, shapes vary between semispherical and pyramidal form, and the size is so small that these quantum dots only contain one three-dimensional fully quantized level for conduction electrons.

11.3.1.2 Hartree–Fock Approximation and Hund's Rules

The natural extension of the atomic model for independent three-dimensional confined electrons is the consideration of the Coulomb interaction between particles in the Hartree–Fock (HF) approximation. The HF scheme has the advantage of conserving the single-particle picture for the many-body state of the system by representing the total wave function as a product of single-particle wavefunctions in a

© 2003 by CRC Press LLC

Slater determinant that obeys Fermi statistics. The main consequence of the HF approximation for the Coulomb interaction among particles is a correction of two terms to the single-particle energies derived from the H_o Hamiltonian[39]

$$E_i = E_{\nu, n_x, n_z} + \frac{e^2}{\varepsilon}\sum_{j\neq i}\int d\vec{r}_i d\vec{r}_j \frac{\left|\Psi_i(\vec{r}_i)\right|^2\left|\Psi_j(\vec{r}_j)\right|^2}{\left|\vec{r}_i - \vec{r}_j\right|}$$
$$-\frac{e^2}{\varepsilon}\sum_{j\neq i}\int d\vec{r}_i d\vec{r}_j \frac{\Psi_i^*(\vec{r}_i)\Psi_j(\vec{r}_i)\Psi_i(\vec{r}_j)\Psi_j^*(\vec{r}_j)}{\left|\vec{r}_i - \vec{r}_j\right|}$$

where the first sum is the Hartree energy carried on all occupied *j*-states different from the *i*-state, irrespective of their spins, and accounts for the classical repulsion between electrons. The second term is the attractive exchange interaction that occurs among carriers with parallel spins. In this scheme, the wavefunctions $\Psi_i(\vec{r}_i)$ satisfy the HF integro-differential equation where the Coulomb interaction term depends upon all the other single-particle wavefunctions of the occupied states. The HF equation is therefore nonlinear and must be solved self-consistently for all wavefunctions of occupied states.

One of the important consequences of the HF approximation for interelectron interaction Equation (11.2) is the prediction of spin effects in the shell filling of artificial atoms similar to Hund's rules in atomic physics.[28] These effects are illustrated in the charging energy of a few electron quantum dots with a cylindrical parabolic potential as achieved in planar or vertical quantum structures.[26] In Figure 11.7a, we show schematically the Coulomb staircase resulting from charging a quantum dot with a few electrons as a function of the charging energy or voltage between the metal electrode or gate and the semiconductor substrate. The relative step sizes of the staircase represent the amount of energy needed to put an additional electron in the dot. The arrows on each step represent the spin of each individual electron on the successive orbitals during the charging process. The filling of the first shell (s-orbital with 2 electrons) consists of one electron with spin-up followed by an electron with spin-down. The step size of the spin-up electron measures the charging energy needed to overcome the Coulomb repulsion against the spin-down electron, which is only the Hartree energy between the two particles. The larger step size of the second (spin-down) electron is due to the fact that the charging of the third electron requires the charging energy augmented by the energy to access the next quantized level, which is the first p-orbital. The latter

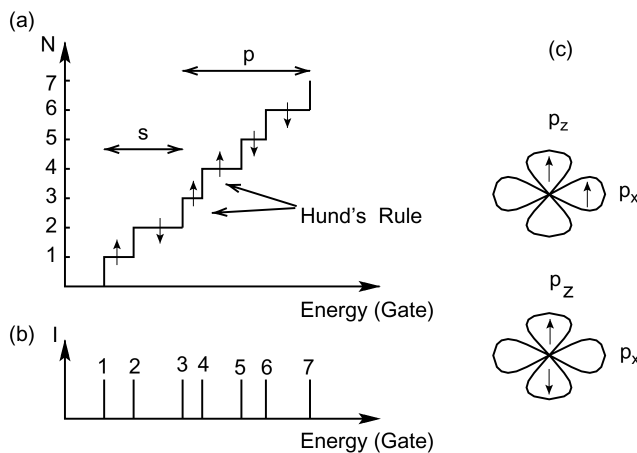


FIGURE 11.7 (a) Coulomb staircase as a function of the charging energy with the spin states of each electron. N is the number of electrons, and the horizontal two-head arrows indicate the occupation of the s- and p-orbitals in the dot. (b) Electron current through the dot vs. the charging energy. (c) Two-dimensional p-orbitals illustrating the two possible occupations of two electrons with parallel (top diagram) and anti-parallel spins (bottom diagram).

© 2003 by CRC Press LLC

process starts the second shell filling with the third electron on either one of the degenerate $l = \pm1$ orbitals of either spin (here we choose the $l = -1$ and the spin-up). At this stage, the configuration with the fourth electron on the $l = 1$ orbital with a parallel spin becomes more favorable because it minimizes the Hartree energy between orbitals of different quantum numbers and results in an attractive exchange energy between the two electrons (Figure 11.7c). This is the reason the third step is smaller than the first and the second steps, requiring less energy and demonstrating Hund's rule in the electron filling of the two-dimensional artificial atom. The fourth step is long because the addition of the fifth electron on either of the p-orbitals with $l = \pm1$ must correspond to a spin-down electron that undergoes a repulsion from the two other p-electrons without benefiting from the exchange because its spin is anti-parallel. Figure 11.7b shows the current peaks resulting from the single-electron charging of the quantum dot, which is obtained by differentiating the Coulomb staircase. Current characteristics with similar structure have recently been observed in gated double-barrier GaAs/AlGaAs/InGaAs/AlGaAs/GaAs vertical quantum dot tunnel devices, which revealed the shell structure for a cylindrical parabolic potential as well as spin effects obeying Hund's rule in the charging of the dot.[28]

Hence the HF approximation provides a reasonable picture of the contribution of electron–electron interaction and spin effects in the spectrum of quantum dots. However, it is well known from atomic physics and theoretical condensed matter physics that this approximation suffers from two important drawbacks: neglect of electron correlation and overestimation of the exchange energy.[40] Moreover, it leads to tedious solution of the self-consistent problem when involving a large number of electrons.

11.3.1.3 Full-Scale Simulation of Quantum Dot Devices

Advances in computer simulation combine the sophistication of realistic device modeling with the accuracy of computational physics of materials based on the density functional theory (DFT).[41–45] These powerful methods provide theoretical tools for analyzing fine details of many-body interactions in nanostructures in a three-dimensional environment made of heterostructures and doping, with realistic boundary conditions. *Microscopic* changes in the quantum states are described in terms of the variation of *macroscopic* parameters such as voltages, structure size, and physical shape of the dots without *a priori* assumption on the confinement profile. Consequently, engineering the exchange interaction among electrons for achieving controllable spin effects in quantum devices becomes possible.

The implementation of a spin-dependent scheme for the electronic structure of artificial molecules involves the solution of the Kohn–Sham equation for each of the spins, i.e., up ($\uparrow$) and down ($\downarrow$). Under the local spin density approximation within the DFT, the Hamiltonian $H^{\uparrow(\downarrow)}$ for the spin $\uparrow$ ($\downarrow$) electrons reads[40,46]

$$\hat{H}^{\uparrow(\downarrow)} = -\frac{\hbar^2}{2}\nabla\left[\frac{1}{m^*(\vec{r})}\nabla\right] + E_c(\vec{r}) + \mu_{xc}^{\uparrow(\downarrow)}[n]$$

where $m^*(r)$ is the position-dependent effective mass of the electron in the different materials, $E_c(r) = e\phi(r) + \Delta E_{os}$ is the effective conduction band edge, $\phi(r)$ is the electrostatic potential which contains the Coulomb interaction between electrons, and ΔE_{os} is the conduction band offset between GaAs and AlGaAs. The respective Hamiltonians are identical in all respects, except for the exchange-correlation potential, which is given by

$$\mu_{xc}^{\uparrow(\downarrow)} = \frac{d(n\varepsilon_{xc}[n])}{dn^{\uparrow(\downarrow)}}$$

where ε_{xc} is the exchange-correlation energy as a function of the total electron density $n(r) = n^{\uparrow}(r) + n^{\downarrow}(r)$ and the fractional spin polarization $\xi = (n^{\uparrow} - n^{\downarrow})/n$, as parameterized by Ceperley and Alder.[47] While it is known that the DFT underestimates the exchange interaction between electrons, which leads to incorrect energy gaps in semiconductors, it provides a realistic description of spin–spin interactions in quantum nanostructures, as shown in the prediction of the addition energy of vertical quantum dots (see Figure 11.8).

© 2003 by CRC Press LLC

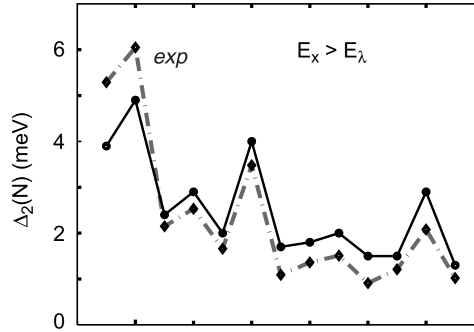


FIGURE 11.8 Addition energy of a vertical single quantum dot. (Data from Tarucha, S., Austing, D.G., Honda, T., van der Hage, R.J., and Kouwenhoven, L.P., *Phys. Rev. Lett.* 77, 3613, 1996; Nagaraja, S., Leburton, J.P., and Martin, R.M., *Phys. Rev. B* 60, 8759, 1999.)

The three-dimensional Poisson equation for the electrostatic potential $\phi(r)$ reads

$$\vec{\nabla}[\varepsilon(\vec{r})\vec{\nabla}\phi(\vec{r})] = -\rho(\vec{r})$$

Here $\varepsilon(r)$ is the permittivity of the material, and the charge density ρ is comprised of the electron and hole concentrations as well as the ionized donor and acceptor concentrations present in the respective regions of the device. The dot region itself is undoped or very slightly p-doped. At equilibrium, the electron concentrations for each spin in the dots are computed from the wavefunctions obtained from the respective Kohn–Sham equations, i.e., $\rho(r) = en(r)$ with $n^{\uparrow(\downarrow)} = \Sigma_\iota |\psi_\iota^{\uparrow(\downarrow)}(r)|^2$. In the region outside the dots, a Thomas–Fermi distribution is used so that the electron density outside the dot is a simple local function of the position of the conduction band edge with respect to the Fermi level, ε_F. The various gate voltages — V_{back}, V_t, and those on the metallic pads and stubs — determine the boundary conditions on the potential $\phi(r)$ in the Poisson equation. For the lateral surfaces in the x–y plane on Figure 11.5, vanishing electric fields are assumed.

Self-consistent solution of the Kohn–Sham and Poisson equations proceeds by solving the former for both spins, calculating the respective electron densities and exchange correlation potentials, solving the Poisson equation to determine the potential $\phi(r)$, and repeating the sequence until the convergence criterion is satisfied.[44] Typically, this criterion is such that variations in the energy levels and electrostatic potential between successive solutions are below 10^{-6} eV and 10^{-6} V, respectively.

The determination of N_{eq}, the number of electrons in the dots at equilibrium for each value of the gate and tuning voltages, is achieved by using Slater's transition rule:[48]

$$E_T(N+1) - E_T(N) = \int_0^1 \varepsilon_{LOA}(n)dn \simeq \varepsilon_{LOA}\left(\frac{1}{2}\right) - \varepsilon_F$$

where $E_T(N)$ is the total energy of the dot for N electrons and $\varepsilon_{LOA}(1/2)$ is the eigenvalue of the lowest-available orbital when it is occupied by 0.5 electron. From the latter equation, it is seen that if the right-hand side is positive, $N_{eq} = N$; otherwise, $N_{eq} = N+1$. Thus the $N \rightarrow N+1$ transition points are obtained by populating the system with $N+0.5$ electrons and varying V_{back} until $\varepsilon_{LOA}(1/2) - \varepsilon_F$ becomes negative. It should be noted that the approximation made in the latter equation is valid only if ε_{LOA} varies linearly with N. This approach has been very successful in the analysis of the electronic spectra and charging characteristics of vertically confined quantum dots.[49] Figure 11.8 shows the addition energy spectrum of a single vertical quantum dot as a function of the number N of electrons in the dot. The addition energy measures the energy required to add a new electron in the dot given the presence of other electrons already in the dot and the restriction imposed by the Pauli principle on the electron energy spectrum.

© 2003 by CRC Press LLC

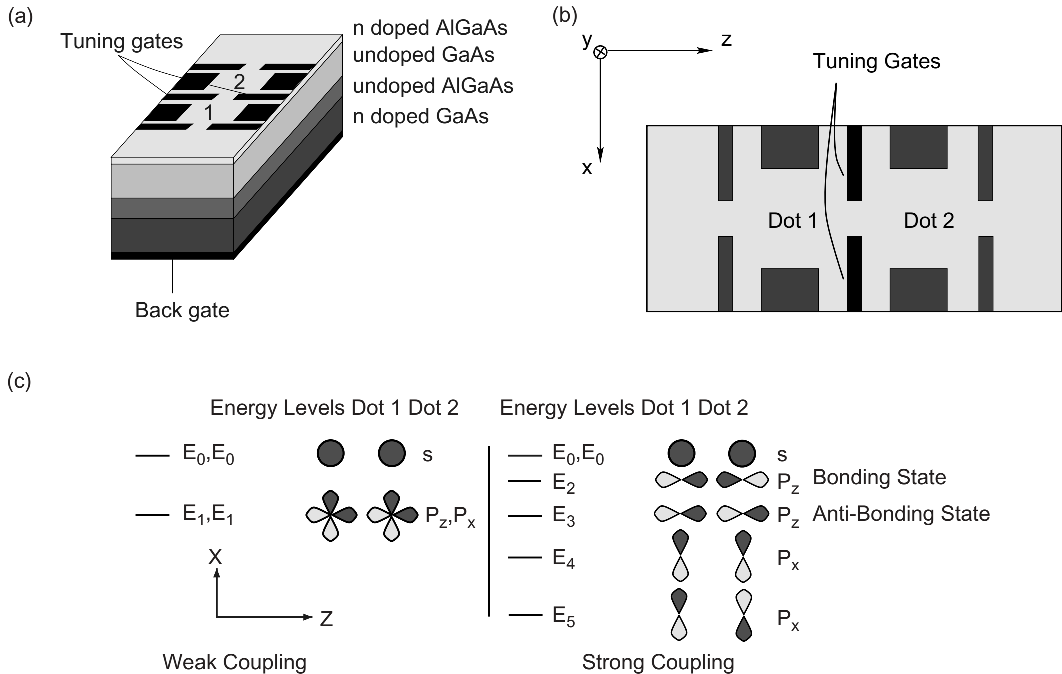


FIGURE 11.9 Schematic representation of the planar coupled quantum dot device. (a) Layer structure with top and back gates; (b) top view of the metal gate arrangement with sizes and orientations; (c) schematic representation of the six lowest orbitals in the weak (left-hand side) and strong (right-hand side) coupling regimes. In both cases, the s-states are strongly localized in their respective dot. Left: p_x- and p_z-like orbitals are degenerate within each dot and decoupled from the corresponding state in the other dot. Right: Increasing coupling lifts the p-orbitals degeneracy with a reordering of the states. The dark and light orbitals indicate positive and negative parts of the wave functions, respectively.

The peaks at N = 2, 6, and 12 are the signature of the existence of a two-dimensional shell structure in the dot, while the secondary peaks at N = 4 and 9 reflect the existence of Hund's rule at half-filled shells. The agreement between theory and experimental data is excellent for the position of the peaks as well as for their magnitude.

The technique is useful in designing double quantum dots with variable interdot barrier for controlling electron–electron interactions.[50] The devices have a planar geometry made of GaAs/AlGaAs heterostructure that contains a two-dimensional electron gas (2DEG). The dots are defined by a system of gate pads and stubs that are negatively biased to deplete the 2DEG, leaving two pools of electrons that form two quantum dots connected in series (Figure 11.9).[46] In these *artificial diatomic molecules*, electron states can couple to form covalent states that are delocalized over the two dots, with electrons tunneling between them without being localized to either.[51] These *bonding states* have lower energy than the constituent dot states by an amount that is equivalent to the binding energy of the molecule. In our case, the dimensions are such that the electron–electron interaction energy is comparable to the single-particle energy level spacing. The number of electrons in the dot, N, is restricted to low values in a situation comparable to a light diatomic molecule such as H-H or B-B. The coupling between dots can be adjusted by varying the voltage on the tuning gates V_t to change the height of the barrier between the two dots. The number of electrons N in the double dot is varied as the 2DEG density, with the back gate voltage V_{back} for a fixed bias on the top gates. Hence, controllable exchange interaction that gives rise to spin polarization can be engineered with this configuration by varying N and the system spin, independently.*

*This is not the case in shell filling of single quantum dots because the total spin of the electronic system is directly related to the number N of electrons in the dot.

© 2003 by CRC Press LLC

11.3.2 Quantum Modeling of Artificial Molecules and Exchange Engineering

In order to simulate these effects, we consider a structure that consists of a 22.5 nm layer of undoped $Al_{0.3}Ga_{0.7}$ as followed by a 125 nm layer of undoped GaAs, and finally an 18 nm GaAs cap layer (Figure 11.9a). The latter is uniformly doped to $5.10^{18}cm^{-3}$ so that the conduction band is immediately above the Fermi level at the boundary between the GaAs cap layer and the undoped GaAs. The inverted heterostructure is grown on the GaAs substrate. The lateral dimensions of the gates and spacing are shown in Figure 11.9b. Figure 11.9c shows the schematic of the lowest four states with their wavefunctions in the double dot for two different tuning voltages.[46] For both values of V_t, the ground state in the individual dots is s-like and forms a degenerate pair. Here, we borrow the terminology of atomic physics to label the quantum dot states. The first excited states, which are px- and pz-like, are degenerate for weak interdot coupling, whereas for strong coupling, the pz-like states mix to form symmetric (*bonding*) and antisymmetric (*anti-bonding*) states that are lower in energy than the px-like states as seen in Figure 11.9c. This reordering of the states has an important bearing on the spin polarization of the double-dot system, as shall be fully explained below.

In the present double-dot structure, we focus on the spin states of the electron system and allow N to vary from zero to eight for two values of V_t: $V_t = -0.67$V, defined as the weak coupling regime, and $V_t = -0.60$V, defined as the strong coupling regime. Electron spin states that are relevant in this analysis are designated by $s_1^{\uparrow(\downarrow)}$, $s_2^{\uparrow(\downarrow)}$ (lower energy s-states in dots 1 and 2) and $p_{x1}^{\uparrow(\downarrow)}$, $p_{x2}^{\uparrow(\downarrow)}$, $p_{z1}^{\uparrow(\downarrow)}$, and $p_{z2}^{\uparrow(\downarrow)}$- (higher energy p-states where the x- and z-indices indicate the orientation of the wavefunctions). For $N = 0$, in the weak coupling limit, the computer model shows that s-states in dots 1 and 2 have negligible overlap because of the relatively high and wide barrier. Indeed, the bonding–anti-bonding energy separation resulting from the coupling between these states is orders of magnitude smaller than the Coulomb charging energy so that s-electrons are practically localized in each dot. A similar situation arises for the p_x- and p_z-states, which, although experiencing slight overlap because of higher energy, they see a lower and thinner barrier and are quasi-degenerate within each dot. In fact, in the weak coupling limit, p_{z1}- and p_{z2}-states that are oriented along the coupling direction between dots experience a bigger overlap than the corresponding p_x-states and consequently lie slightly lower in energy than the latter. Hence, as far as the lower s- and p-states are considered, the double-dot system behaves as two quasi-independent dots (Figure 11.9c, left). In addition, because of the large distance separating the two lower s-states for $V_t = -0.67$V, Coulomb interaction between electrons in dots 1 and 2 is negligibly small, and both dots can be charged simultaneously through double charging[46] to completely fill the s1 and s2 states. Therefore, for $N = 4$, there is no net spin polarization in the double dot, because both contain equal numbers of spin ↑ and spin ↓ electrons.

When the double dot is charged with a fifth electron, the latter occupies either the p_{z1-} or p_{z2-}state (e.g., ↑ spin i.e., $p_{z1}^{\uparrow}$or $p_{z2}^{\uparrow}$) that has the lowest available energy. The sixth electron takes advantage of the nonzero p-state overlap and occupies the other $p_z^{\uparrow}$-state with a parallel ↑ spin. The seventh and eighth electrons find it energetically favorable to occupy successively $p_{x1}^{\uparrow}$ and $p_{x2}^{\uparrow}$, but not any of the spin ↓ states, because of the attractive nature of the exchange-correlation energy among the spin ↑ electrons that results from the nonzero p-state overlap, lowering the energy of the double dot. This particular high-spin configuration among p-orbitals in the "artificial" diatomic molecule deserves special attention because it appears to violate one of the Zener principles on the onset of magnetism in transition elements; this principle forbids spin alignment for electrons on similar orbitals in adjacent atoms.[52] Therefore, it could be argued that the high-spin configuration obtained in the calculation is the consequence of a DFT artifact. Recently, however, Wensauer et al. confirmed the DFT results based on a Heitler–London approach.[53] Similar conclusions have also been obtained by an "exact" diagonalization technique on vertically coupled quantum dots for $N = 6$ electrons.[54,55] Let us point out that Zener's principle is purely empirical, as it is based on the observation of the magnetic properties of natural elements that lacks the tunability of artificial systems. Therefore, the total spin of the double dot can possibly steadily increase by $1/2\hbar$ for each electron added after the fourth electron to $2\hbar$ for $N = 8$, and there is no contradiction with Zener's principle applied to

© 2003 by CRC Press LLC

TABLE 11.1 Spin of the Double Dot for Various Occupation Numbers in the Two Coupling Regimes

	V_t \ N	1	2	3	4	5	6	7	8	9	10	11	12
Spin	–0.67 V	1/2	?	1/2	0	1/2	1	3/2	2	3/2	1	1/2	0
($\hbar$)	–0.67 V	1/2	0	1/2	0	1/2	0	1/2	0	1/2	0	1/2	0

The question mark at N = 2 in the weak coupling regime indicates that the spins are uncorrelated.

natural elements. After all, high-spin configurations have been shown to compete for the ground state of light diatomic molecules such as B_2.[56]

The variation of the total spin S in the double dot with N is shown in Table 11.1. It is also seen that as N increases above eight electrons, the spin ↓ states start to be occupied, thereby decreasing S by $1/2\hbar$ for each additional electron, forming anti-parallel pairs to complete the shell until $N = 12$ when S is reduced to zero. The sequence of level filling with the occupation of degenerate states by electrons of parallel spin is observed in atoms and is governed by Hund's rules; it is therefore impressive that similar rules successfully govern level filling in the double dot in the weak coupling regime. Let us point out that even though $S = 2\hbar$ is the most favored state of the double dot, energetically it is not significantly lower than other competing states for $N = 8$. For instance, the excited states with $S < 2\hbar$ for $N = 8$ are only about 0.1 meV higher in energy. Consequently, for this particular double-dot structure, any attempt to observe experimentally the parallel alignment of the spins of unpaired electrons is restricted to low temperatures for which $k_BT << 0.1$ meV, or any kinds of electrostatic fluctuations smaller than this value. However, it must be noted that the structure is not optimized and that the evidence of spin polarization among p-states in the double-dot system may be achieved in smaller dots with stronger exchange interaction. The key issue here is the fact that the quantum mechanical coupling between the two dots in this bias regime is not strong enough to lift the spatial *quasi-degeneracy* among p_x- and p_z-states which, for our particular configuration, were separated by no more than a few microelectron volts. Stronger coupling between the quantum dots eliminates this effect. Accordingly, if V_t increases to –0.60 V, also referred to as the strong coupling regime, the p-state spatial degeneracy is completely lifted, while deeper s-states also couple, although to a slighter extent to lead to the spectrum of Figure 11.9c, right. Therefore, the spin sequence as a function of N is alternatively $S = 1/2\hbar$ for odd N when the last occupying electron is unpaired and $S = 0$ for even N when it pairs up with an electron of the opposite spin (Table 11.1).

The variation of inter-dot coupling by varying V_t provides a control of direct exchange interaction between p-like electrons in the two dots that may be more robust than for s-electrons. Hence, a lowering of the inter-dot barrier results in a reordering of the single-particle levels, thereby transforming the double-dot (for $N = 8$) from a spin-polarized $S = 2\hbar$ to an unpolarized state $S = 0$. An important result from Table 11.1 is that the Loss–Di Vincenzo scheme for quantum computing with double dots could also be achieved for $N = 6$ electrons, where the control of qubit entanglement for a quantum control-not (XOR) gate operation would be realized with the $S = 1/2\hbar$ spin states of two p-electrons instead of two single electrons ($N = 2$) in the original scenario.[57]

The electrostatic nature of the confinement potential, specifically the coupling barrier, is central to the occurrence of the effects mentioned above. Indeed, the barrier is not uniform but is wider (and higher) for the lower quantum dot s-states than for the higher p-states (Figure 11.9c). This situation is similar to the electronic properties of natural diatomic molecules, where the strongly localized s-states correspond to atomic core states and the delocalized p-states to covalent bonding states. It is therefore possible to engineer exchange interaction in the *artificial molecule* by suitably tailoring the coupling barrier between quantum dots. This is achievable by proper device design, i.e., by adjusting gate size and shape, the doping profiles, and the distance between the GaAs/AlGaAs heterojunction and the control gates, and possibly by choosing other III-V semiconductor systems to optimize the energy separation between singlet and multiplet states.

© 2003 by CRC Press LLC

11.4 Carbon Nanotubes and Nanotechnology

Device aspects of carbon nanotubes represent an interesting new area of nanoscience and nanotechnology. Simulation in this area requires special methods that lie in between the methods used for the periodic solid and the methods applied to quantum dots. Various (carbon, nitride, and chalcogenide) nanotubes are promising for applications because of their unusual mechanical and electronic properties, stability, and functionality.

The lattice structure of single-wall carbon nanotubes follows the lattice structure of graphene (monolayer of graphite): a hexagonal pattern is repeated with translational symmetry along the tube axis and with axial (chiral) symmetry along the tube circumference. Nanotubes are labeled using two numbers [n,m]. These are components of the vector that generates the tube circumference after scrolling, in terms of basic vectors of the graphene lattice (see Figure 11.10A). It is easy to find that only two types of single-wall nanotubes (SWNTs) have a pure axial symmetry: so-called armchair (A) and zigzag (Z) nanotubes. The graphene rectangle shown in Figure 11.10A, gives an armchair (A) nanotube when wrapped from top to bottom (Figure 11.10B) and a zigzag (Z) nanotube when wrapped from left to right. Any other type of nanotube is chiral, which means that it belongs to a screw-axis symmetry group.

Graphite-like systems and materials, such as fullerenes, nanotubes, nanographites, and organic macromolecules, are well known to have valence/conduction band systems generated by pi and sigma valence electrons.[58] The latter ones are localized and, normally, contribute only to the mechanical properties of the graphitic material. In contrast, pi-electrons are mobile and highly polarizable and define transport, electrical, and electromechanical properties. The pi-electronic structure of a monolayer of graphite (graphene) is shown in Figure 11.11. It has a six Fermi points that separate an empty conduction band from an occupied (symmetrical) valence band. A simple but correct picture of the electronic structure of a SWNT follows from a band-folding argument: an additional space quantization for the pi-electrons appears due to confinement in the circumferential direction. It can be thought of as a mere cross sectioning of the electronic structure of graphene along the nanotube symmetry direction. Depending on the lattice symmetry of the tube, three different situations can be realized:

1. The armchair SWNT has a cross-section passing through the Fermi point (Figure 11.12, left). In this case the SWNT is metallic and the conduction band merges with the valence band (Figure 11.13, left).

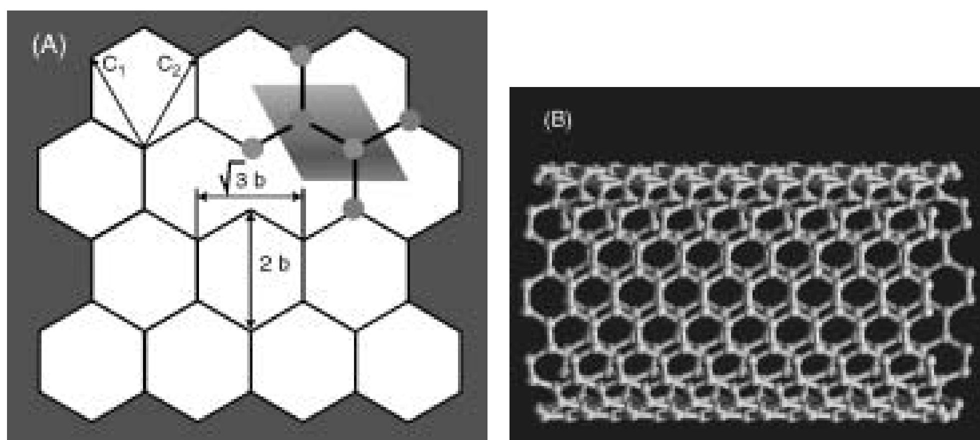


FIGURE 11.10 (A) Honeycomb lattice structure of graphene has a rhombic unit cell with two carbon atoms. Translated along basal vectors, c_1 and c_2, it forms two interconnected sublattices. The carbon–carbon bond length, b, is ~0.14 nm. The edge direction, in basal vectors, is denoted by two integers (shown fragment has left/right edge of type [2,2]). (B) Lattice structure of [10,10] armchair SWNT. Wrapping honeycomb lattice along some chosen axis will form a nanotube.

© 2003 by CRC Press LLC

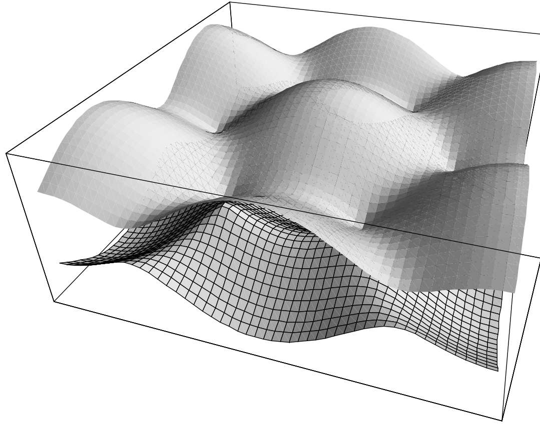


FIGURE 11.11 TBA electronic structure of valence bands of a monolayer of graphite. Threefold symmetry of the lattice results in six Fermi points where the conduction band meets the valence band.

FIGURE 11.12 Lowest conduction sub-band and highest valence sub-band of a metallic tube (left) compared with semiconductor nanotube (right). A nanotube quantization condition makes cut from the cone-shaped bands of graphite. In the case of metallic nanotube, this cross-section passes through the Fermi point and no gap develops between sub-bands. The electron dispersion is linear in longitudinal wave vector. In the case of semiconductor nanotube, the cross-section is shifted away from Fermi point. The carrier dispersion is a hyperbola.

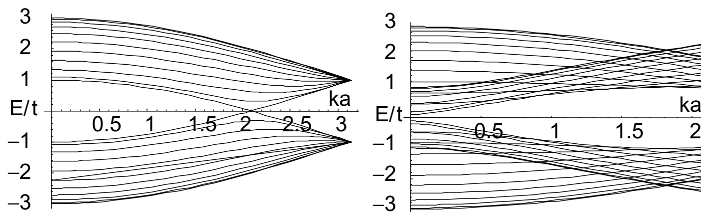


FIGURE 11.13 Electronic structure of a metallic armchair [10,10] nanotube (left) and a semiconductor zigzag [17,0] nanotube (right). The pi-electron energy is plotted in units of hopping integral, t~2.7 eV vs. dimensionless product of the longitudinal wave vector, k, and the bond length, a~0.14 nm (half of the Brilloine zone is shown).

© 2003 by CRC Press LLC

2. The zigzag/chiral nanotube cross-section is distant from the Fermi point (Figure 11.12, right). This tube has a nonzero gap as shown in Figure 11.13, right, and it is a semiconductor tube.
3. One third of zigzag and chiral nanotubes have a very small gap, which follows from arguments other than simple band-folding. In our simplified picture these SWNTs will have a zero band gap and be metals.

The band gap of a semiconductor nanotube depends solely on the tube radius, R — a simple rule follows from the band-folding scheme: $E_g = t\, b/R$, here t~2.7 eV is the hopping integral for pi-electrons. This gap dependence was experimentally measured by Scanning Tunneling Spectroscopy[59] and resonance Raman spectroscopy.[60]

Nanotubes with conductivity ranging from metallic to semiconductor were indeed synthesized. The temperature dependence of the conductivity of the nanotube indicates reliably the metallic or semiconductor character. The field effect is also very useful to distinguish between two types. To measure this effect, the nanotube is placed between two electrodes on top of a back-gate contact that is covered with an insulating layer. After synthesis and purification, a SWNT is normally p-type, i.e., the majority of carries are holes. A typical density of $\sim 10^7$ cm^{-1} holes defines the conductivity in the ON state of a SWNT when operated as a field-effect transistor at zero gate voltage. External positive voltage, applied to the back gate, can deplete the holes and switch the SWNT–FET into the OFF state. Experiments have shown a drop of 5 orders of magnitude of the source-drain current when the gate voltage was changed by 3 volts for very thin insulator layers (thickness < 2 nm).[61] In the case of metallic nanotube bridging two electrodes, only a weak dependence, if any, of the conductivity on the gate voltage is seen.

The electronic structure of SWNT is highly sensitive to external fields, and lattice distortions cause changes in the electronic structure. A lattice distortion moves the Fermi point of graphite and results in the closing/opening of an energy gap, a change in the electron density, and charging of the tube. This opens many possibilities for application of nanotubes to use as nano-biosensors, mesoscopic devices, and nanoelectromechanical systems.

In this section we focus on a particular application of carbon nanotubes — nanoelectromechanical (NEM) switches (Figure 11.14) and nanotweezers (Figure 11.15). The three basic energy domains that

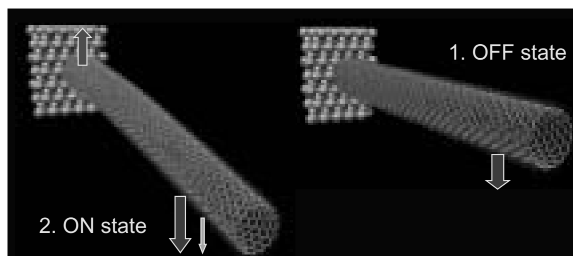


FIGURE 11.14 ON and OFF states of a nanotube electromechanical switch. Arrows show applied forces: electrostatic, van der Waals, and elastic forces.

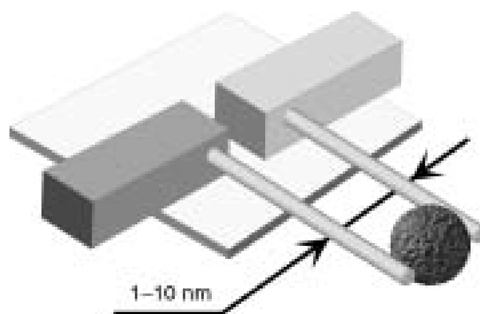


FIGURE 11.15 Nanotube nanotweezers device.

© 2003 by CRC Press LLC

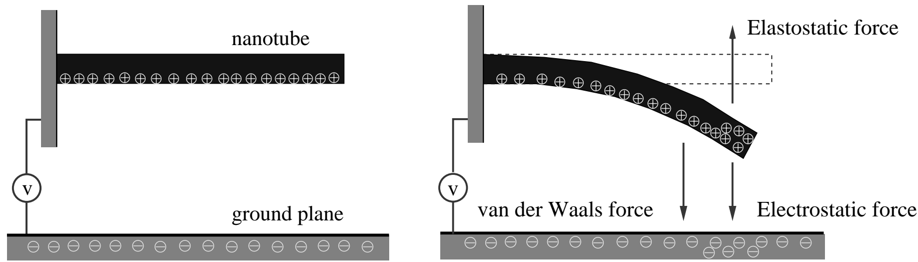


FIGURE 11.16 Force balance for a nanotube over a ground plane: (left) Position of the tube when V = 0; (right) deformed position of the tube when V/0.

describe the physical behavior of NEM switches — mechanics, electrostatics, and van der Waals — are described below.

11.4.1 Operation of Nanoelectromechanical Switches

Shown in Figure 11.16 is the nanoelectromechanical operation of a carbon nanotube-based cantilever switch. The key components are a moveable structure, which can be a single wall or a multiwall carbon nanotube, and a fixed ground plane, which is modeled by a graphite bulk. When a potential difference is created between the moveable structure and the ground plane, electrostatic charges are induced on both the movable structure and the ground plane. The electrostatic charges give rise to electrostatic forces, which deflect the movable tube. In addition to electrostatic forces, depending on the gap between the moveable tube and the ground plane, the van der Waals forces also act on the tube and deflect it. The directions of the electrostatic and van der Waals forces are shown in Figure 11.16. Counteracting the electrostatic and van der Waals forces are elastic forces, which try to restore the tube to its original straight position. For an applied voltage, an equilibrium position of the tube is defined by the balance of the elastic, electrostatic, and van der Waals forces. As the tube deflects, all forces are subject to change, and a self-consistent analysis is necessary to compute the equilibrium position of the tube.

When the potential difference between the tube and the ground plane exceeds a certain critical value, the deflection of the tube becomes unstable and the tube collapses onto the ground plane. The *potential*, which causes the tube to collapse, is defined as the *pull-in voltage* or the *collapse voltage*. When the pull-in voltage is applied, the tube comes in contact with the ground plane, and the device is said to be in the ON state (Figure 11.14.2). When the potential is released and the tube and the ground plane are separated, the device is said to be in the OFF state (Figure 11.14.1).

When compared with microelectromechanical switches, the operation of nanoelectromechanical switches is different because of the importance of the van der Waals forces, which can be neglected at the micrometer scale. The sticking of NEM devices becomes an increasing problem at the nanoscale and can limit the range of operability of NEMS. If the gap between the cantilever tube and the ground plane is very small, even without an applied voltage, the tube can collapse onto the ground plane because of the van der Waals forces. In addition, the separation of the tube from the ground plane after the contact becomes an issue as the van der Waals forces will tend to keep the tube and the ground plane together.

11.4.2 Nanotube Mechanics

Mechanical and structural properties of nanoscale systems have been studied both theoretically and experimentally over the last decade.[62] The strong correlation between the structure and electronic properties of a nanosystem requires a proper understanding of the nanomechanical and nanoelectromechanical behavior of nanotubes. Such studies can lead to new design tools for microscopy and characterization studies as well as development of highly sensitive detectors. The mechanical behavior of a small structure differs from that of a bulk structure. New phenomena such as super-low friction,[63] super-high stiffness,[64] and high cohesion at small distances[65] are encountered.

© 2003 by CRC Press LLC

The mechanical behavior of nanotubes can be modeled either by simple continuum approaches or by more complex atomistic approaches based on molecular dynamics simulations. The elastic properties of pure single-wall and multiwall nanotubes were studied by, for example, Sanchez–Portal et al.[66] and Yakobson and Avouris.[67] Atomistic approaches have the advantage of capturing the mechanical behavior accurately; however, they require large computational resources. Continuum theories, when properly parameterized and calibrated, can be more efficient to understand the mechanical behavior of nanotubes. A simple continuum approach to model the mechanical behavior of nanoelectromechanical switches is based on the beam theory. The beam equation is given by:

$$EI\frac{\partial^4 r}{\partial x^4} = q$$

where r is the gap between the conductor and the ground plane, x is the position along the tube, q is the force per unit length acting normal to the beam, E is the Young's modulus, and I is the moment of inertia and, for nanotubes, can be estimated as:

$$I = \frac{\pi}{4}(R_{ext}^4 - R_{int}^4)$$

where R_{int} is the interior radius and R_{ext} is the exterior radius of the nanotube.

The beam theory can, however, suffer from several limitations. For very large loads, the stress concentration at the edges of the nanotubes may cause the tube to buckle and form kinks. In such cases, the deflection deviates from the beam theory locally. The buckling happens at a certain strain depending on the device geometry, the nanotube symmetry, and the load. If buckling is to be simulated, one can try advanced continuum theories such as a shell theory or a full elasticity theory.[68]

Many-body corrections to van der Waals interactions from semiclassical Casimir forces were calculated and applied in the continuum modeling of nanotube mechanics.[69] The basic analysis of a role of van der Waals terms in electromechanical systems has demonstrated its significance at the sub-nanometer scale.[70] A recent theory[71] of van der Waals interaction for shells of pure carbon is based on universal principles formulated in 1930s.[72] The new approach is based on the quantum electrodynamical description of the van der Waals/Casimir forces. A simple and effective model has been developed to estimate the many-body contribution due to collective modes (plasmons).[69] This contribution is believed to be a major portion of the total van der Waals energy because of the high oscillator strength of the plasmons. The theory reveals many-body terms that are specific for various low-dimensional graphite nanostructures and are not taken into account by standard one-body calculations within the dispersionless model by Lennard–Johns.[72] We have demonstrated the use of the model for several systems (shown in Figure 11.17): a double-wall nanotube (A), a nanotube on the surface (B), and a pair of single-wall tubes (C). A significant difference has been shown for the dependence of the van der Waals energy on distance, which is a consequence of our quantum correction.[69]

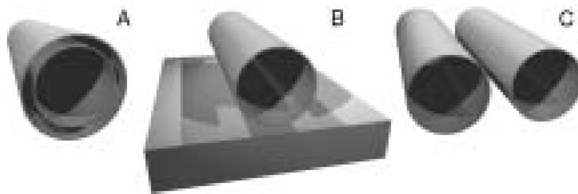


FIGURE 11.17 Geometry of nanotube systems for which a quantum correction to van der Waals forces has been calculated: (A) Double-wall nanotube; (B) single-wall nanotube on a surface; and (C) two single-wall nanotubes.

© 2003 by CRC Press LLC

11.4.3 Electrostatics

The three-dimensional character of the electromagnetic eigenmodes and one-dimensional charge density distribution of a SWNT system result in a weak screening of the Coulomb interaction and the external field. We present a quantum mechanical calculation of the polarizability of the metallic [10,10] tube. The nanotube polarizability is not defined solely by the intrinsic properties of the tube.[73] It depends also on the geometry of the nanotube and closest gates/contacts. Hence, the charge distribution has to be treated self-consistently. Local perturbations of the electronic density will influence the entire system unlike in common semiconductor structures. For example, a point charge placed near the tube surface will generate an induced-charge density along the tube length, which decays very slowly with the distance from the external charge.

Figure 11.18 (after[73]) is a sketch of the depolarization of the tube potential (induced-charge density) by the side electrode and the back-gate (the right part of Figure 11.18 shows the geometry of the device simulated). The continuous line is the statistical approximation (Boltzmann–Poisson equations) which coincides well with the quantum mechanical result (dotted line) except for the quantum beating oscillations at the tube end. The depolarization manifests itself as a significant nonuniformity of the charge along the tube length. This effect is described by the self-consistent compact modeling, which is outlined below.

The potential ϕ^{act} that is induced by a charge density, ρ^{ind}, in one-dimensional systems is proportional to the charge density. Thus, for a degenerate electronic structure of a metallic nanotube in the low-temperature limit, the Poisson equation is effectively reduced to[73,74]

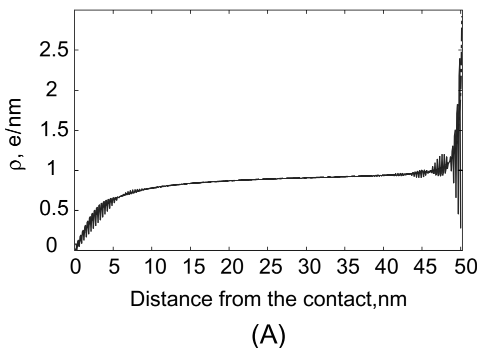
$$\rho^{ind}(z) = -e^2\nu_M\phi^{act}(z).$$

Here ν_M stands for the nanotube density of states, which is constant in a studied voltage range. We have demonstrated that $e^2\nu_M$ acts as an atomistic capacitance of an SWNT:

$$C_A^{-1} = \frac{1}{e^2\nu_M}$$

(a similar quantity for a two-dimensional electron gas system has been introduced by Luryi[75]) and the geometric capacitance:

$$C_m^{-1} = 2\log\left(\frac{2h}{R}\right)$$



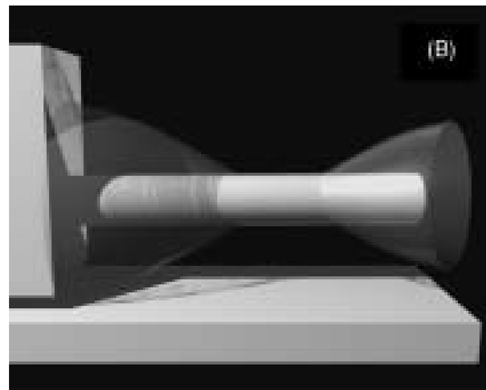


FIGURE 11.18 (A) Self-consistent charge density of a [10,10] armchair nanotube at 5 V voltage applied between side and back-gate contacts. (B) Sketch of the simulated device geometry. The distance between the tube center and the back gate is 5 nm, and the tube radius and length are 0.6 nm and 60 nm.

© 2003 by CRC Press LLC

is a function of distance to the back-gate and SWNT radius. In case of the straight SWNT (as in Figure 11.18) the geometric capacitance is a logarithmic function of the distance between the tube and the gate. In equilibrium, we have the following relation between the equilibrium charge density and external potential (gate voltage), which comprises both the atomistic and geometric capacitances:

$$\rho_\infty = \left(-\frac{\varphi^{xt}}{C_m^{-1} + C_A^{-1}}\right) \simeq -\varphi^{xt} C_m \left(1 - \frac{C_m}{C_A}\right)$$

This equation is still valid for a nanotube of arbitrary shape although no simple expression for the geometric capacitance can be written.

11.4.4 Analytical Consideration for the Pull-In

We finish this section with an analytical model that can be used for a quick estimation of pull-in voltages of the nanotube system within continuum modeling. Assuming that the elastic energy of the NEMS device is given by

$$T = k(h - x)^2/2$$

and the external (electrostatic) force is the gradient of the energy component given by

$$V = C\varphi^2/2$$

we can calculate elastic and electrostatic forces. Then we include the van der Waals energy term:

$$W \simeq \varepsilon x^{-\alpha}$$

and write analytically the pull-in voltage and pull-in gap as functions of the device stiffness, k, the device capacitance, C, and van der Waals energy, W:

$$x_o = hA_1 \frac{1}{2}\left(1 + \sqrt{1 + A_2 \frac{W(x_o)}{kh^2}}\right)$$

$$V_o = B_1 \frac{\sqrt{2}kh}{\sqrt{C(x_o)}} \sqrt{\frac{1}{2} - B_2 \frac{W(x_o)}{kh^2} + \frac{1}{2}\sqrt{1 + A_2 \frac{W(x_o)}{kh^2}}}$$

Here four constants A_1, B_1, A_2, and B_2 are describing the specific dependence of C and W on x, the dynamic gap or the internal coordinate of the NEMS device. In case of a planar switch and the Lennard–Jones potential, these constants are 3/2, $\sqrt{2/3}$, 36, and 36, respectively.

As a result of the van der Waals attraction to the gate, the NEMS device cannot operate at very small gaps, h. The critical gap, h_c, (at which $x_o = 0$) is about 2 nm for the switch with $k \sim W/1\ \text{nm}^2$, and C ~2 $k^{1/2}/(3\ \text{V/nm})$. Next, Figures 11.19 and 11.20 show that, by neglecting the van der Waals correction to the pull-in gap, x_o, one underestimates the critical pull-in voltage by 15%.

The self-consistent solution for the pull-in gap is plotted in Figure 11.20. Again, neglecting the van der Waals terms results in an un-physical divergence of the pull-in gap when approaching the critical distance h_c.

11.4.5 Outlook

Development of fast and precise approaches for three-dimensional device modeling of nanotube systems becomes clearly important after recent successes of the IBM and Delft groups in creating prototypes for nanotube electronics.[61,76] The physics of carbon nanotube devices is rather distinct from the physics of

© 2003 by CRC Press LLC

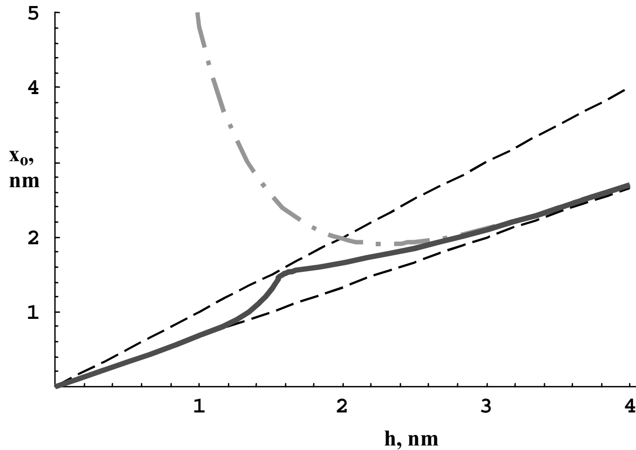


FIGURE 11.19 The pull-in gap as a function of the initial device gap. Solid curve represents the self-consistent analytical result. Dash-dotted curve shows the dependence in neglecting the van der Waals correction.

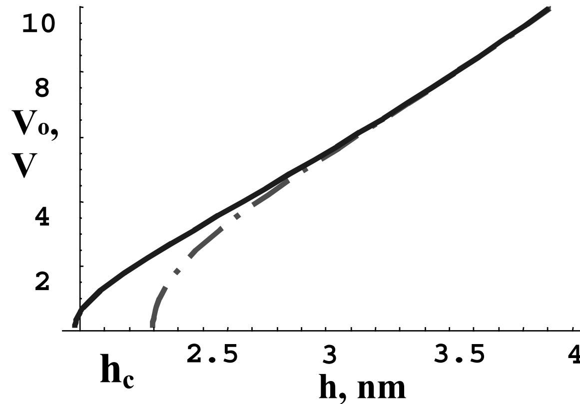


FIGURE 11.20 The pull-in voltage as a function of the gap. Solid curve represents the analytical result explained in the text. Dash-dotted curve shows the dependence in neglecting the van der Waals correction for the pull-in gap.

standard semiconductor devices, and it is unlikely that semiconductor device modeling tools can be simply transferred to nanotube device modeling.

Development of device modeling tools for nanotubes can be very complicated because of the breakdown of continuum theories. Molecular mechanics (MM) and molecular dynamics (MD) can be used reliably when continuum theories break down. However, both MM and MD can be very computer time-consuming. A good compromise is to develop a multiscale approach where continuum theories are combined with atomistic approaches. Multiscale methods can be accurate and more efficient compared with atomistic approaches. The highest level in the multiscale hierarchy is represented by the quantum mechanical result for the single-tube polarizability, which is the atomistic analog of the bulk dielectric function. It contains the complete information for the electronic structure and charge distribution and gives the means for calculating the screened Coulomb and the van der Waals/Casimir forces.[69] The main difficulty here is the requirement to solve the problem for device structures. The electronic structure and the polarizability change during device operation, and this requires a self-consistent treatment. At the intermediate level, classical molecular dynamics provides a detailed knowledge for geometry and material parameters of the system. This is a prerequisite for calculating the mechanical response of the system. It also supplies proper boundary conditions for electrostatic calculations through the actual device geometry. At the lowest level of the simulation hierarchy, the only level that can be used to simulate and

© 2003 by CRC Press LLC

understand larger systems of devices, continuum theories must and can be applied. The parameters of the continuum models will, of course, need to be derived from the higher level simulations.

Modeling and simulation of large-scale nanoscale circuits where carbon nanotubes are interconnected with other nanoelectronic, nanomechanical, chemical, and biological molecules are beyond the capability of currently existing supercomputers. Development of compact models for nanotubes and other nanodevices can enable the design of large-scale nanocircuits for breakthrough engineering applications.

11.5 Simulation of Ionic Channels

Nature has created many forms of nanostructures. Ion channels are of particular importance and have become accessible to the simulation methods that are widely used in computational electronics and for the nanostructures that have been described above. We therefore add this section to emphasize the importance of merging the understanding of biological (i.e., carbon-based) and silicon-based nanostructures.

Found in all life forms, ion channels are in a class of proteins that forms nanoscopic aqueous tunnels in the otherwise almost impermeable membranes of biological cells. An example of an ion channel, *ompF* porin, which resides in the outer membrane of the *E. coli* bacterium, is illustrated in Figure 11.21. Every ion channel consists of a chain of amino acids carrying a strong and rapidly varying permanent electric charge. By regulating the passive transport of ions across the cell membrane, ion channels maintain the correct internal ion composition that is crucial to cell survival and function. Ion channels directly control electrical signaling in the nervous system, muscle contraction, and the delivery of many clinical drugs.[77] Most channels have the ability to selectively transmit or block a particular ion species, and many exhibit switching properties similar to electronic devices. From a device point of view, ion channels can be viewed as transistors with unusual properties: exquisite sensitivity to specific environment factors, ability to self-assemble, and desirable properties for large-scale integration such as the infinite ON/OFF current ratio. By replacing or deleting one or more of the amino acids, many channels can be mutated, altering the charge distribution along the channel.[78] Engineering channels with specific conductances and selectivities are thus conceivable, as well as incorporating ion channels in the design of novel bio-devices.

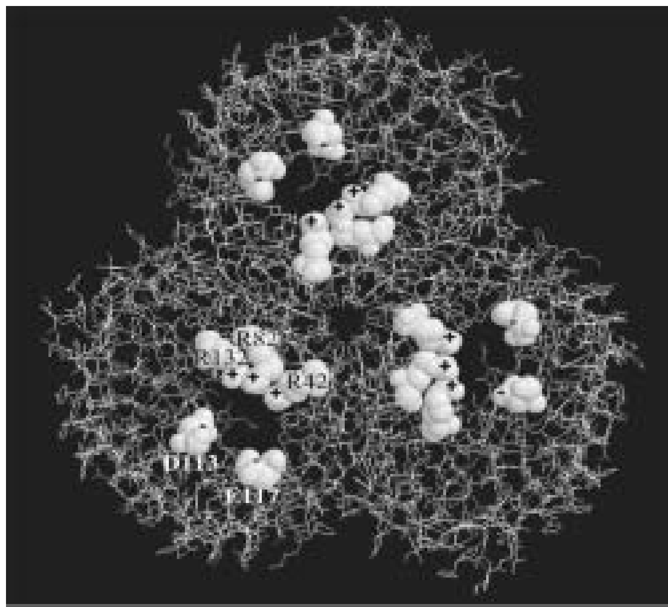


FIGURE 11.21 Molecular structure of *ompF,* a porin channel found in the outer membrane of the *E. coli* bacterium. This projection along the length of the channel shows the threefold symmetry of the trimer. Several ionized amino acids in the constriction region of each pore are highlighted.

© 2003 by CRC Press LLC

Experimentally, the electrical and physiological properties of ion channels can be measured by inserting the channel into a lipid bilayer (membrane) and solvating the channel/membrane in an electrolyte solution. An electrochemical gradient is established across the membrane by immersing electrodes and using different concentrations of salt in the baths on either side of the membrane.

11.5.1 Hierarchical Approach to Modeling Ion Channels

Detailed simulation of ion transport in protein channels is very challenging because of the disparate spatial and temporal scales involved. A suitable model hierarchy is desirable to address different simulation needs. Continuum models, based on the drift-diffusion equations for charge flow, are the fastest approach; but they require large grids and extensive memory to resolve the three-dimensional channel geometry. Ion traversal of the channel is a very rare event on the usual time scale of devices, and the flow is actually a granular process. Continuum models, therefore, are useful mainly to probe the steady-state of the system. We suppose the system to be ergodic. At a given point in the simulation domain, the steady-state ion concentration represents the probability of ion occupation at that position, averaged over very long times or, equivalently, averaged over many identical channels at any given instant. Despite some limitations, continuum models can be parameterized to match current-voltage characteristics by specifying a suitable space and/or energy-dependent diffusion coefficient, which accounts for the ions' interactions with the local environment.

A step above in the hierarchy we find particle models, where the trajectories of individual ions are computed. The simpler model is based on a Brownian Dynamics description of ion flow, in which ion trajectories evolve according to the Langevin Equation. Ions move in the local electric field, calculated from all the charges in the system as well as any externally applied fields. The energy dissipated via ion-water scattering is modeled by including a simple frictional term in the equation of motion, while the randomizing effect of the scattering is accounted for by including a zero-mean Gaussian noise term.[79] Ionic core repulsion can also be included by adding a suitable repulsive term (e.g., Lennard–Jones) to the total force acting on the ion. When the latter is neglected, the simulation is equivalent to a discrete version of the drift-diffusion model.[80] If the ion motion is assumed to be strongly overdamped, relatively long time steps can be used (e.g., picoseconds), making this a very practical approach.

At the next level in the hierarchy are particle models, where the ion flow is resolved with a self-consistent transient, following Monte Carlo or MD approaches, as they are known in semiconductor device simulation. MD simulations resolve the motion and forces among all particles, both free (ions and water molecules) and bound (e.g., protein atoms) in the system. Bound particles are modeled as charged balls connected by springs (chemical bonds). The entire system is brought to a simulated experimental temperature and then equilibrated by allowing the system to evolve according to Newtonian mechanics.[81] While this methodology is the most complete, due to the extreme computational costs involved, it can only be applied today to very small systems on very short time scales of simulation. Monte Carlo methods, originally developed for semiconductor device simulation, provide a more practical compromise. Water and protein are treated as a background dielectric medium, as is done with Brownian Dynamics, and only the individual ion trajectories are resolved. The key difference between Brownian Dynamics and Monte Carlo techniques lies in the way the ion dynamics are handled. In Monte Carlo models the ion trajectories evolve according to Newtonian mechanics; but individual ion-water collision interactions are replaced with an appropriate scattering model, which is resolved on the natural time scales of the problem.[82] In the limit of high friction, both approaches should give the same result.

11.5.2 Drift-Diffusion Models

Drift-diffusion models are useful for studying ion transport in open-channel systems over time scales that cannot be resolved practically by detailed particle models. Water, protein, and membrane are treated as uniform background media with specific dielectric constants; and the macroscopic ion current in the water is resolved by assigning an appropriate space or energy-dependent mobility and diffusion coefficient

© 2003 by CRC Press LLC

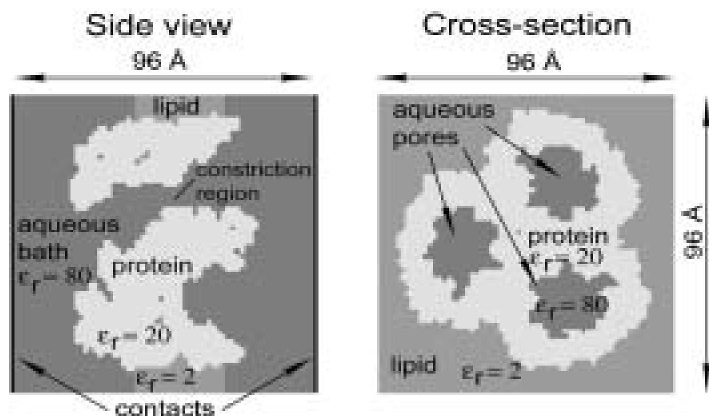


FIGURE 11.22 Mesh representation of the *ompF* trimer *in situ* in a membrane, immersed in a solution of potassium chloride — longitudinal and cross-sectional slices through the three-dimensional computational domain generated on a uniform rectilinear grid (1.5Å spacing). Electrodes immersed in the baths maintain a fixed bias across the channel/membrane system.

to each ionic species. The solution of Poisson's equation over the entire domain provides a simple way to include external boundary conditions and image force effects at dielectric discontinuities. Complete three-dimensional models of flow in ionic channels can be implemented with the established tools of semiconductor device simulation.

In order to define the various regions of the computational domain, the molecular structure of the protein must be mapped onto a grid. Protein structures are known with atomic resolution for a number of important channels, but considerable processing is still necessary to determine the charge and the dielectric permittivity distribution corresponding to the individual molecular components. With this information, one can assemble a grid defining the boundaries between water and protein, as illustrated in Figure 11.22. The current density $j_\pm$ arising from the flow of ions down the electrochemical gradient in the aqueous region of the domain is given by the drift-diffusion equation:

$$\vec{j}_\pm = -(\mu_\pm \rho_\pm \vec{\nabla}\varphi - D_\pm \vec{\nabla}\rho_\pm)$$

where $\rho_\pm$ are the ionic charge densities and $\mu_\pm$ and $D_\pm$ are, respectively, the mobilities and diffusion coefficients of each ionic species. For the purposes of this discussion, we restrict ourselves to systems with only two ionic species of opposite charge, but the same treatment can be extended to allow for multiple ionic species by including a drift-diffusion equation for each additional species. Conservation of charge is enforced by a continuity equation for each species, given by

$$\vec{\nabla} \cdot \vec{j}_\pm + \frac{\partial \rho_\pm}{\partial t} = S_\pm$$

The term $S_\pm$ is set to zero for simple transport simulation, but it can be set to any functional form to describe higher order effects, such as the details of ion binding and other chemical phenomena that populate or deplete the ion densities. The electrostatic potential φ is described by Poisson's Equation:

$$\vec{\nabla} \cdot (\varepsilon \vec{\nabla}\varphi) = -(\rho_{fixed} + \rho_+ + \rho_-)$$

where ρ_{fixed} represents the density of fixed charge residing within and on the surface of the protein. When solved simultaneously, this system of coupled equations provides a self-consistent description of ion flow in the channel. The equations are discretized on the grid and solved iteratively for steady-state conditions,

© 2003 by CRC Press LLC

subject to specific boundary conditions for applied potential and for ionic solution concentrations in the baths at the ends of the channel. In a typical semiconductor device, the mobile charge in the contacts is originated by fixed ionized dopants. In an ion-channel system, the salt concentration in the electrolyte far from the protein determines the density of mobile ionic charges, which, from an electrical point of view, behave similarly to the intrinsic electron/hole concentrations in an undoped semiconductor at a given temperature.

11.5.2.1 Application of the Drift-Diffusion Model to Real Ion Channels

Complete three-dimensional drift-diffusion models have been implemented using the computational platform PROPHET[83] and used to study transport in ion channels like gramicidin and porin, for which detailed structure and conductivity measurements are available. Porin in particular presents a very challenging problem because, as shown in Figure 11.21, the channel is a *trimer* consisting of three identical parallel channels, connected through a common anti-chamber region. Memory requirements for continuum simulations of porin are currently at the limit of available workstation resources; however, simulations are now performed routinely on distributed shared memory machines. Figure 11.23 compares the current–voltage curves computed with a three-dimensional drift-diffusion simulation with those measured experimentally.[84] These results were generated in approximately 8 hours on an SGI origin2000.

11.5.3 Monte Carlo Simulations

The Monte Carlo simulation technique, as it is known in the tradition of semiconductor device simulation, can be coupled with a particle-mesh model to provide a self-consistent, time-resolved picture of ion dynamics in a channel system.[82,85] The starting point is the grid, which defines the regions accessible to ions as well as the dielectric topography of the system. In reality the boundaries between aqueous, protein, and membrane regions are not static but move over atomic length scales due to the thermal fluctuations of the atoms of the protein. Such fluctuations, which are resolved in MD simulations, are ignored in Monte Carlo simulations (although in principle they could be included).

11.5.3.1 Resolving Single-Ion Dynamics

Ions are distributed throughout the aqueous region according to a given initial concentration profile. The charge of each mobile ion, and of each static charge within the protein, is interpolated to the grid using a prescribed weighting scheme to construct a charge density at the discrete grid points. The

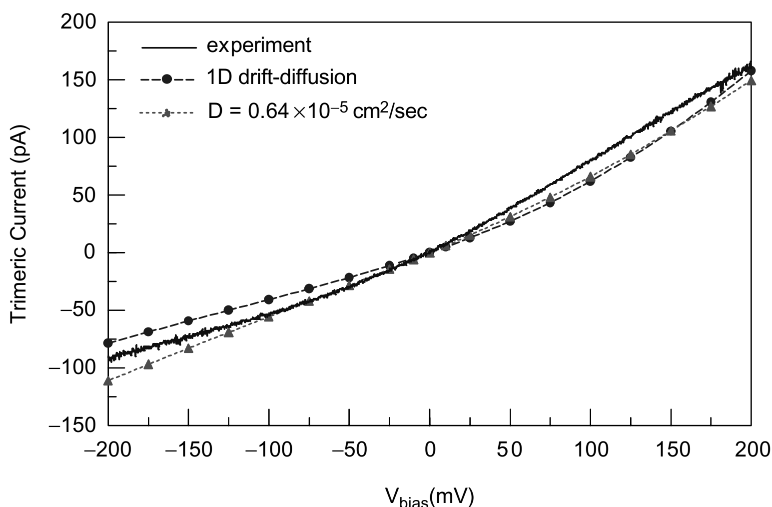


FIGURE 11.23 Comparison of measured and computed current–voltage curves for *ompF* in 100 mM potassium chloride, assuming a spatially dependent diffusion coefficient and a spatially uniform diffusion coefficient.

© 2003 by CRC Press LLC

electrostatic field due to the charge density distribution, as well as any externally applied field, is found by solving Poisson's equation on the grid. The field at the grid points is interpolated back to the ion positions and used to move the ions forward in time by integrating Newton's Second Law over small time steps. At the end of the time step, the new ion positions are used to recalculate a new charge density distribution and hence a new field to advance the ions over the next timestep. This cycle is iterated either until a steady-state is reached or until quantities of interest (e.g., diffusion coefficient) have been calculated. The effects of ion volume can also be incorporated by including an ionic core repulsive term in the force, acting on each ion as is done in Brownian Dynamics.

11.5.3.2 Modeling Ion-Water Interactions

Ion motion is treated as a sequence of free flights interrupted by collisions with water molecules, which are modeled by assuming a particular ion-water scattering rate $\nu(E_{ion}(t))$, generally a function of ion energy. The scattering rate represents the average number of collisions per unit time that an ion would experience if it maintained a constant energy. The probability for an ion to travel for a time t without scattering is given by

$$P(t) = \exp\left(-\int_0^t \nu(E_{ion}(t'))dt'\right)$$

The probability density function (probability per unit time) for a flight to have duration t is given by

$$p(t) = \nu(E_{ion}(t))P(t)$$

Ion flight times can be randomly selected from the probability density function by integrating the latter over the (unknown) flight time T_f and equating the integral to a uniformly distributed random number r on the unit interval. Thus,

$$-\log(r) = \int_0^{T_f} \nu(E_{ion}(t))dt$$

The integral on the right-hand side is trivial only for constant scattering rates, but in general it cannot be performed analytically. A number of methods have been introduced to solve the integral; an extended discussion is given at the Internet location given in Reference 82.

11.6 Conclusions

The combination of a three-dimensional drift-diffusion and three-dimensional Monte Carlo approach provides the essential hierarchy for looking at biological systems from the point of view of device-like applications. There are, however, significant differences between solid-state devices and biological systems, which require different choices in the definition of a Monte Carlo simulation strategy. In a typical device, the ensemble must include many thousands of particles; but a reasonable steady-state is reached after several picoseconds of simulation (on the order of ten to twenty thousand time steps). In a practical simulation domain for a biological channel, only a very small number of ions is present in the system; but because the ion traversal of the channel is a rare event, measurable current levels can only be established by extending the simulation to the millisecond range. Because the number of time steps required to resolve ion dynamics is typically on the order of tens of femtoseconds, this would require a number of iteration steps on the order of 10^{12}, which is still extremely expensive. For a fully self-consistent simulation, the solution of Poisson's Equation in three-dimensions presents the real bottleneck, while the computational cost of resolving the few particle trajectories is minimal. Alternative schemes for evaluating the electrostatic potential self-consistently include precalculating the potential for various ion

© 2003 by CRC Press LLC

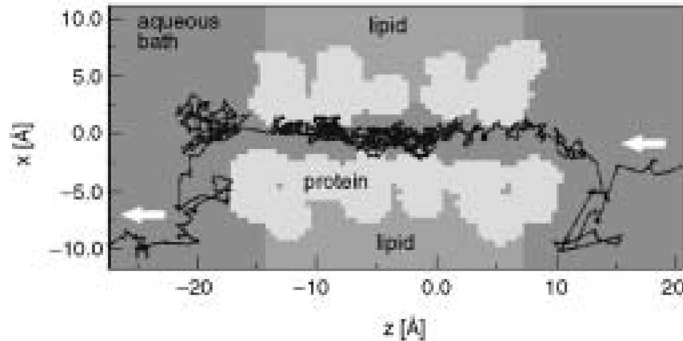


FIGURE 11.24 Geometric representation of the gramicidin channel as used in three-dimensional Monte Carlo transport model. The successful trajectory of a single sodium ion traversing the channel is shown.

pair configurations, storing the results in look-up tables, and employing the superposition principle to reconstruct the potential at the desired point by interpolating between table entries.[86] A prototype Monte Carlo simulation of sodium chloride transport in the gramicidin channel has been successfully implemented,[87] adapting the grid developed for the continuum simulations, as shown in Figure 11.24. For this simulation Poisson's Equation is solved approximately every 10 time steps using an accurate conjugate gradient method.

Acknowledgments

N. R. Aluru is thankful to M. Dequesnes for technical assistance and allowing the use of some of his results. J.-P. Leburton is indebted to G. Austing, R. M. Martin, S. Nagaraja, and S. Tarucha for fruitful discussions, and to P. Matagne for technical assistance. U. Ravaioli and T. van der Straaten are indebted to Bob Eisenberg (Rush Medical College) and Eric Jakobsson (University of Illinois) for introducing them to the subject of ion channels and for many useful discussions; to R. W. Dutton, Z. Yu, and D. Yergeau (Stanford University) for assistance with PROPHET; and T. Schirmer (University of Basel) for providing molecular structures for the porin channel. S. V. Rotkin is indebted to L. Rotkina and I. Zharov for fruitful discussions and to K. A. Bulashevich for technical support. N. R. Aluru, J.-P. Leburton, and S. V. Rotkin acknowledge the support of the CRI grant of UIUC. The work of J.-P. Leburton was supported by NSF grants DESCARTES ECS-98-02730, the Materials Computational Center (DMR-99–76550) and the DARPA-QUIST program (DAAD19–01–1–0659). The work of U. Ravaioli and T. van der Straaten was funded in part by NSF Distributed Center for Advanced Electronics Simulation (DesCArtES) grant ECS 98–02730, DARPA contract F30 602-012-0513 (B.E.), and NSF KDI grant to the University of Illinois. S. V. Rotkin acknowledges support of DoE grant DE-FG02-01ER45932 and the Beckman Fellowship from the Arnold and Mabel Beckman Foundation. For his calculations, B. R. Tuttle primarily utilized the SGI-ORIGIN2000 machines at the National Center for Supercomputing Applications in Urbana, IL. K. Hess acknowledges the Army Research Office (DAAG55-98-1-03306) and the Office of Naval Research (NO0014-98-1-0604) and ONR–MURI.

References

1. Hess, K., *Advanced Theory of Semiconductor Devices*, IEEE Press, 2000.
2. Hu, C., Tam, S.C., Hsu, H., Ko, P., Chan, T., and Terrill, K.W., *IEEE Trans. Electron. Devices* ED-32, 375, 1985.
3. Tuttle, B.R., Hydrogen and PB defects at the Si(111)-SiO2 interface: an *ab initio* cluster study, *Phys. Rev. B* 60, 2631, 1999.
4. Tuttle, B.R. and Van de Walle, C., Structure, energetics and vibrational properties of Si-H bond dissociation in silicon, *Phys. Rev. B* 59, 12884, 1999.

© 2003 by CRC Press LLC

5. Hess, K., Tuttle, B.R., Register, L.F., and Ferry, D., Magnitude of the threshold energy for hot-electron damage in metal oxide semiconductor field-effect transistors by hydrogen desorption, *Appl. Phys. Lett.* 75, 3147, 1999.
6. Tuttle, B.R., Register, L.F., and Hess, K., Hydrogen related defect creation at the Si-SiO2-Si interface of metal-oxide-semiconductor field-effect transistors during hot electron stress, *Superlattices Microstruct.* 27, 441, 2000.
7. Tuttle, B.R., McMahon, W., and Hess, K., Hydrogen and hot electron defect creation at the Si(100)-SiO2 interface of metal-oxide-semiconductor field effect transistors, *Superlattices Microstruct.* 27, 229, 2000.
8. Tuttle, B.R., Energetics and diffusion of hydrogen in SiO2, *Phys. Rev. B* 61, 4417, 2000.
9. Hess, K., Register, L.F., McMahon, W., Tuttle, B.R., Ajtas, O., Ravaioli, U., Lyding, J., and Kizilyalli, I.C., Channel hot carrier degradation in MOSFETs, *Physica B* 272, 527, 1999.
10. Hess, K., Haggag, A., McMahon, W., Cheng, K., Lee, J., and Lyding, J., The physics of determining chip reliability, *IEEE Circuits Device* 17 (3), 33–38, 2001.
11. Van de Walle, C. and Tuttle, B.R., Microscopic theory of hydrogen in silicon devices, *IEEE Trans. Electron. Devices* 47, 1779, 2000.
12. Haggag, A., McMahon, W., Hess, K., Cheng, K., Lee, J., and Lyding, J., *IEEE Intl. Rel. Phys. Symp. Proc.*, 271, 2001.
13. Staedele, M., Tuttle, B.R., and Hess, K., Tunneling through ultrathin SiO_2 gate oxide from microscopic models, *J. Appl. Phys.* 89, 348, 2002.
14. Staedele, M., Fischer, B., Tuttle, B.R., and Hess, K., Influence of defects on elastic gate tunneling currents through ultrathin SiO2 gate oxides: predictions from microscopic models, *Superlattices Microstruct.* 28, 517, 2000.
15. Staedele, M., Tuttle, B.R., and Hess, K., Tight-binding investigation of tunneling in thin oxides, *Superlattices Microstruct.* 27, 405, 2000.
16. Staedele, M., Tuttle, B.R., Fischer, B., and Hess, K., Tunneling through ultrathin thin oxides — new insights from microscopic calculations, *Intl. J. Comp. Electr.*, 2002.
17. Staedele, M., Fischer, B., Tuttle, B.R., and Hess, K., Resonant electron tunneling through defects in ultrathin SiO2 gate oxides in MOSFETs, *Solid State Electronics*, 46, 1027–1032 (2001).
18. Klimeck, G., Bowen, R.C., and Boykin, T., Off-zone-center or indirect band-gap-like hole transport in heterostructures, *Phys. Rev. B* 63, 195310, 2001.
19. DiVentra, M. and Pantelides, S., Hellmann–Feynman theorem and the definition of forces in quantum time-dependent and transport problems, *Phys. Rev. B* 61, 16207, 2000.
20. Damle, P.S., Ghosh, A.W., and Datta, S., Unified description of molecular conduction: from molecules to metallic wires, *Phys. Rev. B* 64, 16207, 2001.
21. Kasner, M.A., *Phys. Today* 46, 325, 1993.
22. Averin, D.V. and Likharev, K.K., in Lee, P.A., and Webb, R.A. (Eds.), *Mesoscopic Phenomena in Solids*, Elsevier, Amsterdam, 1991, pp. 173–271.
23. Devoret, M.H. and Grabert, H., in Devoret, M.H. and Grabert, H. (Eds.), *Single Charge Tunneling: Coulomb Blockade Phenomena in Nanostructures*, Plenum Press, 1991, pp. 1–19.
24. Likharev, K.K., *Proc. IEEE* 87, 606, 1999.
25. Meirav, V. and Foxman, E.B., *Semicond. Sci. Technol.* 10, 255, 1995.
26. Ashoori, R.C., Stoermer, H.L., Weiner, J.S., Pfeiffer, L.N., Pearton, S.J., Baldwin, K.W., and West, K.W., *Phys. Rev. Lett.* 68, 3088, 1992.
27. Ashoori, R., *Nature* 379, 413, 1996.
28. Tarucha, S., Austing, D.G., Honda, T., van der Hage, R.J., and Kouwenhoven, L.P., *Phys. Rev. Lett.* 77, 3613, 1996.
29. Austing, D.G., Honda, T., and Tarucha, S., *Jpn. J. Appl. Phys.* 36, 4151, 1997.
30. Sze, S.M., *Physics of Semiconductor Devices*, 2nd ed., John Wiley & Sons, New York, 1981.
31. Vincenzo, D.P.D., *Nature* 393, 113, 1998.
32. Arakawa, Y. and Yariv, A., *IEEE J. Quant. Electron.* 22, 1887, 1986.

© 2003 by CRC Press LLC

33. Bimberg, D., Grundmann, M., and Ledentsov, N.N., *Quantum Dot Heterostructures*, Wiley, London, 1998.
34. Johnson, N.F., *J. Phys. Condens. Matter* 7, 965, 1995.
35. Maccucci, M., Hess, K., and Iafrate, G.J., *J. Appl. Phys.* 77, 3267, 1995.
36. Nagaraja, S., Matagne, P., Thean, V.Y., Leburton, J.-P., Kim, Y.-H., and Martin, R.M., *Phys. Rev. B* 56, 15752, 1997.
37. Leonard, D., Pond, K., and Petroff, P.M., *Phys. Rev. B* 50, 11687, 1994.
38. Miller, M.S., Malm, J.O., Pistol, M.E., Jeppesen, S., Kowalski, B., Georgsson, K., and Samuelson, L., *Appl. Phys. Lett.* 80, 3360, 1996.
39. Madelung, O., *Introduction to Solid State Physics*, Springer Verlag, Berlin, 1978.
40. Jones, R.O. and Gunnarson, O., *Rev. Mod. Phys.* 61, 689, 1989.
41. Kumar, A., Laux, S.E., and Stern, F., *Phys. Rev. B* 42, 5166, 1990.
42. Stopa, M., *Phys. Rev. B* 54, 13767, 1996.
43. Koskinen, M., Manninen, M., and Rieman, S.M., *Phys. Rev. Lett.* 79, 1389, 1997.
44. Jovanovic, D. and Leburton, J.-P., *Phys. Rev. B* 49, 7474, 1994.
45. Lee, I.H., Rao, V., Martin, R.M., and Leburton, J.-P., *Phys. Rev. B* 57, 9035, 1998.
46. Nagaraja, S., Leburton, J.-P., and Martin, R.M., *Phys. Rev. B* 60, 8759, 1999.
47. Perdew, J.P. and Zunger, A., *Phys. Rev. B* 23, 5048, 1981.
48. Slater, J.C., *Quantum Theory of Molecules and Solids*, McGraw-Hill, New York, 1963.
49. Matagne, P., Leburton, J.-P., Austing, D.G., and Tarucha, S., to be published.
50. Blick, R.H., Haug, R.J., Weis, J., Pfannkuche, D., Klitzing, K., and Eberl, K., *Phys. Rev. B* 53, 7899, 1996.
51. Kouwenhoven, L.P., *Science* 268, 1440, 1995.
52. Zener, C., *Phys. Rev.* 81, 440, 1951.
53. Wensauer, A., Steffens, O., Suhrke, M., and Roessler, U., *Phys. Rev. B* 62, 2605, 2000.
54. Imamaura, H., Maksym, P.A., and Aoki, H., *Phys. Rev. B* 59, 5817, 1999.
55. Rotani, M., Rossi, F., Manghi, F., and Molinari, E., *Solid State Commun.* 112, 151, 1999.
56. Bender, C.F. and Davidson, E.R., *J. Chem. Phys.* 46, 3313, 1967.
57. Loss, D. and Di Vincenzo, D.P., *Phys. Rev. A* 57, 1998.
58. Dresselhaus, M.S., Dresselhaus, G., and Eklund, P.C., *Science of Fullerenes and Carbon Nanotubes,* Academic Press, 1996.
59. Wildoer, J.W.G., Venema, L.C., Rinzler, A.G., Smalley, R.E., and Dekker, C., Electronic structure of atomically resolved carbon nanotubes, *Nature* 391 (6662), 59–62, 1998.
60. Kuzmany, H., Plank, W., Hulman, M., Kramberger, C., Gruneis, A., Pichler, T., Peterlik, H., Kataura, H., and Achiba, Y., Determination of SWCNT diameters from the Raman response of the radial breathing mode, *Eur. Phys. J. B* 22 (3), 307–320, 2001.
61. Bachtold, A., Hadley, P., Nakanishi, T., and Dekker, C., Logic circuits with carbon nanotube transistors, *Science* 294 (5545), 1317–1320, 2001.
62. Ebbesen, T.W., Potential applications of nanotubes, in *Carbon Nanotubes*, CRC Press, Boca Raton, FL, 1997, pp. 296.
63. Falvo, M.R., Steele, J., Taylor, R.M., and Superfine, R., Gearlike rolling motion mediated by commensurate contact: carbon nanotubes on HOPG, *Phys. Rev. B* 62 (16), R10665–R10667, 2000.
64. Yu, M.F., Files, B.S., Arepalli, S., and Ruoff, R.S., Tensile loading of ropes of single wall carbon nanotubes and their mechanical properties, *Phy. Rev. Lett.* 84 (24), 5552–5555, 2000.
65. Hertel, T., Walkup, R.E., and Avouris, P., Deformation of carbon nanotubes by surface van der Waals forces, *Phys. Rev. B-Condensed Matter* 58 (20), 13870–13873, 1998.
66. Sanchez–Portal, D., Artacho, E., Solar, J. M., Rubio, A., and Ordejon, P., *Ab initio* structural, elastic, and vibrational properties of carbon nanotubes, *Phys. Rev. B-Condensed Matter* 59 (19), 12678–12688, 1999.
67. Yakobson, B.I. and Avouris, P., Mechanical properties of carbon nanotubes, *Carbon Nanotubes: Synthesis, Structure, Properties, Applications* 80, 287–327, 2001.

© 2003 by CRC Press LLC

68. Yakobson, B.I., Brabec, C.J., and Bernholc, J., Nanomechanics of carbon tubes — instabilities beyond linear response, *Phys. Rev. Lett.* 76 (14), 2511–2514, 1996.
69. Rotkin, S.V. and Hess, K., Many-body terms in van der Waals cohesion energy of nanotubules, *J. Comp. Electr.*, 1, 294–297 (2002).
70. Dequesnes, M., Rotkin, S.V., and Aluru, N.R., Calculation of pull-in voltages for carbon nanotube-based nanoelectromechanical switches, *Nanotechnology* 13, 120–131, 2002.
71. Girifalco, L.A., Hodak, M., and Lee, R.S., Carbon nanotubes, buckyballs, ropes, and a universal graphitic potential, *Phys. Rev. B-Condensed Matter* 62 (19), 13104–10, 2000.
72. Lennard–Jones, J.E., Perturbation problems in quantum mechanics, *Proc. R. Soc London, Ser. A* 129, 598–615, 1930.
73. Rotkin, S.V., Bulashevich, K.A., and Aluru, N.R., Atomistic models for nanotube device electrostatics, in *ECS Centennial Meet,* ECS, Philadelphia, 2002, p. V5–1164.
74. Odintsov, A.A. and Tokura, Y., Contact phenomena and Mott transition in carbon nanotubes, *J. Low Temp. Phys.* 118 (5–6), 509–518, 2000.
75. Luryi, S., Quantum capacitance devices, *Appl. Phys. Lett.* 52 (6), 501–503, 1988.
76. Derycke, V., Martel, R., Appenzeller, J., and Avouris, P., Carbon nanotube inter- and intramolecular logic gates, *Nano Lett.* 1 (9), 453–456, 2001.
77. Hille, B., *Ionic Channels of Excitable Membranes*, Sinauer Associates, Massachusetts, 1992.
78. Phale, P.S., Philippsen, A., Widmer, C., Phale, V.P., Rosenbusch, J.P., and Schirmer, T., Role of charged residues at the ompF porin channel constriction probed by mutagenesis and simulation, *Biochemistry* 40, 6319–6325, 2001.
79. Reif, F., *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill, Singapore, 1987.
80. Schuss, Z., Nadler, B., and Eisenberg, R.S., Derivation of Poisson and Nernst-Planck equations in a bath and channel from a molecular model, *Phys. Rev. E* 64, 036116-(1–14), 2001.
81. Allen, M.P. and Tildesley, D.J., *Computer Simulation of Liquids,* Clarendon, Oxford, 1987.
82. http://www.ceg.uiuc.edu/ncce.htm
83. http://www-tcad.stanford.edu/
84. van der Straaten, T., Varma, S., Chiu, S.-W., Tang, J., Aluru, N., Eisenberg, R., Ravaioli, U., and Jakobsson, E., Combining computational chemistry and computational electronics to understand protein ion channels, in *The 2002 Intl. Conf. Computational Nanosci. Nanotechnol.*, 2002.
85. Hockney, R.W. and Eastwood, J.W., *Computer Simulation Using Particles,* McGraw-Hill, 1981.
86. Chung, S.-H., Allen, T.W., Hoyles, M., and Kuyucak, S., Permeation of ions across the potassium channel: Brownian dynamics studies, *Biophys. J.* 77, 2517–2533, 1999.
87. van der Straaten, T. and Ravaioli, U., Self-consistent Monte-Carlo/P3M simulation of ion transport in the gramicidin ion channel, unpublished.

© 2003 by CRC Press LLC