Introduction to Nanoscale Science and Technology

August 9, 2001

1 About the course

The study and manipulation of matter on the nanometer scale is a thriving area of research, with profound implications for technology (e.g. nanoelectronics, nanostructured materials, nanobiology) and pure science (e.g What is the nature of the transition from quantum to classical behavior?). The aim of this class is to familiarize students with the main issues and techniques relevant to physics on the nanometer scale.

We'll begin with some introductory lectures, in which I'll review some basic condensed matter physics concepts, possibly in a context that not all of you have seen. The sequence of topics will be:

- A brief review of solid state physics
- Fabrication of nanostructures
- Characterization and nanoscale probes

For the rest of the course, we will frame our discussions around particular scientific questions, and using recent papers in the literature to answer them. (I give you fair warning that the selection of papers will have a definite experimental slant. In part this is because of my own bias as an experimentalist; generally, however, the reason is that experiment has really been the driving innovative force in nanoscale work over the last twenty years. Very few of the phenomena we will study were predicted in advance of experimental observation.) The idea is to do this in a seminar format. That is, everyone will get copies of the relevant article(s) for the current topic, and one person will give a presentation (around 30 minutes, with transparencies) describing the work and how it relates to the question at hand. Before the presentation, I will give a brief review of the essential physics issues. After the talk, we will discuss the findings and assess what we've learned regarding our question. I'm going to do the first couple of these presentations, and then you're going to take turns doing them. Giving clear, effective talks like this is a tremendously valuable skill, and the more practice you get at it, the better!

When reading the papers we'll be discussing, besides learning some cool nanoscale physics, I want you to ask yourself some questions about each paper. Try to answer:

- *What are the energy/time/length scales of interest here?*

  As you will come to appreciate, the relative sizes of important energies, characteristic times, and characteristic lengths strongly influence both the physics that we study, and the way in which we study it.

- *What is new about this measurement?*

  New physics results and insights in this field have consistently been driven by the development of new measurement techniques, fabrication methods, and material systems. As we read various papers, ask yourself what distinguishes this measurement from previous work: why hadn't the knowledge gained here been gleaned in previous investigations?

- *What do we know now that we didn't know before?*

  That is, what is the main result?

- *Does the conclusion depend critically on theory?*

  It's useful to pay attention to the relationship between the data taken in the experiment, and the conclusions drawn from that data.

- *What are the next steps?*

  Often very good experiments raise more questions than they really answer. Consider what the next logical move is, given the results in the paper.

Some of the questions we will examine include:

- How can we fabricate devices on the nanometer scale?

- What measurement techniques allow us to study such devices?

- What length, energy, and time scales are relevant?

- What are quantum contributions to electrical properties?
- Why don't we see such quantum effects all the time?
- Do we understand electron correlations on these scales?
- Can individual molecules be used as circuit elements?

As the end of the term approaches, you shall pick a topic (from a selection I'll provide, though you're welcome to suggest one of your own) for a final paper (< 10 pages). The choices are going to reflect some of the major outstanding problems and active research areas. Your task will be to do some literature research, briefly describe the main physical ideas, and summarize the state of our understanding.

2 Why nanoscale?

There are two main reasons for the vigor of research at nanometer length scales, technology and new physics.

2.1 Technology

A clear motivation behind much nanoscale work is the ongoing drive toward miniaturization of electronic components. Since its invention in 1948, the transistor has shrunk in size by almost seven orders of magnitude, or nearly fourteen orders of magnitude in areal density. The pace of this change has been truly remarkable, and prompted Gordon Moore, founder of Intel, to point out what has now become popularly known as Moore's Law: the typical linear dimension of a feature on an integrated circuit decreases in size by a factor of two roughly every eighteen months.

Obviously the size trend cannot continue forever, since matter is composed of atoms on the order of 0.1 nm in radius. Further, years before feature sizes are extrapolated to hit the atomic limit, typical device designs will cease to work correctly for two reasons, one mundane, the other interesting. The boring reason is that reliability of traditional device designs will begin to suffer terribly as the positions of individual atoms become increasingly important. The interesting reason is that as devices shrink their length scales start to cross physically significant threshholds; physical processes that could be neglected in larger devices can become dominant, crucially affecting device performance.

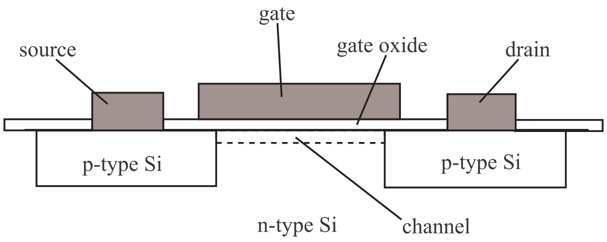

Figure 1: Schematic cross-section of a MOSFET.

One example of both of these issues is the classic metal oxide semiconductor field effect transistor (MOSFET) shown in Figure 1. This three-terminal device, the mainstay of the microelectronics industry, consists of source and drain electrodes separated by a channel region. Isolated from this channel region by a very thin SiO_2 layer is a gate electrode. When a bias voltage is applied to this gate, charge is either depleted or accumulated (depending on the sign of the bias) in an extremely thin layer in the channel. Ignoring some subtleties for now, when the charge layer is present, current can flow between the source and drain, and the transistor is "on"; similarly, when the charge layer is depleted completely, the transistor is "off." There are compelling engineering reasons for wanting to make these devices as small as possible. The shorter the channel length, the smaller the capacitance being charged by the gate electrode, which means the smaller the effective RC time constant of the gate, which in turn implies faster switching. Similarly, a thinner gate oxide, while implying higher gate capacitance, also means lower voltages are necessary to fully turn the transistor on or off, implying lower power consumption. The channel length of the MOSFETs in a state-of-the-art microprocessor is now a little over 100 nm, while the insulating gate oxide in such a device is currently around 3 nm thick.

With continuing miniaturization, however, several failure modes begin to become important. From the device reliability standpoint, thinner gate oxides mean trouble. Because of the chemical bonding structure of the Si/SiO_2 interface, even for atomically flat interfaces oxides thinner than 1.3 nm will be leaky to tunneling processes and unable to meet reliability criteria. Further, for finite source-drain bias shorter channel lengths can run into leakage current problems from tunneling, meaning transistors won't turn off all the way. Finally, as operating voltages are decreased to avoid gate oxide leakage problems, individual charged impurities in the oxide can strongly influence device performance. These issues and others like them have led some to make dire predictions of the imminent end of Moore's Law. As you can see, understanding the properties of matter on these scales is crucial to the future development of the microelectronics industry!

Other technologies besides the semiconductor industry are deeply interested in nanoscale science. The potential for interfacing biology and nanoscience is only beginning to be tapped. The fabrication and characterization techniques that we will be discussing are beginning to be applied to biological systems with impressive results. For example, companies like Affymetrix are developing "gene chips", microfabricated devices with chemically and biologically tailored sites for sorting DNA, RNA, and other biomaterials. A number of researchers are developing biologically compatible microdevices for applications like timed drug release, biomonitoring, and even prosthetic retinas. Knowledge has flowed in both directions also. Researchers are exploring biologically inspired ideas like self-assembly for the fabrication of nanostructed devices and materials. Some pie-in-the-sky pundits have gone so far as to predict that bio-like nanotechnology, with molecular-scale assemblers and other machines, will soon lead to nanomachines repairing the cells in our bodies.

Materials scientists are also closely involved in nanoscale work. Designer nanostructured materials have the potential to offer unprecidented combinations of mechanical, electrical, and optical properties. Examples include nanotube-based composites, ultrahard nanostructured alloys, and photonic band-gap materials.

2.2 New physics

By "new physics" I mean that our ability to controllably fabricate, manipulate, and examine strutures on the nanometer scale has lead to the observation of physical phenomena that simply don't arise elsewhere. Really distinguishing this from techological motivation is sometimes difficult, since technologists are always on the lookout for properties that may be adapted to some useful end.

Often these new phenomena are related to the crossing of characteristic energy scales. An example of this that we will discuss at some length is *Coulomb blockade*, an electrical transport effect that is sensitive to the motion of single charges. In this case the competing energy scales are thermal ($k_{\mathrm{B}}T$), the charging energy of a small capacitor C ($U_C = e^2/2C$), and the spacing of individual electronic levels (Δ). By shrinking device lengths, both the charging energy and level spacing can become comparable to the thermal energy scale, and Coulomb blockade effects can be detected.

Similarly, length scale crossovers can also have dramatic effects. Perhaps the best example of this is found in quantum corrections to electrical conduction. For now, let us define a "phase coherence length" L_ϕ as the distance over which a particular electron in a conductor can exhibit quantum interference. Once conductors are made with dimensions comparable to L_ϕ, dramatic effects such as *universal conductance fluctuations* and *localization* can be seen.

As the course progesses, we'll come across a number of interesting physical effects and get a better idea of the sorts of new phenomena that we can expect at these scales.

3 Length, energy, and time scales

Here we'll briefly list and define some of the length, energy, and time scales that are relevant to the papers we'll be reading. This is also a convenient time to introduce notation for these quantities; I'll try to be consistent throughout the course and avoid using the same symbol for multiple meanings. Note that I tend to refer to "electrons" rather than the more general expression "charge carriers." I'll try to be careful about this in situations where this might cause confusion (such as hole-doped systems).

3.1 Length scales, and what do we mean by "nanoscale"?

Literally, the term "nanoscale" implies that one or more dimension of the systems under consideration will be on the order of nanometers, $\sim 10^{-9}$ m. Just to give a sense of perspective, here's a list of some size scales for reference.

- $\sim 10^{-15}$ m
  atomic nucleus radius

- $\sim 10^{-10}$ m
  typical atomic radius

- $\sim 10^{-9}$ m
  diameter of single walled carbon nanotube
  thickness of thinnest insulating layer on Pentium

- $\sim 10^{-8}$ m
  limits of electron beam lithography
  size of grains in typical evaporated Au film
  Bohr radius of donor in semiconductor

- $\sim 10^{-7}$ m
  deep UV light wavelength, limits of photolithography
  width of smallest lateral features on Pentium
  Fermi wavelength in typical 2D semiconductor

- $\sim 10^{-6}$ m
  IR light wavelength

- $\sim 10^{-4}$ m
  roughly 0.004 inches
  diameter of a thick human hair
  wavelength of THz radiation

Here are a number of length scales that we'll run into during this course.

a $\equiv$ interatomic spacing in solids. Typically ~ 2 Å.

λ_{F} $\equiv$ the Fermi wavelength; wavelength of carriers that dominate electrical transport. In semiconductors, may be as long as 100 nm, while in traditional metals is more like 1 Å.

ℓ_{e} $\equiv$ electron elastic mean free path; classically, how far the electron travels before elastically scattering off a defect, impurity, or grain boundary. When considering electrons as plane waves, the typical distance over which the wave propogates before diffracting off a scatterer. Generally considered a temperature-independent property of a material, intimately related to the microstructure. Can be as large as 20 μm in clean two-dimensional electron gas (2DEG), or as short as a few Å in a highly disordered alloy.

ℓ_{i} $\equiv$ electron *inelastic* mean free path. Typical distance the electron travels before experiencing an inelastic scattering event. Definitely temperature-dependent because inelastic process rates depend on the product of temperature-dependent factors, as we shall see.

$\lambda_{\mathrm{T}} \equiv$ wavelength of typical thermal phonon. Depends on speed of sound and temperature.

$L_\phi \equiv$ phase coherence length. At distances shorter than L_ϕ, quantum interference of electron paths must be considered, while over longer distances, the electrons lose their phase memory. Can be tens of μm in 2DEG and up to several μm in regular metals. Usually strongly temperature dependent. We'll go over this in considerably more detail.

$L_{\mathrm{T}} \equiv$ thermal length. Given a wavepacket built out of electrons with a spread in energy $k_{\mathrm{B}}T$, this is the distance the wavepacket travels before significantly spreading out. This thermal smearing can mask quantum effects; again, more on this later. Often on the same order as L_ϕ.

$\xi \equiv$ the localization length. This is the typical size of a localized wavefunction (one that doesn't have wave-like components stretching off to infinity). Depends crucially on the nature of the disorder, of course.

$L_H \equiv$ the magnetic length. Within a factor of order unity, in problems involving magnetic fields, a square of side L_H will enclose one flux quantum: $H \times L_H^2 = \hbar/e$ in SI units. Roughly 25 nm at 1 Tesla, though some people use slightly different definitions.

$r_{\mathrm{c}} \equiv$ the cyclotron radius.

$r_{\mathrm{e}} \equiv$ typical electron-electron separation.

$d \equiv$ a symbol representing some useful length in the problem under consideration.

$L \equiv$ sample length.

3.2 Energy scales

As you might guess, there are also a large number of energy scales that must be considered in understanding nanoscale systems. Below is a sample of some of the most commonly encountered. Be aware that people often refer to energies in other systems of units, such as Kelvin (tacitly dividing by k_{B}), Volts (dividing by e), or even Hz (dividing by $\hbar$).

$k_{\mathrm{B}}T \equiv$ the thermal energy scale. At room temperature, about 4×10^{-21} J, or 26 meV.

E_F $\equiv$ the Fermi energy; energy of the highest occupied electron state. We'll review this shortly. For typical metals like Au, E_F is on the order of an eV

μ $\equiv$ the chemical potential; the amount of energy it takes to add a particle to the system

E_v $\equiv$ energy of the top of the valence band.

E_c $\equiv$ energy of the bottom of the conduction band. The difference between E_c and E_v is the *band gap*, and can vary between zero and several eV.

Φ $\equiv$ the work function; the energy needed to take a carrier from the Fermi level and promote it into the continuum. Usually on the order of an eV.

U_C $\equiv$ charging energy of a capacitance C, $e^2/2C$.

Δ_L $\equiv$ the spacing between (single-particle) energy levels in a box of size L.

E_T $\equiv$ the Thouless energy; roughly speaking, the energy scale of coherence effects. When the boundary conditions of a coherent region change, the energy levels in that region shift by an amount on the order of E_c.

$\mu_\mathrm{B}B$ $\equiv$ the Zeeman energy of a spin in a magnetic field.

eV $\equiv$ the energy an electron gains crossing a potential difference V.

3.3 Time scales

Unsurprisingly, there are many time scales that must be considered when analyzing the physics we'll be discussing. Often one can freely convert between lengths and times using relevant velocities, and any energy can be converted into a time by dividing by $\hbar$. Sometimes, though, physical intuition for a system is best developed by thinking in terms of times. Below we list some commonly encountered time scales.

τ $\equiv$ typical time between elastic scattering events. Can vary anywhere from 10^{-10} s in exceedingly clean systems to 10^{-15} s in very dirty metals.

$\tau_{\mathrm{i}} \equiv$ time between inelastic scattering events. Strongly temperature dependent since inelastic processes tend to be.

$\tau_{\phi} \equiv$ the *coherence* or *dephasing* time. The typical time scale over which quantum coherence of a single particle spatial state is lost due to interactions with environmental degrees of freedom. *Not* necessarily the same as τ_{i}. Can be as long as 10^{-7} s at very low temperatures in some materials. Under certain very special circumstances, can even be long at room temperature in rare cases; we'll get to this later in the course.

$\tau_{\mathrm{s}} \equiv$ the spin-flip scattering time. The time between scattering events involving an electron and a magnetic impurity such that the electron's spin is flipped. Depends in detail on the

$\omega_{\mathrm{c}}^{-1} \equiv$ the inverse cyclotron frequency; how long it takes an electron to complete a cyclotron orbit.